THE IMPACT OF TOBACCO PRICES ON SMOKING ONSET IN VIETNAM: DURATION ANALYSES OF RETROSPECTIVE DATA

G. Emmanuel Guindon[1,2*]

1. Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Ontario, Canada
2. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

* Corresponding author: Centre for Health Economics and Policy Analysis
Health Sciences Centre 3H1
McMaster University
1200 Main Street West
Hamilton, Ontario, L8N 3Z5
Tel: 1-905-525-9140, x22986
Fax: 1-905-546-5211
E-mail: guindoge@mcmaster.ca

Word count: 7632 words

JEL classification: I18; I12; C41

Keywords: Duration analysis; Smoking initiation; Tobacco prices; Vietnam

Acknowledgements: I thank Jeremiah Hurley, Michel Grignon and Michael Boyle for their comments and guidance, Joy de Beyer, Jinhu Li, Nguyen Tuan-Lam, Noori Akhtar-Danesh and members of the health polinomics seminar at McMaster University for helpful comments and Martin Forster and Andrew Jones for making their Stata codes publicly available.

Funding: During this study, GEG was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Canada Graduate Scholarship (CGS) and a Centre for Health Economics and Policy Analysis (CHEPA) Doctoral Fellowship. The SAVY price dataset was acquired with generous funding from the Bloomberg Philanthropies.

Competing Interests: GEG is a former employee of the World Health Organization—Tobacco Free Initiative whose objective is to reduce the global burden of disease and death caused by tobacco.

The views expressed in this paper are those of the author and do not represent the views of the above organizations.

Last updated: 27 October 2009

ABSTRACT

The benefits of preventing smoking onset are well known and even just delaying smoking onset conveys benefits. Tobacco control policies are of critical importance to low-income countries with high smoking prevalence such as Vietnam. Using a survey of teens and young adults (the 2003 Survey Assessment of Vietnamese Youth), I conduct duration analyses to explore the impact of tobacco prices on the onset of smoking. Results suggest that tobacco prices in Vietnam have a statistically significant and fairly substantial effect on the age of starting smoking. Increases in average tobacco prices (measured by an index of tobacco prices) and in the prices of two popular brands are found to delay smoking onset. Of particular interest is the finding that Vietnamese youth are more sensitive to changes in prices of a popular international brand that has had favorable tax treatment since the late 1990s.

Word count: 143 words

1. BACKGROUND AND RATIONALE

The benefits of preventing smoking onset are well known. The effects of smoking on adult health are substantial. Worldwide, tobacco use causes 5.4 million deaths each year, and the number is expected to rise to 6.4 million in 2015 and 8.3 million by 2030 (Mathers and Loncar, 2006). Tobacco use is directly responsible for lung cancer and is an important risk factor for a plethora of other illnesses, including numerous cancers, tuberculosis and diseases of the cardiovascular, respiratory and digestive systems (Doll, 1998; US Department of Health and Human Services, 2004). Smoking during childhood and adolescence is the cause of serious contemporaneous health problems such as cough and phlegm production, an increased number and severity of respiratory illnesses, decreased physical fitness, and potential retardation in the rate of lung growth and the level of maximum lung function (US Department of Health and Human Services, 1994).

Even just delaying smoking onset conveys benefits. Youth who initiate smoking at an early age are at an increased risk for developing long-term health consequences. In the United States, early smoking initiation has been associated with increased daily consumption of tobacco, longer smoking duration and nicotine dependence (Breslau, Fenn and Peterson, 1993; Breslau and Peterson, 1996; Everett et al., 1999). The association between early initiation and duration held even when nicotine dependence was taken into account (Breslau and Peterson, 1996).

Theory suggests that youth should respond as strongly, if not more strongly, than adults to changes in the price of tobacco products. This is due, firstly, to the addictive nature of tobacco products: longer term users are less able to curb consumption and therefore adjust less rapidly to changes in tobacco prices, compared to younger individuals who may not yet be addicted to nicotine (Lewit, Coate and Grossman, 1981). Second, peer behaviour determines youth smoking more than it affects adult smoking. An increase in price would reduce the number of young smokers and consequently reduce the impact of peer smoking on other young smokers, thus multiplying the effect of price changes (Lewit et al., 1981). Third, since young and poor smokers spend a larger share of their relatively smaller disposable income on tobacco than older and richer users they tend to be more responsive to tobacco price increases (Grossman and Chaloupka, 1997). Fourth, smoking is associated with an individual's rate of time preference. The young and poor tend to display higher rates of time preference –they tend to be more concerned with current costs (money prices) at the expense of longer term costs (health consequences) (Becker, Grossman and Murphy, 1991).

DeCicca et al. (DeCicca, Kenkel and Mathios, 2002), however, argue that a number of influences can offset these tendencies. If the impact of the income effect of a price change on youth versus adult differs, the total effect (income and substitution) of a price change could lead to a larger response among adults than youth. In addition, within the rational addiction framework, past, current and future consumption of an addictive good are complements (Becker et al., 1991). As past consumption reinforces current consumption which reinforces future consumption, the price response grows over time. Hence the greater the addiction, the greater the long-run price response. Finally, the peer effect can go in the opposite direction if cigarettes are the only inputs into the production of peer acceptance.

In 1999, the World Bank concluded, after an extensive review of the evidence, that total price elasticity estimates clustered about -0.4 in high-income countries and -0.8 in low- and middle-income countries (World Bank, 1999). In a meta-analysis of 86 studies (and more than 500 point estimates), Gallet and List (Gallet and List, 2003) find a mean (total) price elasticity of -0.48. Similarly, in an extensive review of both theoretical and empirical evidence, Chaloupka and Warner (Chaloupka and Warner, 2000) write:

> As the now-substantial body of economic research demonstrates, however, the demand for cigarettes clearly responds to changes in prices and other factors, as found in applications of both traditional models of demand and more recent studies that explicitly account for the addictive nature of smoking (Chaloupka and Warner, 2000 p. 1546).

The above reviews have important methodological limitations and have weaker generalizability to low- and middle-income countries: they provide limited quality assessment of data and methods utilized by the studies and include a relatively small number of studies conducted in low- and middle-income countries. As only a small number of studies examine the decision to initiate smoking, they do not provide definitive evidence regarding the impact of tobacco prices on smoking onset (as compared to smoking participation, smoking intensity or smoking cessation).

Approaches to examine the factors that influence youth smoking may focus on the decision to initiate smoking or the decision to be a current smoker (i.e. participation conditional on having initiated). The distinctions in these approaches are important. Approaches that model participation do not allow one to distinguish between former smokers who have quit smoking and those who have never smoked. The addictive nature of nicotine plays a critical role in the decision to continue smoking. In contrast, the role of addiction in the decision to initiate smoking is of lesser importance. Differences between participation and smoking onset elasticities vary with age –younger individuals are substantially more likely to initiate smoking (Douglas and Hariharan, 1994).

Determining the impact of tobacco prices on smoking onset in low- and middle-income countries is of critical importance given how young their populations are and given that many low- and middle-income countries are experiencing a rise in non-communicable diseases associated with tobacco use (Beaglehole and Yach, 2003). In Vietnam, 45% of the population is below the age of 25 (General Statistics Office, 2008b) and it is estimated that about 40 000 deaths are already attributable to tobacco use each year (Levy et al., 2006).

This paper reports the results of two analyses: the first is a systematic search for and critical review of studies that examine the effect of prices (or taxes) on the decision to initiate smoking; the second is an original analysis of the impact of tobacco prices on the onset of smoking in Vietnam. As secondary objectives, the second analysis also explores the impact of cigarette brand-specific and waterpipe tobacco prices and of factors other than prices that may be associated with smoking onset including but not limited to income/wealth, urban/rural status and smoking behaviour of family and peers.

2. REVIEW OF EXISTING EVIDENCE

I systematically searched for and critically reviewed studies that examine the effect of prices (or taxes) on the decision to initiate smoking. I searched two computerized bibliographic databases (MEDLINE via PubMed and EconLit), hand-searched four specialty journals (*Health Economics*, *Journal of Health Economics*, *Nicotine & Tobacco Research* and *Tobacco Control*) and examined references of recent reviews (Cameron, 1998; Chaloupka et al., 2000; Chaloupka and Warner, 2000; Gallet and List, 2003; Guindon, Perucic and Boisclair, 2003; Laporte, 2006; Thomas et al., 2008) and to identify unpublished studies, I also searched the unpublished literature via Google and Google Scholar. This systematic search yielded a total of twenty-five articles related to the effect of prices (or taxes) on the decision to initiate smoking, a substantially larger number than reviewed in any other single study. Nearly all studies, however, are conducted using data from the United States and to a lesser extent data from other high-income OECD countries (Australia, Canada, France, Great Britain, Ireland, Spain and Sweden). Only one study (Laxminarayan and Deolalikar, 2004) uses data from a low-income country (Vietnam). The generalizability of studies conducted in high-income economies to low-income settings, notably to countries with dramatically different patterns of tobacco use and tobacco control environments such as Vietnam, is fairly limited.

Most studies have important limitations, some serious enough that considerable caution is needed when interpreting results. As first highlighted by Forster and Jones (Forster and Jones, 2001), there is a general failure to apply diagnostic tests to assess the fit of the empirical models of smoking onset. For example, standard duration models (continuous and discrete) rest on the assumption that each individual will eventually fail (i.e. start smoking).[1] Similarly, how the issue of duration is incorporated into discrete-time model is seldom discussed. This can be important, as ignoring time dependency in the baseline hazard produces a model that is more or less equivalent to an exponential model (i.e., the hazard probability is flat with respect to time) (Box-Steffensmeier and Jones, 2004). When using retrospective data, imperfect recall by respondents (e.g., heaping around rounded values) can introduce substantial measurement errors (Tauras and Chaloupka, 1999). This can be problematic when respondents are asked to recall the exact year or age at which they initiated smoking when such events occurred decades earlier.[2] Even if recall is accurate, current location may

[1] Such an assumption is reasonable when mortality is the outcome under study but is problematic when smoking onset is the outcome being modeled because large proportion of individuals never start smoking. López Nicolás (López Nicolás, 2002) and Kidd and Hopkins (Kidd and Hopkins, 2004) compare models that relax the assumption that each individual will eventually fail with standard duration models (i.e. split and non-split population models) and find vastly different effects.

[2] Studies that may be particularly affected by this data limitation include Douglas and Hariharan (Douglas and Hariharan, 1994), Douglas (Douglas, 1998), Forster and Jones (Forster and Jones, 2001), Grignon and Pierrard (Grignon and

not match location at time of decision (Tauras, O'Malley, and Johnston 2001).[3] When using longitudinal data, current prices may not match those at the time of decision. With one exception (Cawley, Markowitz and Tauras, 2004), longitudinal data utilized have not been collected on an annual basis. Hence, most authors regress smoking onset, measured over a two-year period, on contemporaneous prices.[4]

Additional limitations include the limited variation in cigarette prices within US states[5]; a small number of panels available for studies that use longitudinal data; possible correlations between taxes and tobacco control measures and/or antismoking sentiment[6]; the inclusion of covariates that are, by construction, related to age of initiation[7]; and especially for studies using retrospective data a limited number of covariates that are exogenously determined before or when individuals initiated smoking.

Despite the relatively large number of studies identified, the considerable heterogeneity in their methodological approaches and the limitations described above greatly limits the ability to make conclusive statements about the impact of tobacco prices on smoking onset. Additionally, several studies use the same data, so the number of independent estimates is substantially smaller than the number of studies. When considering studies that use a split population duration approach with retrospective data and treat price as a

---

Pierrard, 2004), López Nicolás (López Nicolás, 2002), Madden (Madden, 2007) and Malhotra and Boudarbat (Malhotra and Boudarbat, 2008).

[3] Hence, studies that use a price indicator measured at sub-national level (e.g. state, province) and that experience high levels of within-country migration and/or that use long time series will be disproportionally affected.

[4] A revealing example is DeCicca et al. (2008a) that regress smoking initiation that occurred at any time between 1992 and 2000 on cigarette tax rates in 2000.

[5] Cawley et al. (Cawley et al., 2004) illustrate this issue by regressing, using ordinary least squares, cigarette prices on state and time fixed effects and find a coefficient of determination of 0.99. This can be of consequence. For example, the limited price variation renders the use of state fixed effects (FE) problematic as including FE disallows any of the average unit-to-unit variation in regressors from being used to estimate the parameters of the model. The main result, for example, of DeCicca et al.'s (DeCicca et al., 2002) discrete time hazard models that taxes are not associated with smoking onset are dependent on the inclusion of state FE.

[6] DeCicca et al. (DeCicca et al., 2002) argue that if taxes are correlated with tobacco control measures such as advertising bans and smoke free policies or antismoking sentiment, estimates of the price or tax responsiveness will be inaccurate. This issue is mostly relevant to studies conducted in federated states such as the United-States and Canada where taxes and tobacco control measures may differ substantially across states or provinces.

[7] Several studies (e.g. Arzhenovskiy, 2006; Cawley et al., 2004; Coppejans et al., 2007; Douglas and Hariharan, 1994) that use duration analysis techniques include a measure of age as an explanatory variable, in the duration component of the model. This can problematic as age is, by construction, related to age of initiation.

time-variant covariate, the evidence is fairly limited.[8] Douglas (Douglas, 1998) using data from the United States (1954-1987), Forster & Jones (Forster and Jones, 2001; Forster and Jones, 2003) using data from Great Britain (1920-1984) and Madden (Madden, 2007) using data from a survey of Irish women (1960-1998) find small effect sizes that are not statistically significant; López Nicolás (López Nicolás, 2002) using Spanish data (1957-1990) and Kidd and Hopkins (Kidd and Hopkins, 2004) using data from Australia (1963-1990) find statistically significant but relatively small effect sizes: a 10% increase in prices would delay starting by about one to one and a half months. Studies that use a binary approach (e.g., probit or logit) provide mixed evidence. DeCicca et al. (DeCicca et al., 2002; DeCicca, Kenkel and Mathios, 2005) using longitudinal data from the United States find, for some specifications, large and statistically significant effect sizes, and for other specifications, small and not statistically significant effect sizes. Zhang et al. (Zhang et al., 2006) using Canadian data find large and statistically significant effect sizes (elasticity of initiation with respect to cigarette price is -3.36) while Cawley et al. (Cawley, Markowitz and Tauras, 2006) using US data find large and statistically significant effect sizes, but for boys only (elasticity of initiation with respect to cigarette price is -1.2).

As mentioned earlier, only one study uses data from a low-income country. Laxminarayan and Deolalikar (Laxminarayan and Deolalikar, 2004) using a sample of Vietnamese smokers and non-smokers, study the association between the odds of initiating cigarette smoking and waterpipe tobacco smoking between 1993 and 1998 and changes in the prices of the two tobacco products. They find that changes in the price of cigarettes are significantly and negatively associated with the decision to initiate cigarette smoking (elasticity of cigarette smoking initiation with respect to cigarette price is -1.18). With respect to the impact of waterpipe tobacco prices on waterpipe smoking initiation, they find large effect sizes (elasticity of waterpipe smoking initiation with respect to waterpipe tobacco price is -1.56) that are, however, not statistically significant.

[8] Several studies include price as time-invariant covariate. For example, Douglas and Hariharan (Douglas and Hariharan, 1994) include measures of prices when respondents were 18 years old and the change in price between the age 15 and 18, Grignon and Pierrard (Grignon and Pierrard, 2004) include prices when respondents were 14 and 18 years old, Malhotra and Boudarbat (Malhotra and Boudarbat, 2008) include prices when respondents were 15 years old, and Glied (Glied, 2002) explores the effect of taxes at 14 years old on 'late' initiation (defined as initiation that occurred after age 16). Treating price as a time-variant variable is conceptually more intuitive as the decision whether or not to start smoking is an ongoing decision, made on the basis of current information (Douglas, 1998).

Laxminarayan and Deolalikar also examine the possible effect that changes in prices may have on substitution between tobacco products and on quitting behaviour. They find that changes in the price of cigarettes are significantly and positively associated with the decision to switch from cigarette smoking to pipe smoking. Changes in the price of waterpipe tobacco are found to be significantly and negatively associated with the decision to quit cigarette smoking but not pipe smoking. Changes in the price of cigarettes are not found to have any significant statistical impact on the decision to quit either cigarette or pipe smoking. An important limitation of the Laxminarayan and Deolalikar's study is that the cigarette price data for 1992-1993 and 1997-1998 are not comparable as they are for different brands that are not in the same price category.[9] Additionally, waterpipe tobacco prices were not measured in 1997-1998 and had to be imputed "from household waterpipe tobacco expenditures and price of tobacco sold obtained from households producing tobacco (p. 1194)". The imputation procedure is not reported.

The mixed evidence for high-income populations and the dearth of evidence for populations of low- and middle-income countries uncovered in the review warrant additional analyses of the impact of tobacco prices or taxes on smoking onset in low- and middle-income settings.

3. TOBACCO TAXES, PRICES AND SMOKING IN VIETNAM

In Vietnam the most recent survey conducted in 2006 points to a smoking prevalence of 49.2% among men, but only 1.5% among women (General Statistics Office, 2008c). This is lower than a decade earlier when more than 60% of men and 4% of women smoked tobacco products, although most of this reduction appears to have taken place between 1993 and 1998 (General Statistics Office, 1994, 2000). Of particular importance is the high smoking prevalence in young men (more than 65% of men aged 25 to 45 years smoked in 2006) (General Statistics Office, 2008c). Smoking prevalence is evenly distributed between urban and rural areas, although tobacco users in different areas tend to smoke different products. Waterpipe smoking is more prevalent in rural areas while cigarettes are more popular in urban areas. Among male smokers in 2001–02, 69.1% smoked cigarettes only, 23.2% smoked waterpipe tobacco only, and 7.7% reported using both

[9] Such information is not provided by the authors. An analysis of the VLSS price data uncovered the deficiency.

products. Among urban males, 48.6% smoked cigarettes only and 3.8% smoked waterpipe tobacco only. Among rural males, 35.6% smoked cigarettes only and 16.0% smoked waterpipe tobacco only. Waterpipe tobacco smoking diminished considerably between the early 1990s and the early 2000s (from 19.7% to 13.0%) and is now concentrated among older men (only 2.7% of young men aged 15-24 smoked waterpipe tobacco in 2001-02) (Ministry of Health, 2003).

Despite recent modest excise tax rate increases and the introduction of VAT in 1999, total taxes on cigarettes in Vietnam account for at most 45% of total retail price (Guindon et al., 2009), a rate well below the rates recommended by the World Bank (65%–80% of the final retail price (World Bank, 1999)). Of interest are the recent inconsistent changes in Vietnam's special consumption tax (SCT) on cigarettes. The SCT on filtered cigarettes made with mainly imported materials (i.e., international brands such as *BAT 555)* decreased from 70% (charged on the pre-tax, ex-factory price) in 1994 to 55% in 2007. Rates on domestic filtered cigarettes were kept stable until 2006, when a rate increase schedule was implemented. In January 2008, the SCT on cigarettes was increased to 65% (charged on the pre-tax, ex-factory price) on all three types of cigarettes (filter, mainly made with imported material; filter, mainly made with domestic material; and non-filter). Vietnam's failure to raise tobacco taxes since the early 1990s partly explains its low tobacco product prices. Cigarettes in Vietnam are among the most affordable in the world (Blecher and van Walbeek, 2004, 2009).

4. DATA

4.1. Tobacco, cigarette and waterpipe tobacco prices

Province-level average prices and price indices were obtained from the General Statistics Office of Vietnam (GSO). To calculate the consumer price index (CPI), GSO collects prices for nearly 400 goods and services each month. The number of outlets (e.g., shops or markets) from which prices are collected differs across geographical areas and is based on the population size of each province. GSO provided data for an index of tobacco products (CPI tobacco) and average prices for waterpipe tobacco (per 1 kg) and for two brands of cigarettes (per pack of 20): i) *Vinataba*, a popular brand manufactured by the Vietnam Tobacco Corporation (Vinataba), the government-owned manufacturer; and ii) *555*, a popular British American Tobacco (BAT) premium brand. The data provided are annual averages at the province-level. They cover 30 provinces and the years 1996-2006.

Figure 1 compares trends in a price index of tobacco products and GDP per capita for Vietnam for the years 1995–2006. Figure 2 presents trends in the prices of *Vinataba* and *BAT 555* cigarettes and waterpipe tobacco (per 100 gram) for the period 1996–2006. All series have been adjusted for inflation (using province-level Consumer Price Index (CPI) all-items indices). In real terms, the prices of tobacco products in Vietnam did not increase between 1996 and 2006. On average, they declined by about 5% over that decade. This stands out against trends in real income in Vietnam. Figure 3 presents cigarette price data for 30 provinces (out of 64 provinces). These data show considerable differences across provinces. For example, prices per pack of 555 cigarettes vary from VND 15 000 in Thanh Hoa province to more than VND 19 000 in the provinces of Dong Nai, Ho Chi Minh and Binh Phuoc.

Figure 1. CPI tobacco and GDP per capita, 1995–2006

Figure 2. Prices of tobacco products, 1996–2006

Figure 3. Prices of Vinataba and BAT 555 cigarettes, by province, 2006

4.2. 2003 Survey Assessment of Vietnamese Youth (SAVY)

In 2003, the Vietnamese Government (Ministry of Health and the General Statistics Office) in collaboration with the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) conducted the Survey Assessment of Vietnamese Youth (SAVY) with objectives to assess education, employment, health and reproductive behaviour and other development issues of adolescent and young people, including HIV/AIDS, substance use, injuries and violence. Using the 2002 Vietnam Living Standards Survey as a sample frame, the SAVY sample was drawn using a multi-staged and stratified design. In the first stage, 42 of Vietnam's 61 provinces[10] were selected. In the second stage, 446 of 1643 enumeration areas (EAs) were selected. These EAs contained 8920 households (40 140 individuals). All youths aged 14-24 were targeted to participate in SAVY. Of 9989 youths, 7584 (3753 men and 3831 women) were successfully interviewed for a

[10] In 2003, Vietnam had 61 provinces. Vietnam now has 64 provinces.

response rate of 75.9%. All interviews, conducted between October 2003 and January 2004, were administered face-to-face (in public and communal place without the presence of parents). Young individuals living in special arrangements such as barracks, re-education centers, social protection centers, factories and dormitories' were excluded. SAVY is nationally representative as well as representative across rural and urban strata.

Smoking prevalence among young Vietnamese women is extremely low with an ever-smoking prevalence of less than 2% and a current smoking prevalence that is almost zero. Such low smoking participation renders the modeling of smoking onset among women infeasible, so the analysis focuses on males only. SAVY was conducted in 42 of Vietnam's 61 provinces while the price data obtained from GSO covers 30 provinces. SAVY data from 23 provinces can be linked with GSO data. Respondents from provinces without corresponding price data are dropped (1291). Given the limited number of years of price data available (1996-2006), it is assumed that individuals are first exposed to the risk of starting to smoke at age 14. Consequently, individuals older than 14 in 1996 (i.e. older than 21 at interview) and those who reported starting smoking before age 14 are excluded (574 and 75, respectively). The data suggest that an age of smoking onset of 14 is a reasonable assumption as very few Vietnamese youths reported having initiated before the age of 14.[11] An additional four respondents are dropped because of missing or nonsensical data (e.g., age of smoking onset greater than age at time of survey). The final sample comprises 1809 male respondents from 7 of Vietnam's 8 regions and 23 provinces.[12] The total population of the 23 provinces represented 51.5% of the total population of Vietnam in 2003 (General Statistics Office, 2008a).

In addition to price measures that enter models as time varying covariates, I create a time varying dummy indicator of student status. All other variables are time-invariant. I use an asset index approach proposed by Filmer and Pritchett (Filmer and Pritchett, 1999, 2001) as a proxy for family wealth (measured at the

[11] Others impose a similar restriction. Douglas (Douglas, 1998) assume individuals are first exposed to the risk of starting at age 11 and Kidd and Hopkins (Kidd and Hopkins, 2004) and Madden (Madden, 2007) at age 10.

[12] Red River Delta region: Ha Noi, Hai Phong, Ha Tay, Nam Dinh, Thai Binh; North East region: Lang Son, Thai Nguyen, Phu Tho; North Central Coast region: Thanh Hoa, Nghe An, Thua Thien Hue; South Central Coast region: Da Nang, Binh Dinh, Phu Yen; Central Highlands region: Dak Lak, Lam Dong; South East: Ho Chi Minh, Binh Phuoc, Dong Nai, Binh Thuan; Mekong River Delta region: Tien Giang, Kien Giang, Can Tho.

household-level). SAVY includes data on 13 asset indicators that can be grouped into two types: household ownership of consumer durables, with 11 indicators (electric fan; television; radio; video cassette/VCD/DVD player; bicycle; motobike; boat; refrigerator; computer; telephone; cell phone); and, characteristics of the household's dwelling with 2 indicators (main power source for lighting; main source of drinking water). The asset index is constructed using the full SAVY sample. That is, this variable is created before any individuals are dropped from the sample. Three dummy indicators of parental, sibling and peer smoking (father, brothers and friends) are constructed. As a measures of social interaction, two dummy variables indicating if youths belong to a group of friends or are members of any mass organizations or clubs are constructed. Other covariates include: urban/rural status, ethnicity, paid work, lived away from home, knows how to read and write, greater than 18 years of age at interview and geographical regions. Definitions and descriptive statistics are provided in Table 1.

4.3. Strengths and limitations

The SAVY data enable this analysis to overcome many of the data limitations of previous studies. Although SAVY is cross-sectional and retrospective, the relatively young age of respondents diminishes the possibility of recall bias. SAVY also allows the inclusion of a large number of covariates that were exogenously determined before or when individuals initiated smoking. For example, SAVY allows the inclusion of a measure of household socio-economic status and a time variant measure of school attendance.

A limitation of SAVY is that it does not differentiate between cigarette and waterpipe smoking. As noted earlier, however, only 2.7% of young men aged 15-24 smoked waterpipe tobacco in 2001-02 so such limitation is unlikely to be of consequence. I nevertheless explore the importance of this limitation using data from the Global Youth Tobacco Survey (GYTS) conducted in 2003, which includes information on waterpipe tobacco use. GYTS also allows the use of two distinct measures of smoking onset: age of first experimentation and age of daily smoking initiation. GYTS's weakness lies with its limited geographical coverage and that it is school-based, although secondary school enrolment is relatively high in Vietnam (75% in boys) (Asian Development Bank, 2006).

The GSO price dataset allows me to take advantage of price variations across time and across a relatively large number of provinces. My primary price measure is a weighted price index of tobacco products at province-level. I am also able to investigate differential effects of three specific price measures: waterpipe tobacco and two popular cigarette brands, *Vinataba* and *BAT 555*. The characteristics of SAVY and GSO price data minimizes the possibility of price-matching errors. Because SAVY was conducted at the end of the calendar year, the measure of smoking onset more or less mirrors the calendar year and can be matched accurately with the annual price data. As pointed out earlier when using retrospective data, current location may not match location at time of decision. This type of price-matching error is unlikely to be of consequence in this setting: there was limited out-province migration in Vietnam during the 1990s; between 1994 and 1999 inter-provincial migration was less than 3% (Anh, Tacoli and Thanh, 2003).

5. METHODS

There are two main methods for modeling the determinants of youth smoking onset (Forster and Jones, 2001). The first treats the decision to initiate smoking as a binary event within a discrete choice framework. The second, duration/survival analysis allows one to investigate both whether and, if so, when an event occurs. Duration analysis allows one to include incomplete information from individuals who may not have experienced the event of interest (e.g., smoking onset) before the end of the data collection. Observations from such individuals are (right) censored and excluding them distorts the distribution of event duration (Singer and Willett, 1993). Duration models are specifically designed to take the duration process into account.

Because a large proportion of individuals never start smoking, the assumption inherent in standard survival (or duration) models that each individual will eventually fail appears very restrictive (i.e., the *P*(eventual failure) > 0 for all individuals). I follow Douglas and Hariharan (Douglas and Hariharan, 1994), Douglas (Douglas, 1998) and Forster and Jones (Forster and Jones, 2001) and use a split population duration model. In the split population survival model, the survival process applies only to individuals who are predicted to eventually begin smoking. The split population model weights the likelihood of each observation by using the estimated probability that the individual will ever start smoking (Douglas and Hariharan, 1994).

Following the notation of Schmidt and Witte (Schmidt and Witte, 1984; Schmidt and Witte, 1989), let F be an *unobservable* variable indicating whether an individual i would or would not eventually start smoking (i.e. fail). Formally:

$$P(\text{eventually fail}) = P(F=1) = \delta$$

$$P(\text{never fail}) = P(F=0) = 1-\delta$$

Let $g(t \mid F=1)$ and $G(t \mid F=1)$ be the conditional density of survival times and its corresponding cumulative distribution function for individuals who eventually fail. Let T be the length of the follow-up period (i.e. T_i indicates censoring time). Let R be an *observable* indicator so that R_i=1 if there is failure by time T and R_i=0 if not. For the starters (R_i=1), the unconditional density is:

$$\begin{aligned} P(R=1) &= P(F=1)P(t<T \mid F=1)g(t \mid t<T, F=1) \\ &= P(F=1)g(t \mid F=1) \\ &= \delta g(t \mid F=1) \end{aligned}$$

For the nonstarters (R_i=0), the unconditional density is:

$$\begin{aligned} P(R=0) &= P(F=0) + P(F=1)P(t>T \mid F=1) \\ &= (1-\delta) + \delta G(t \mid F=1) \end{aligned}$$

Combining the two unconditional densities yields the following likelihood function:

$$L = \prod_{i=1}^{N} \delta_i g(t \mid F=1)^{R_i} \left(1-\delta_i + \delta_i G(t \mid F=1)\right)^{1-R_i}$$

The contribution of individual i to the log-likelihood function is:

$$\ln L = R_i\left(\ln \delta_i + \ln g(t \mid F=1)\right) + (1-R_i)\ln\left(1-\delta_i + \delta_i G(t \mid F=1)\right)$$

The probability δ_i is typically modeled as a logit or a probit but can also be modeled as a complementary log-log (cloglog). Unlike probit or logit for which the response curve is symmetric about δ_i=0.5, the cloglog model has a response curve that is asymmetric (Box-Steffensmeier and Jones, 2004). Formally (Box-Steffensmeier and Jones, 2004):

$$\log it : \delta_i = \frac{e^{\alpha' \mathbf{z}_i}}{\left(1 + e^{\alpha' \mathbf{z}_i}\right)}$$

$$probit : \delta_i = \Phi\left(\alpha' \mathbf{z}_i\right)$$

$$c \log\log : \delta_i = 1 - e^{\left(-e^{\alpha' \mathbf{z}_i}\right)}$$

where $\mathbf{z_i}$ is a vector of time invariant covariates, Φ is the cumulative density function for the standard normal distribution, and α is a parameter vector. When $\delta_i = 1$ for all individuals, the split population duration model reduces to a standard duration model.

Both theory and data indicate that the hazard rate first increases and then decreases, which rules out exponential and Weibull duration models. Log-logistic and lognormal models are typically used. Formally, the probability density function $g(t \mid F = 1)$ and the survival function $G(t \mid F = 1)$ for the log-logistic distributions are (Box-Steffensmeier and Jones, 2004):

$$g(t \mid F = 1) = \lambda p(\lambda t)^{\rho - 1} \Big/ \left(1 + (\lambda t)^{\rho}\right)^2$$

$$G(t \mid F = 1) = 1 \Big/ 1 + (\lambda t)^{\rho}$$

where ρ is a shape parameter, $\lambda = e^{-\beta' \mathbf{x}}$, $\mathbf{x_i}$ is a vector of time invariant and time variant covariates, and β is a parameter vector.

Results are presented for the specification that uses a cloglog for the participation component and log-logistic for the duration component of the model. All models are estimated using Stata/SE 10.1 for Macintosh with maximum likelihood statistical codes (method `lf`) adapted from Forster and Jones (Forster and Jones, 2000). To improve convergence, in addition to Stata's modified Newton-Raphson (NR) algorithm, the Berndt-Hall-Hall-Hausman (BHHH) algorithm is used. Each algorithm is used, in turn, for five iterations.

To ensure models are correctly specified and do not yield biased estimates, I perform several diagnostic checks; to illustrate the importance of distributional assumptions and appropriate specifications, I perform extensive sensitivity analyses. I estimate log-logistic frailty models that account for unobserved heterogeneity. Frailty models introduce a random parameter that accounts for the variability due to unobserved individual-level (or group-level) factors that are otherwise unaccounted for by other covariates. Shared- (or group-) frailty assumes that similar observations share the same frailty (i.e., the same unobserved heterogeneity). The frailty is typically assumed to follow a gamma or inverse-Gaussian distribution with mean 1 and variance theta (θ) (Gutierrez, 2002). Adding shared frailty is analogous to adding a random effect to a linear regression to account for correlation between groups of individuals (Kleinbaum and Klein, 2005). I hypothesize that individuals from different provinces, urban/rural strata and who have different levels of wealth (bottom two wealth quintiles vs. top three wealth quintiles) may have different frailties.

Cawley et al. (Cawley et al., 2004), DeCicca et al. (DeCicca et al., 2002; DeCicca et al., 2008b) and Tauras et al. (Tauras, O'Malley and Johnston, 2001) estimate discrete-time hazard models. Unlike parametric duration models that explicitly account for duration dependence by specifying a known distribution, duration dependency is not directly accounted for in discrete-time hazard models. Ignoring duration dependency in the baseline hazard produces a model that is more or less equivalent to an exponential model (i.e., the hazard probability is flat with respect to time) (Box-Steffensmeier and Jones, 2004). I estimate discrete-time hazard models using different duration specifications and, in order to account for possible unobserved heterogeneity, I also estimate discrete-time gamma frailty models and discrete-time split population models.[13] Finally, I estimate discrete-time hazard models (split and non-split) with province-FE although including FE disallows any of the average unit-to-unit variation in regressors from being used to estimate the parameters of the model.[14] In this analysis, including FE amounts to examining only if intra-province changes in smoking onset are associated with intra-province changes in prices.

[13] I use Stata's `pgmhaz8` and `spsurv` routine developed by Stephen Jenkins. `spsurv` does not allow the inclusion of covariates in the participation component of the model.
[14] I also attempt to estimate split population log-logistic duration models with cloglog, logit and probit links. These models, however, do not converge.

6. RESULTS

Descriptive statistics are presented in Table 1. Most young men in the sample were students at age 14 (0.893), were of Kinh ethnicity (0.895), knew how to read and write (0.962), had a group of friends with whom they kept company (0.926) and had friends or a father who smoked (0.657 and 0.599, respectively); about half had ever worked to earn money (0.455) and were a member of any mass organizations or clubs (0.451) while relatively few had brothers who smoked (0.177) or had ever lived away from home continuously for more than one month (0.227).

Table 1. Descriptive statistics.

Figure 4 plots Kaplan and Meier (KM) product-limit survivor and hazard functions assuming that individuals are exposed to the risk of starting to smoke at age 0. It shows clearly the differences between young men and women. Young women's survival and hazard rates are nearly flat. Figure 4 also shows clearly that the hazard rate among young men is non-monotonic which rules out distributions such as Weibull and exponential. Figure 5 contrasts men's survivor functions for the full sample and smokers only. The survivor function for the full sample reaches a limit at around 0.25. The survivor function for the sub-sample of starters reaches zero when starters are about 25 years of age. Such result provides support to the decision to use a split population approach.

Figure 4. Kaplan-Meier survivor and hazard functions for starting –men and women

Figure 5. Kaplan-Meier survivor functions for starting –men (full sample and smokers only)

Table 2 presents the result of the split population duration models. Model 1 presents the results when CPI tobacco is used as the price measure while model 2 uses all three tobacco product prices (*Vinataba* and *BAT 555* cigarettes, and waterpipe tobacco). Models 3-5 present the results when each tobacco product price measures are included individually.

Table 2. Split population log-logistic–cloglog results for starting smoking

For each model, the first column reports parameter estimates for the duration component. These estimates can be interpreted as the effect of a covariate on the age of starting. That is, the estimates of β from the estimate of the duration function $G(t \mid F = 1)$. A positive coefficient indicates that higher values of an explanatory variable delays smoking onset. As price measures are expressed in natural logarithm, their coefficients can be interpreted as elasticities (Forster and Jones, 2000).[15] The second column reports parameter estimates for the participation component of the model. That is, the estimate of α from the probability of smoking δ_i. Antilogged coefficient from the participation component can be interpreted as hazard ratios (unlike antilogged logit coefficients that are odds ratios). Although estimated simultaneously, the effects of the covariates are not constrained to have the same magnitude or direction in both parts of the model.

6.1. Smoking onset and tobacco prices

Model 1 (using CPI tobacco as the price measure) suggests that tobacco prices in Vietnam have a statistically significant and fairly substantial effect on the age of starting (Table 2). The elasticity implied by the tobacco index is 1.33. At the mean starting age (17 years), doubling tobacco prices is expected to delay smoking onset by nearly four years. The associations between the prices of *Vinataba* and *BAT 555* cigarettes and the age of smoking onset are statistically significant and relatively large, 0.52 and 0.99, respectively (model 2). When introduced individually, the prices of *Vinataba* and *BAT 555* cigarettes are also statistically significant and relatively large, 0.44 and 1.10, respectively (models 3-4). When introduced with the prices of cigarettes (*Vinataba* and *BAT 555)*, prices of waterpipe tobacco are statistically significantly and negatively associated with duration (models 2).

6.2. Control variables

[15] Note that elasticities are likely not comparable across studies as different ‘time origin’ are used. For example, Forster and Jones (2001) and López Nicolás (2002) assume individuals are first exposed to the risk of starting at age 0, Kidd and Hopkins (2004) and Madden (2007) at age 10 and Douglas (1998) at age 11.

Focusing first on the duration component of the models, being a student is statistically significantly and negatively associated with age of smoking onset. Model 1 suggests that being a student hasten smoking onset by 14.8% compared to individuals who are no longer in school (Table 2.). This association is robust across all five models (parameter estimates range from -0.16 to -0.14). Wealth (as proxied by an asset index) is statistically significantly associated with duration (the wealth dummies are jointly statistically significant in all models). The parameter estimates suggest the effect of wealth on age of onset is non-linear –there is evidence of an inverted-U relationship. Compared to those in the bottom quintile, individuals in middle wealth categories tend to initiate smoking earlier. Individuals who are members of a mass organizations or clubs tend to start later. Model 1 suggests that individuals who are members of a mass organizations or clubs and those who have ever worked to earn money and who start smoking delay their smoking onset by 14.7% and 25.7%, respectively. Counter-intuitively, having friends that smoke appears to delay smoking onset.

Focusing now on the participation component of the model, peer influences (having friends and brothers that smoke) appear to influence participation substantially. For example, youths who have friends that smoke have a hazard that is nearly 15 times higher than youths who do not have friends that smoke. Wealth is statistically significantly associated with participation (the wealth dummies are jointly statistically significant in all models). Individuals who live in urban areas, who have ever worked to earn money and who know how to read and write are more likely to initiate while those who were born before 1986 are less likely to initiate smoking.

Region dummies are jointly statistically significantly associated with both duration and participation in all models. The shape parameter (ρ) is less than 1 in all models implying that the hazard rate rises first, before declining.

6.3. Robustness: sensitivity analyses and diagnostic checks

Table 3 presents the results of the log-logistic frailty duration models. Model 1 presents the results of log-logistic duration models without frailty. Model 2 introduces individual frailty while models 3-5 introduce shared frailty. The presence of individual frailty and shared frailty across provinces (evaluated using likelihood ratio (LR) tests) is categorically rejected. LR tests, however, suggest that individuals from different urban/rural strata (when CPI tobacco is used) and of differing wealth (when product specific prices are used) may have different frailties. This suggests that unobserved heterogeneity likely impacts the log-logistic duration models. The results suggest tobacco prices (measured by CPI tobacco) are statistically significantly associated with age of smoking onset. The association between the prices of *BAT 555* cigarettes and age of initiation is relatively large and statistically significant. Finally, the log-logistic frailty duration models suggest little to no association between the prices of *Vinataba* cigarettes and waterpipe tobacco and age of smoking onset.

Table 3. Log-logistic duration results for starting smoking

Tables 4A and 4B presents the results of the discrete-time cloglog duration models. Model 1 ignores time dependency. Model 2, 4 and 6 use a natural logarithmic transformation of time. Models 3, 5 and 7 use a flexible approach and account for duration dependency by including temporal dummies. For covariates measured in natural logarithmic such as price, coefficients represent the elasticity of the hazard with respect to a regressor. The results show clearly the importance of incorporating duration in discrete-time models. Ignoring duration dependency (model 1 in Tables 4A and 4B) yields nonsensical results: coefficients are positive and at times, statistically significant, suggesting, for example, that higher prices of *Vinataba* cigarettes are associated with increased hazard of smoking onset. All other models suggest a large, negative and statistically significant association between prices of tobacco products (as measured by CPI tobacco) and the hazard of smoking onset. Similarly, the results suggest a large, negative and statistically significant association between prices of *BAT 555* cigarettes and the hazard of smoking onset. The results, however, suggest no association between the prices of *Vinataba* cigarettes and waterpipe tobacco and the hazard of smoking onset.

Tables 4A and 4B. Discrete-time cloglog duration results for starting smoking

Introducing province-FEs to the discrete-time duration models does not alter the significant association between prices and smoking onset. All models estimated suggest statistically significant and even stronger association between prices and smoking onset: effect sizes are substantially larger.

To ensure that my results are not sensitive to my choice of link function (i.e., cloglog), covariates and my assumption that individuals are first exposed to the risk of starting to smoke at age 14, I perform sensitivity analyses by re-estimating split population log-logistic models using the following specifications. First, I use a logit and a probit link function. Second, I assume that individuals are first exposed to the risk of starting to smoke at age 15 and 16. Third, I explore the effect of including additional covariates (e.g., parents' highest educational attainment, tobacco control policy indicators). The results are robust to the aforementioned alternative specifications. Lastly, to ensure against programming coding errors, I use Stata's `lncure` routine developed by Mario Cleves, which estimates a split population model with lognormal distribution and logit link.[16]

As indicated earlier, GYTS allows one to explore the impact of prices on the age at which individuals first experimented with smoking and the age at which individuals first became daily smokers. GYTS also allows to include as control an indicator of waterpipe tobacco use. Elasticity estimates using GYTS data are qualitatively similar, albeit slightly smaller than the estimates obtained when the SAVY dataset is used.

Figures 6 and 7 plot the cumulative Cox-Snell residuals for the observed failures in order to assess if models are correctly fitted. A properly fitted model should yield cumulative Cox-Snell residuals that resemble a sample from an exponential distribution. A plot of the cumulative hazard function for the Cox-Snell residuals should lie on a straight line from the origin with slope equal to one (Collett, 2003). The log-logistic models using the full sample of smokers indicate serious misspecification: deviations from the 45° are substantial. Such deviations are less serious, although non-negligible, in the split population models that use smokers

[16] One disadvantage of this approach is that it is not possible include covariates in the participation component of the model. See http://www.stata.com/users/mcleves/lncure/

only. Some deviations in the right-hand tail are expected because of the reduced sample due to prior failures and censoring (Cleves, Gould and Gutierrez, 2004).[17] These results provide support in favor of using a split population approach.

Figure 6. Cumulative hazard plot of the Cox-Snell residuals –tobacco price index.

Figure 7. Cumulative hazard plot of the Cox-Snell residuals –Vinataba, BAT 555 and waterpipe tobacco prices.

7. DISCUSSION

Results of the duration analyses suggest that tobacco prices in Vietnam have a statistically significant and fairly substantial effect on the age of starting smoking. Increases in average tobacco prices (measured by an index of tobacco prices) and in the prices of two popular brands (*Vinataba* and *BAT 555*) are found to delay smoking onset. Of particular interest, given their favorable tax treatment since the late 1990s, is the finding that Vietnamese youth are more sensitive to changes in prices of an international brand. Vietnam's special consumption tax (SCT) on international brands decreased from 70% in 1993 to 55% in 2007. During the same period, the SCT on local brands such as *Vinataba* increased from 52% to 55%. Results from the split population model suggest that increases in the prices of waterpipe tobacco may hasten smoking onset. The likely interpretation is that increases in the prices of waterpipe tobacco may hasten cigarette smoking onset among potential waterpipe tobacco users. This finding suggests that changes in the taxation of cigarettes should not be decided in isolation.[18]

Wealth (as proxied by an asset index) is significantly associated with smoking onset. Compared to those in the bottom quintile, individuals in higher wealth categories –and especially individuals in middle wealth categories– tend to initiate smoking earlier. These results suggest that tobacco products are normal goods in Vietnam. The effect of having friends or brothers who smoke suggests peer influences play an important role

[17] It is worth noting that examining plots based on the Cox-Snell residuals to assess model fit has limitations. Collett (2003) argues that in practice a straight line plot is often obtained even when the model fitted is known to be incorrect.
[18] Waterpipe tobacco is not subject to taxation in Vietnam. Much of the tobacco consumed in this form is home-produced, which renders tax collection difficult.

in the decision to start smoking. Unfortunately, the dataset does not allow to explore asymmetric peer effects (Harris and Gonzalez Lopez-Valcarcel, 2008).

This study's contributions to the smoking literature are three-fold. To my knowledge, this is the first application of duration analysis to study the impact of tobacco prices on the decision to initiate smoking in a low-income country. Only two studies have attempted to examine the association between tobacco prices and smoking onset in low- and middle-income countries (Arzhenovskiy, 2005; Laxminarayan and Deolalikar, 2004). This is important because of the limited generalizability of studies conducted in high-income economies to low-income settings, because populations in most low- and middle countries are very young and because several low- and middle-income countries are already experiencing a rise in non-communicable diseases associated with tobacco use (Beaglehole and Yach, 2003).

Second, the data utilized are less subject to many data and measurement limitations found in existing studies: (1) Linking data from GSO and SAVY minimizes the possibility of recall bias and price-matching errors and allows me to take advantage of price variations across time and across a large number of provinces; (2) SAVY is a rich dataset that allows the inclusion of a number of covariates that are exogenously determined before or when individuals initiated smoking; (3) in addition to using a price index of tobacco products, I use three additional price measures and I also use a second dataset to further explore the association between tobacco prices and smoking onset. Consequently, on the whole, my estimates are less prone to measurement errors and omitted variable and endogeneity biases.

Third, I apply methods –split-population duration analyses – that are better suited to measure the impact of tobacco prices on the onset of smoking. I also perform extensive sensitivity analyses and diagnostic checks. The finding that prices of tobacco products have a significant impact on smoking onset is robust across all specifications.

There are, however, some limitations that merit discussion. Vietnam's socioeconomic and cultural characteristics limit the generalizability of my findings. Smoking prevalence among Vietnamese women is extremely low. As such, the findings cannot be assumed to extend to young women. Vietnam has a fast growing economy with a relatively high level of human development but nevertheless remains a poor country. GDP per capita was just about 17.1 million dong (1050 USD) in 2008 (International Monetary Fund, 2008). As such, the study's generalizability may not extend to higher income economies. The study's main findings, however, are likely generalizable to several low- and middle-income countries such as Cambodia, China and Laos. As is the case in Vietnam, smoking prevalence in men in these countries is in excess of 50%. And as is the case in Vietnam, smoking rates in women in Cambodia, China and Laos are orders of magnitude lower than in men (6%, 3% and 16%, respectively) (China Ministry of Health, 2006; National Institute of Statistics Cambodia and Ministry of Planning, 2006; World Health Organization, 2004).

Although the relatively young age of respondents diminishes the possibility of recall bias, the possibility that recall bias impacts the finding remains. Glied (Glied, 2002) reports evidence of recall bias among a young cohort in the United States. For example, she reports that 10% of those who reported in 1984 they were currently smoking at least one cigarette daily, recalled in 1992 that they had started smoking only after 1984 or had never started.

Using a tobacco price index has the benefit of reflecting changes in the prices of various brands of cigarettes and in Vietnam's case, the prices of waterpipe tobacco. However, if individuals switch across price categories (e.g., from low to high priced tobacco products or conversely, from high to low priced tobacco products), the tobacco index will likely reflect changes in Laspeyre's weights. Given Vietnam's rapid income growth, it is unlikely that individuals have switched to cheaper cigarette brands or tobacco products. The possibility, however, that individuals have switched to higher priced brands or products exists. If such price switching is pervasive, the price index of tobacco products may not be a good indicator of changes in tobacco prices in Vietnam. The similarities between the results obtained using prices of specific cigarette brands and the tobacco index suggest such limitation is unlikely to be of consequence.

Findings from this study provide additional evidence of the effectiveness of tobacco prices at reducing tobacco use initiation. Vietnam ratified the Framework Convention on Tobacco Control in December 2004, which demonstrates a commitment to tobacco control.[19] However, Vietnam's recent price and tax policies have not been in line with its commitment under the FCTC: inflation-adjusted prices of tobacco products in Vietnam declined by about 5% between 1995 and 2006.[20] These results suggest that living up to its commitments under the FCTC would have a measurable effect on smoking onset in Vietnam.

[19] http://www.who.int/fctc/signatories_parties/en/index.html

[20] Article 6 of the FCTC states (World Health Organization, 2005 pp. 7-8):
1. The Parties recognize that price and tax measures are an effective and important means of reducing tobacco consumption by various segments of the population, in particular young persons.
2. Without prejudice to the sovereign right of the Parties to determine and establish their taxation policies, each Party should take account of its national health objectives concerning tobacco control and adopt or maintain, as appropriate, measures which may include:
(a) implementing tax policies and, where appropriate, price policies, on tobacco products so as to contribute to the health objectives aimed at reducing tobacco consumption;

REFERENCES

Anh, D. N., C. Tacoli, and H. X. Thanh. 2003. "Migration in Vietnam: A review of information on current trends and patterns, and their policy implications." *Migration, Development, and Pro-Poor Policy Choices in Asia.* Dhaka: Refugee and Migratory Movements Research Unit, Bangladesh and the Department for International Development, UK.

Arzhenovskiy, S. 2006. "Socioeconomic determinants of smoking in Russia [in Russian]." *Quantile Journal* 1: 81-100.

Asian Development Bank. 2006. *Key Indicators 2006: Measuring Policy Effectiveness in Health and Education*: Asian Development Bank.

Beaglehole, R. and D. Yach. 2003. "Globalisation and the prevention and control of non-communicable disease: the neglected chronic diseases of adults." *Lancet* 362(9387): 903-8.

Becker, G. S., M. Grossman, and K. M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *The American Economic Review* 81(2): 237-41.

Blecher, E. H. and C. P. van Walbeek. 2004. "An international analysis of cigarette affordability." *Tob Control* 13(4): 339-46.

Blecher, E. H. and C. P. van Walbeek. 2009. "Cigarette affordability trends: an update and some methodological comments." *Tob Control* 18(3): 167-75.

Box-Steffensmeier, J. M. and B. S. Jones. 2004. *Event history modeling: a guide for social scientists*. Cambridge: Cambridge University Press.

Breslau, N., N. Fenn, and E. L. Peterson. 1993. "Early smoking initiation and nicotine dependence in a cohort of young adults." *Drug Alcohol Depend* 33(2): 129-37.

Breslau, N. and E. L. Peterson. 1996. "Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences." *Am J Public Health* 86(2): 214-20.

Cameron, S. 1998. "Estimation of the demand for cigarettes: a review of the literature." *Econ Issues* 3(2): 351–72.

Cawley, J., S. Markowitz, and J. A. Tauras. 2004. "Lighting up and slimming down: the effects of body weight and cigarette prices on adolescent smoking initiation." *J Health Econ* 23(2): 293-311.

Cawley, J., S. Markowitz, and J. A. Tauras. 2006. "Obesity, Cigarette Prices, Youth Access Laws, and Adolescent Smoking Initiation." *Eastern Economic Journal* 32(1): 149-70.

Chaloupka, F. J., T.-w. Hu, K. E. Warner, R. Jacobs, and A. Yurekli. 2000. "The taxation of tobacco products." In *Tobacco control policies in developing countries*, edited by P. Jha and F. J. Chaloupka, pp. 237-72. New York: Oxford University Press.

Chaloupka, F. J. and K. E. Warner. 2000. "The Economics of Smoking." In *Handbook of Health Economics*, edited by A. J. Culyer and J. P. Newhouse, pp. 1539-627. Amsterdam: Elsevier Science, North-Holland.

China Ministry of Health. 2006. "China Smoking and Health Report 2006. ." Beijing: Ministry of Health.

Cleves, M. A., W. W. Gould, and R. G. Gutierrez. 2004. *An introduction to survival analysis using Stata*. College Station, Tex.: Stata Press.

Collett, D. 2003. *Modelling survival data in medical research*. Boca Raton, Fla.: Chapman & Hall/CRC.

Coppejans, M., D. Gilleskie, H. Sieg, and K. Strumpf. 2007. "Consumer Demand under Price Uncertainty: Empirical Evidence from the Market for Cigarettes." *Review of Economics and Statistics* 89(3): 510-21.

DeCicca, P., D. Kenkel, and A. Mathios. 2002. "Putting Out the Fires: Will Higher Taxes Reduce the Onset of Youth Smoking?" *Journal of Political Economy* 110(1): 144-69.

DeCicca, P., D. Kenkel, and A. Mathios. 2005. "The Fires Are Not Out Yet: Higher Taxes and Young Adult Smoking." In *Substance Use: Individual Behaviour, Social Interactions, Markets and Politics*, edited by B. Lindgren and M. Grossman, pp. 293-312. Advances in Health Economics and Health Services Research series, vol. 16. Amsterdam and San Diego: JAI Elsevier.

DeCicca, P., D. Kenkel, and A. Mathios. 2008a. "Cigarette taxes and the transition from youth to adult smoking: Smoking initiation, cessation, and participation." *J Health Econ* 27(4): 904-17.

DeCicca, P., D. Kenkel, A. Mathios, Y. J. Shin, and J. Y. Lim. 2008b. "Youth smoking, cigarette prices, and anti-smoking sentiment." *Health Econ* 17(6): 733-49.

Doll, R. 1998. "Uncovering the effects of smoking: historical perspective." *Stat Methods Med Res* 7(2): 87-117.

Douglas, S. 1998. "The Duration of the Smoking Habit." *Economic Inquiry* 36(1): 49-64.

Douglas, S. and G. Hariharan. 1994. "The hazard of starting smoking: estimates from a split population duration model." *J Health Econ* 13(2): 213-30.

Everett, S. A., C. W. Warren, D. Sharp, L. Kann, C. G. Husten, and L. S. Crossett. 1999. "Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students." *Prev Med* 29(5): 327-33.

Filmer, D. and L. H. Pritchett. 1999. "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries." *Population and Development Review* 25(1): 85-120.

Filmer, D. and L. H. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India." *Demography* 38(1): 115-32.

Forster, M. and A. M. Jones. 2000. "The role of tobacco taxes in starting and quitting smoking: duration analysis of British data." *Centre for Health Economics Discussion Paper Series, Discussion Paper 176.* York: Centre for Health Economics, University of York.

Forster, M. and A. M. Jones. 2001. "The role of tobacco taxes in starting and quitting smoking: Duration analysis of British data." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164: 517-47.

Forster, M. and A. M. Jones. 2003. "Corrigendum: The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 166(3): 441-42.

Gallet, C. A. and J. A. List. 2003. "Cigarette demand: a meta-analysis of elasticities." *Health Econ* 12(10): 821-35.

General Statistics Office. 1994. “Vietnam Living Standard Survey (VLSS), 1992–1993.” Hanoi: Statistical Publishing House.

General Statistics Office. 2000. “Vietnam Living Standard Survey (VLSS), 1997–1998.” Hanoi: Statistical Publishing House.

General Statistics Office. 2008a. *Statistical Handbook of Vietnam, 2007*. Hanoi: General Statistics Office, Statistical Publishing House.

General Statistics Office. 2008b. *The 2007 Population Change and Family Planning Survey: Major Findings*. Hanoi: General Statistics Office, Statistical Publishing House.

General Statistics Office. 2008c. “Vietnam Household Living Standard Survey (VHLSS), 2006.” Hanoi: Statistical Publishing House.

Glied, S. 2002. “Youth tobacco control: reconciling theory and empirical evidence.” *J Health Econ* 21(1): 117-35.

Grignon, M. and B. Pierrard. 2004. “The impact of price on young people's smoking behaviour [in French].” *Journal d'Economie Medicale* 22(3): 119-30.

Grossman, M. and F. J. Chaloupka. 1997. “Cigarette taxes. The straw to break the camel's back.” *Public Health Rep* 112(4): 290-7.

Guindon, G. E., N. Hien, H. Kinh, and E. McGirr. 2009. “Tobacco Taxation in Vietnam.” Paris: International Union Against Tuberculosis and Lung Disease.

Guindon, G. E., A.-M. Perucic, and D. Boisclair. 2003. “Higher tobacco prices and Taxes in South-East Asia: an effective tool to reduce tobacco use, save lives and increase government revenue.” *HNP Discussion Paper. Economics of Tobacco Control Paper No. 11.* Washington: The World Bank.

Gutierrez, R. G. 2002. “Parametric frailty and shared frailty survival models.” *Stata Journal* 2(1): 22-44.

Harris, J. E. and B. Gonzalez Lopez-Valcarcel. 2008. “Asymmetric peer effects in the analysis of cigarette smoking among young people in the United States, 1992-1999.” *J Health Econ* 27(2): 249-64.

International Monetary Fund. 2008. “World Economic Outlook Database.” IMF.

Kidd, M. P. and S. Hopkins. 2004. “The Hazards of Starting and Quitting Smoking: Some Australian Evidence.” *Economic Record* 80(249): 177-92.

Kleinbaum, D. G. and M. Klein. 2005. *Survival analysis: a self-learning text*. New York: Springer.

Laporte, A. 2006. “Price responsiveness of demand for cigarettes: does rationality matter?” *Subst Use Misuse* 41(4): 511-31.

Laxminarayan, R. and A. Deolalikar. 2004. “Tobacco Initiation, Cessation, and Change: Evidence from Vietnam.” *Health Econ* 13(12): 1191-201.

Levy, D. T., S. Bales, N. T. Lam, and L. Nikolayev. 2006. “The role of public policies in reducing smoking and deaths caused by smoking in Vietnam: results from the Vietnam tobacco policy simulation model.” *Soc Sci Med* 62(7): 1819-30.

Lewit, E. M., D. Coate, and M. Grossman. 1981. "The Effects of Government Regulation on Teenage Smoking." *Journal of Law and Economics* 24(3): 545-69.

López Nicolás, Á. 2002. "How important are tobacco prices in the propensity to start and quit smoking? An analysis of smoking histories from the Spanish National Health Survey." *Health Econ* 11(6): 521-35.

Madden, D. 2007. "Tobacco taxes and starting and quitting smoking: does the effect differ by education?" *Applied Economics* 39(5): 613-27.

Malhotra, N. and B. Boudarbat. 2008. "The Hazards of Starting the Cigarette Smoking Habit." *Working paper.*

Mathers, C. D. and D. Loncar. 2006. "Projections of global mortality and burden of disease from 2002 to 2030." *PLoS Med* 3(11): e442.

Ministry of Health. 2003. "Vietnam National Health Survey (VNHS), 2001-02 ". Hanoi: MoH.

National Institute of Statistics Cambodia and Ministry of Planning. 2006. "Cambodia Smoking Behavior Survey, 2004." Phnom Penh: National Institute of Statistics.

Schmidt, P. and A. D. Witte. 1984. *An economic analysis of crime and justice : theory, methods, and applications*. Orlando, Fla.: Academic Press.

Schmidt, P. and A. D. Witte. 1989. "Predicting criminal recidivism using 'split population' survival time models." *Journal of Econometrics* 40(1): 141-59.

Singer, J. D. and J. B. Willett. 1993. "It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics* 18(2): 155-95.

Tauras, J. A. and F. J. Chaloupka. 1999. "Determinants of Smoking Cessation: An Analysis of Young Adult Men and Women." National Bureau of Economic Research Inc NBER Working Papers: 7262.

Tauras, J. A., P. M. O'Malley, and L. D. Johnston. 2001. *Effects of Price and Access Laws on Teenage Smoking Initiation: A National Longitudinal Analysis*: National Bureau of Economic Research, Inc, NBER Working Papers: 8331.

Thomas, S., D. Fayter, K. Misso, D. Ogilvie, M. Petticrew, A. Sowden, M. Whitehead, and G. Worthy. 2008. "Population tobacco control interventions and their effects on social inequalities in smoking: systematic review." *Tob Control* 17(4): 230-7.

US Department of Health and Human Services. 1994. *Preventing tobacco use among young people. A report of the Surgeon General*. Rockville, Maryland: Public Health Service, Centers for Disease Control, Center for Health Promotion and Education, Office in Smoking and Health.

US Department of Health and Human Services. 2004. *The health consequences of smoking: a report of the Surgeon General*. Atlanta: United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

World Bank. 1999. *Curbing the epidemic: governments and the economics of tobacco control*. Washington: The World Bank.

World Health Organization. 2004. “Lao People's Democratic Republic World Health Survey, 2003.” Geneva: WHO.

World Health Organization. 2005. “Framework Convention on Tobacco Control.” Geneva: World Health Organization.

Zhang, B., J. Cohen, R. Ferrence, and J. Rehm. 2006. “The impact of tobacco tax cuts on smoking initiation among Canadian young adults.” *Am J Prev Med* 30(6): 474-9.

Figure 1. CPI tobacco and GDP per capita, 1995–2006

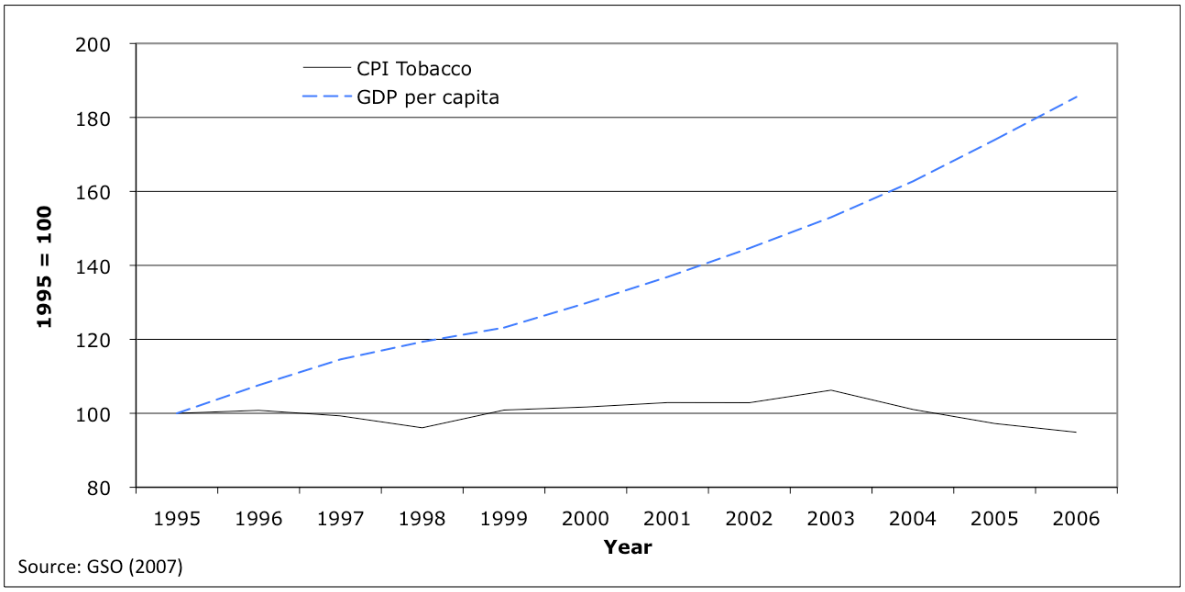

Figure 2. Prices of tobacco products, 1996–2006

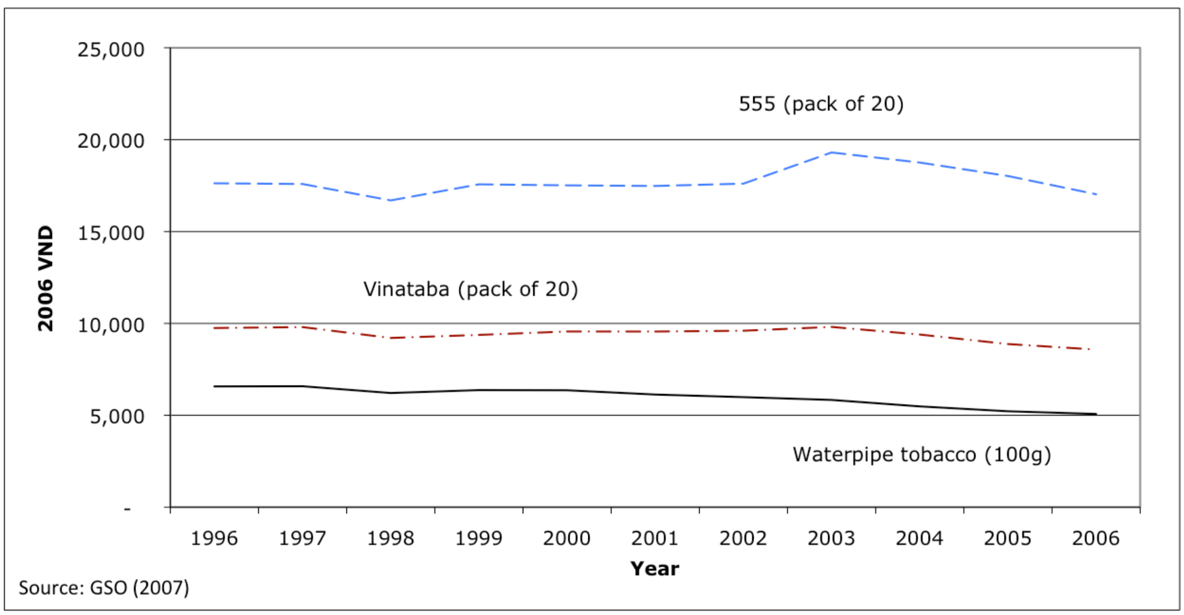

Figure 3. Prices of Vinataba and BAT 555 cigarettes, by province, 2006

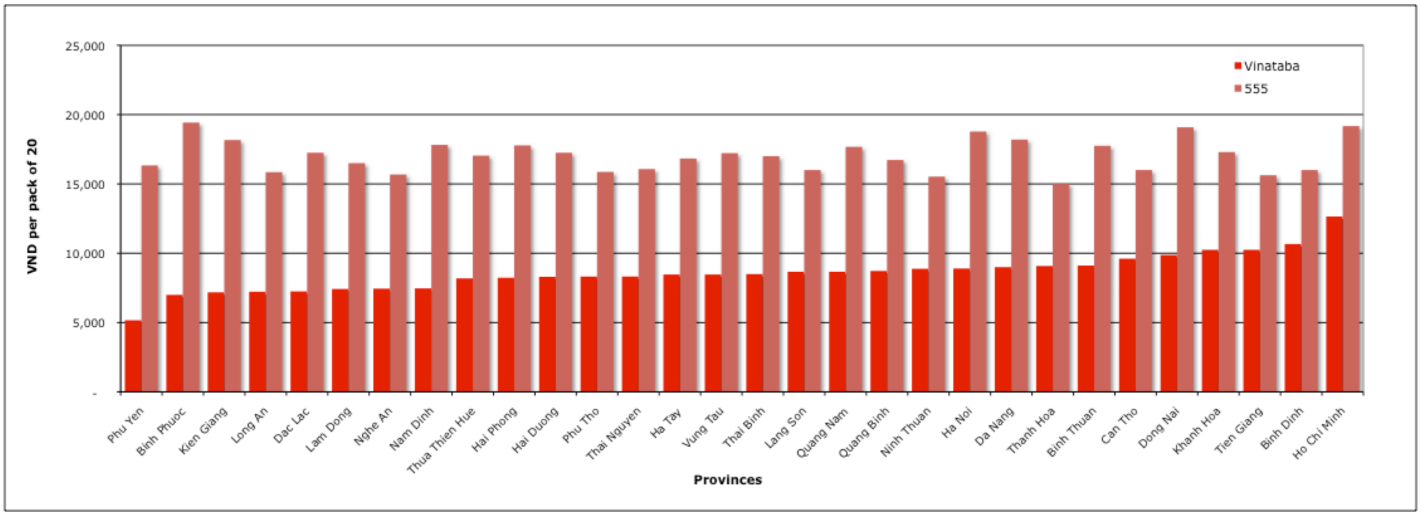

Figure 4. Kaplan-Meier survivor and hazard functions for starting –men and women

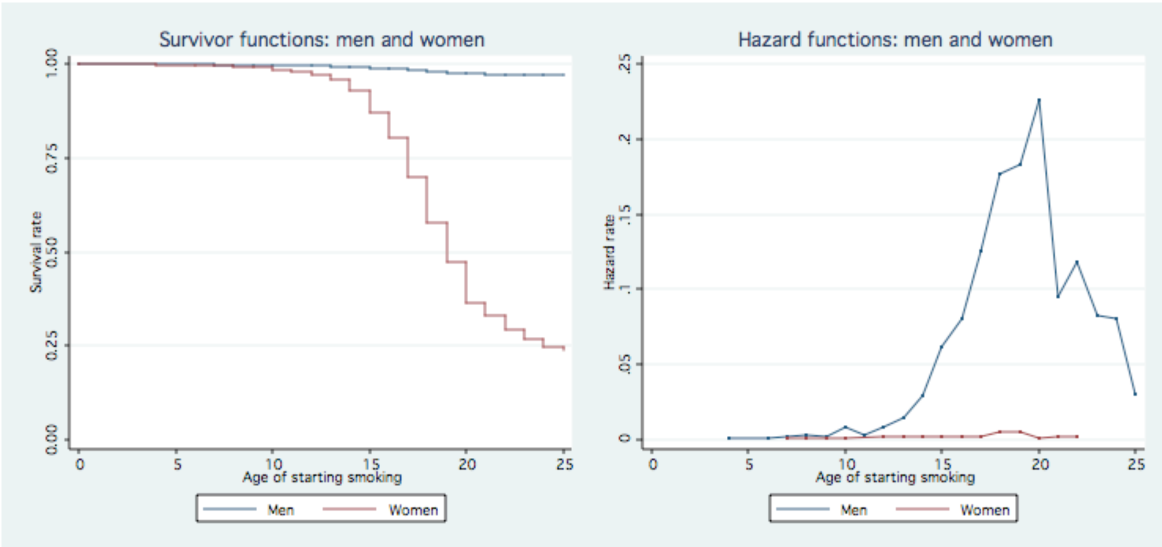

Figure 5. Survivor functions for starting –men (full sample and smokers only)

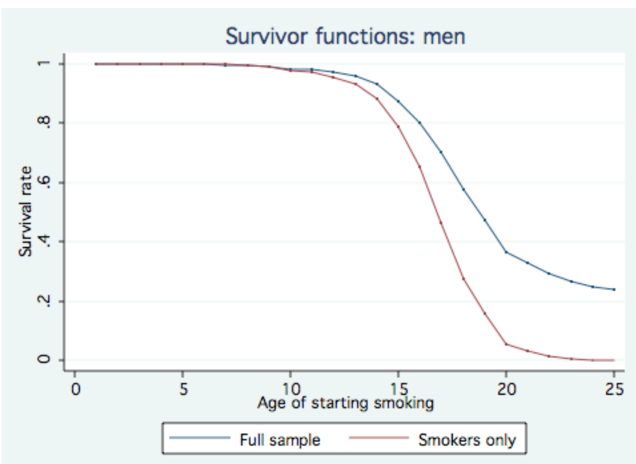

Figure 6. Cumulative hazard plot of the Cox-Snell residuals –tobacco price index.

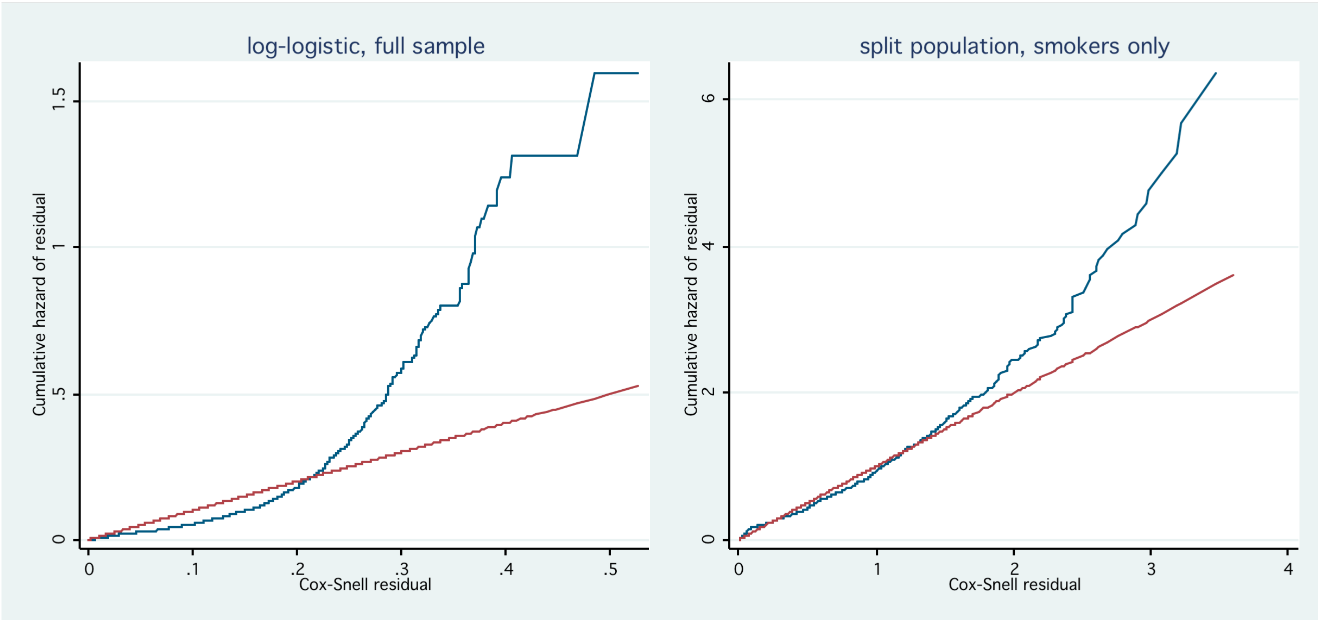

Figure 7. Cumulative hazard plot of the Cox-Snell residuals –*Vinataba, BAT 555* and waterpipe tobacco prices.

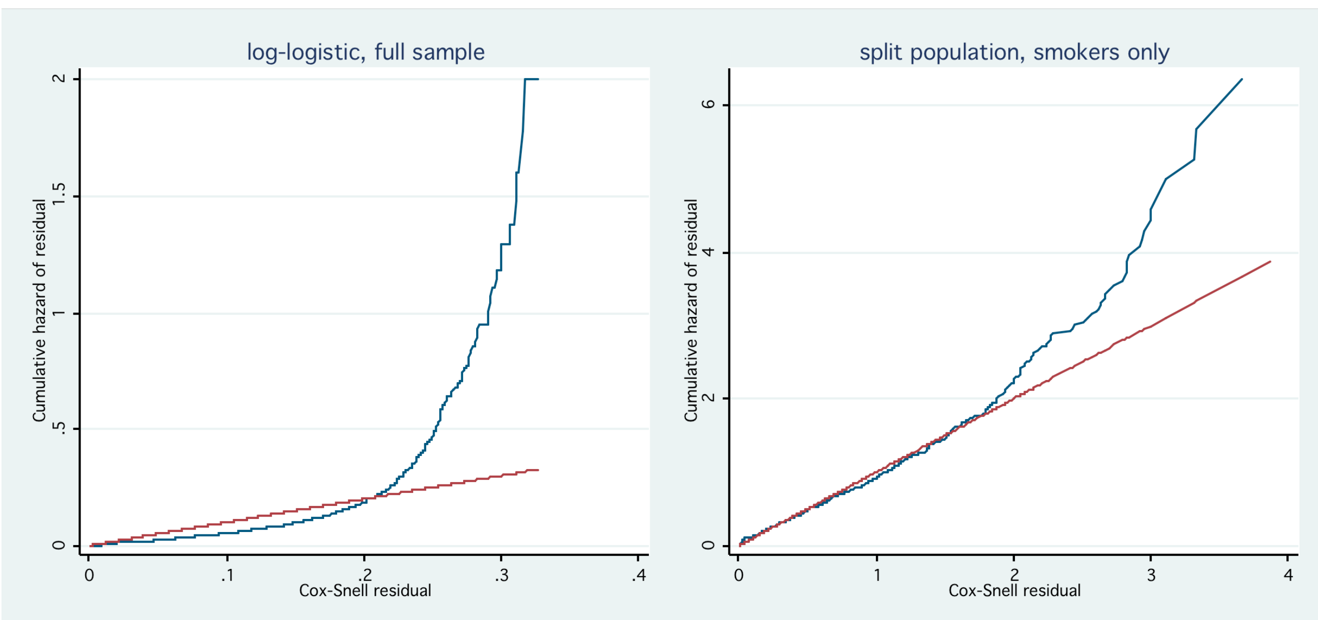

Table 1. Variables definitions and descriptive statistics

Variable	Definition/Question	Mean	SD
Age of starting smoking (smokers only)	How old were you when you first smoked?	17.03	1.63
Student	The variable 'student' is time-varying and is constructed from the following questions: How old were you when you started going to school? Do you currently go to school, college, university? How old were you when you stopped going to school? Student at age 14:	0.893	
Urban	Rural area=0; 1 otherwise (urban area of big cities; urban area of other cities; two (district level))	0.395	
Wealth	Asset index		
Quintile 1 (low)	-Reference category	0.135	
Quintile 2		0.184	
Quintile 3		0.211	
Quintile 4		0.239	
Quintile 5 (high)		0.231	
Smoking behaviour			
Friends	Do you have any close friends who smoke?	0.657	
Father	Does anyone in your family smoke? -Father	0.599	
Brothers	Does anyone in your family smoke? -Brothers	0.177	
Ethnicity -Kinh	What is your ethnicity? Kinh=1; 0 otherwise	0.895	
Lived away	Have you ever lived away from home continuously for more than one month?	0.227	
Read/write	Do you know how to read and write?	0.962	
Paid work	Have you ever worked to earn money?	0.455	
Friends	Do you have a group of friends with whom you often keep company?	0.926	
Clubs	Are you a member of any mass organizations or clubs?	0.451	
Born < 1986	How old are you according to solar calendar?	0.466	
Region			
Red River Delta	-Reference category	0.292	
North East		0.096	
North Central Coast		0.123	
South Central Coast		0.070	
Central Highlands		0.068	
South East		0.258	
Mekong River Delta		0.093	
Prices			
1996			
CPI tobacco	Index; 1995=100	97.71	1.76
Vinataba	Per pack of 20, in 2000 VND	6931	940
BAT 555	Per pack of 20, in 2000 VND	12618	1326
Waterpipe	Per kilogram, in 2000 VND	4344	2049
2003			
CPI tobacco	Index; 1995=100	100.65	7.76
Vinataba	Per pack of 20, in 2000 VND	7092	1317
BAT 555	Per pack of 20, in 2000 VND	14102	2044
Waterpipe	Per 100 g in 2000 VND	3947	1761

Table 2. Split population log-logistic–cloglog results for starting smoking

Models	*(1)* Duration	Participation	*(2)* Duration	Participation	*(3)* Duration	Participation	*(4)* Duration	Participation	*(5)* Duration	Participation
Prices (in logs)										
CPI tobacco	1.328	–	–	–	–	–	–	–	–	–
	(0.251)***									
Vinataba	–	–	0.521	–	0.437	–	–	–	–	–
			(0.151)***		(0.126)***					
BAT 555	–	–	0.994	–	–	–	1.095	–	–	–
			(0.171)***				(0.171)***			
Waterpipe	–	–	-0.120	–	–	–	–	–	-0.001	–
			(0.062)*						(0.052)	
Student	-0.148	–	-0.142	–	-0.155	–	-0.146	–	-0.156	–
	(0.036)***		(0.035)***		(0.036)***		(0.036)***		(0.037)***	
Urban	0.040	0.785	0.070	1.012	0.054	0.943	0.067	0.970	0.052	0.864
	(0.045)	(0.317)**	(0.043)	(0.297)***	(0.048)	(0.314)***	(0.044)	(0.299)***	(0.050)	(0.325)***
Wealth										
Quintile 2	-0.124	-0.428	-0.181	-0.724	-0.185	-0.732	-0.173	-0.777	-0.161	-0.632
	(0.075)*	(0.539)	(0.072)**	(0.469)	(0.083)**	(0.557)	(0.074)**	(0.471)*	(0.081)**	(0.546)
Quintile 3	-0.161	-1.045	-0.231	-1.354	-0.227	-1.324	-0.214	-1.306	-0.214	-1.262
	(0.076)**	(0.525)**	(0.070)***	(0.433)***	(0.076)***	(0.484)***	(0.071)***	(0.441)***	(0.077)***	(0.500)**
Quintile 4	-0.113	0.121	-0.193	-0.274	-0.183	-0.203	-0.181	-0.309	-0.154	-0.148
	(0.073)	(0.516)	(0.068)***	(0.441)	(0.080)**	(0.541)	(0.069)***	(0.446)	(0.081)*	(0.559)
Quintile 5 (high)	-0.009	-0.369	-0.107	-0.982	-0.094	-0.899	-0.101	-1.023	-0.069	-0.870
	(0.086)	(0.622)	(0.084)	(0.534)*	(0.103)	(0.698)	(0.085)	(0.535)*	(0.107)	(0.735)*
Smoking behaviour										
Friends	0.192	2.746	0.202	2.795	0.184	2.723	0.196	2.748	0.177	2.656
	(0.096)**	(0.400)***	(0.092)**	(0.366)***	(0.096)*	(0.381)***	(0.093)**	(0.363)***	(0.097)*	(0.376)***
Father	-0.022	0.275	-0.009	0.336	-0.007	0.368	-0.012	0.315	-0.001	0.392
	(0.039)	(0.237)	(0.037)	(0.218)	(0.039)	(0.229)	(0.038)	(0.221)	(0.041)	(0.227)*
Brother(s)	0.039	0.592	0.036	0.510	0.024	0.533	0.037	0.546	0.028	0.564
	(0.043)	(0.273)**	(0.042)	(0.258)**	(0.045)	(0.284)*	(0.043)	(0.259)**	(0.046)	(0.276)**
Ethnicity -Kinh	-0.052	-0.309	-0.090	-0.202	-0.077	-0.168	-0.096	-0.214	-0.102	-0.248
	(0.070)	(0.381)	(0.072)	(0.379)	(0.075)	(0.394)	(0.073)	(0.377)	(0.074)	(0.400)
Lived away	0.028	0.294	0.048	0.306	0.019	0.262	0.055	0.348	0.023	0.334
	(0.039)	(0.256)	(0.038)	(0.244)	(0.040)	(0.253)	(0.039)	(0.245)	(0.040)	(0.250)
Read/write	0.088	0.955	0.089	0.761	0.056	0.702	0.127	0.880	0.096	0.822
	(0.104)	(0.512)*	(0.107)	(0.502)	(0.110)	(0.528)*	(0.108)	(0.496)*	(0.110)	(0.524)*
Paid work	0.257	2.220	0.265	2.195	0.232	2.110	0.258	2.121	0.237	2.050
	(0.047)***	(0.321)***	(0.046)***	(0.283)***	(0.049)***	(0.306)***	(0.046)***	(0.282)***	(0.049)***	(0.300)***
Friends	-0.108	0.348	0.103	1.351	0.021	1.022	0.116	1.346	0.009	0.922
	(0.101)	(0.566)	(0.121)	(0.474)***	(0.167)	(0.718)	(0.123)	(0.471)***	(0.185)	(0.822)
Clubs	0.147	-0.001	0.165	0.207	0.163	0.151	0.168	0.164	0.159	0.055
	(0.044)***	(0.299)	(0.040)***	(0.257)	(0.044)***	(0.287)	(0.041)***	(0.256)	(0.047)***	(0.298)
Born < 1986	–	-2.484	–	-2.645	–	-2.451	–	-2.569	–	-2.381
		(0.335)***		(0.325)***		(0.325)***		(0.323)***		(0.324)***
Region										
North East	-0.055	-0.479	-0.240	-0.615	-0.134	-0.537	-0.230	-0.591	-0.122	-0.522
	(0.075)	(0.464)	(0.073)***	(0.429)	(0.078)*	(0.467)	(0.074)***	(0.434)	(0.079)	(0.472)
North Central Coast	-0.032	0.217	-0.075	-0.036	-0.089	0.041	-0.044	0.004	-0.072	0.073
	(0.065)	(0.445)	(0.063)	(0.428)	(0.065)	(0.462)	(0.063)	(0.430)	(0.068)	(0.475)
South Central Coast	0.027	-0.549	-0.061	-0.885	-0.089	-0.871	-0.152	-0.945	-0.155	-0.843
	(0.077)	(0.460)	(0.077)	(0.414)**	(0.084)	(0.466)*	(0.076)**	(0.420)**	(0.088)*	(0.504)
Central Highlands	0.061	0.120	-0.211	-0.321	0.007	-0.104	-0.221	-0.361	0.002	-0.083
	(0.094)	(0.808)	(0.091)	(0.516)	(0.103)	(0.700)	(0.092)**	(0.520)	(0.107)	(0.767)
South East	-0.147	-1.535	-0.373	-1.290	-0.292	-1.419	-0.348	-1.289	-0.202	-1.477
	(0.062)**	(0.358)***	(0.068)***	(0.333)***	(0.065)***	(0.362)***	(0.061)***	(0.335)	(0.072)***	(0.362)***
Mekong River Delta	0.180	-0.566	0.010	-0.905	-0.067	-0.890	0.030	-0.909	-0.023	-0.862
	(0.086)**	(0.635)	(0.077)	(0.514)*	(0.088)	(0.621)	(0.076)	(0.508)*	(0.086)	(0.577)
Constant	-4.473	-2.026	-11.178	-2.607	-2.138	-2.285	-8.837	-2.668	1.704	-2.266
	(1.188)***	(0.798)**	(1.833)***	(0.832)***	(1.123)	(0.843)**	(1.651)***	(0.832)***	(0.578)***	(0.846)***
Gamma (ρ)	0.287	–	0.277	–	0.289	–	0.280	–	0.291	–
	(0.010)***		(0.009)***		(0.010)***		(0.009)***		(0.010)***	
Log likelihood	-2098.4		-2086.5		-2106.2		-2092.3		-2112.0	

Note: Standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%. Red River Delta is the reference region.
Number of observations = 6721; Number of subjects = 1809; Number of failures = 582.

Table 3. Log-logistic duration results for starting smoking

Models	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Frailty (distribution)		*Gamma*	*Gamma*	*Gamma*	*Gamma*		*Gamma*	*Gamma*	*Gamma*	*Gamma*
Shared			Province	Urban	Wealth			Province	Urban	Wealth
Prices (in logs)										
CPI tobacco	0.927	0.927	0.927	0.916	0.927	–			–	–
	(0.310)***	(0.310)***	(0.310)***	(0.293)***	(0.310)***					
Vinataba	–	–	–	–	–	0.285	0.285	0.285	0.285	0.293
						(0.177)	(0.177)	(0.177)	(0.177)	(0.164)*
BAT 555	–	–	–	–	–	0.763	0.763	0.763	0.763	0.853
						(0.213)***	(0.213)***	(0.213)***	(0.213)***	(0.193)***
Waterpipe	–	–	–	–	–	-0.069	-0.069	-0.069	-0.069	-0.058
						(0.073)	(0.073)	(0.073)	(0.073)	(0.068)
Student	0.018	0.018	0.018	0.024	0.018	0.024	0.024	0.024	0.024	0.033
	(0.047)	(0.047)	(0.047)	(0.044)	(0.047)	(0.046)	(0.046)	(0.046)	(0.046)	(0.041)
Urban	-0.079	-0.079	-0.079	-0.574	-0.079	-0.082	-0.082	-0.082	-0.082	-0.067
	(0.045)*	(0.045)*	(0.045)*	(0.274)**	(0.045)*	(0.044)*	(0.044)*	(0.044)*	(0.044)*	(0.041)
Wealth										
Quintile 2	-0.072	-0.072	-0.072	-0.071	-0.072	-0.081	-0.081	-0.081	-0.081	-0.087
	(0.072)	(0.072)	(0.072)	(0.067)	(0.072)	(0.069)	(0.069)	(0.069)	(0.069)	(0.065)
Quintile 3	-0.071	-0.071	-0.071	-0.067	-0.071	-0.085	-0.085	-0.085	-0.085	-0.163
	(0.069)	(0.069)	(0.069)	(0.064)	(0.069)	(0.067)	(0.067)	(0.067)	(0.067)	(0.287)
Quintile 4	-0.222	-0.222	-0.222	-0.214	-0.222	-0.237	-0.237	-0.237	-0.237	-0.299
	(0.071)***	(0.071)***	(0.071)***	(0.065)***	(0.071)***	(0.069)***	(0.069)***	(0.069)***	(0.069)***	(0.289)
Quintile 5 (high)	-0.115	-0.115	-0.115	-0.102	-0.115	-0.126	-0.126	-0.126	-0.126	-0.184
	(0.084)	(0.084)	(0.084)	(0.080)	(0.084)	(0.081)	(0.081)	(0.081)	(0.081)	(0.292)
Smoking behaviour										
Friends	-0.445	-0.445	-0.445	-0.451	-0.445	-0.425	-0.425	-0.425	-0.425	-0.427
	(0.059)***	(0.059)***	(0.059)***	(0.061)***	(0.059)***	(0.057)***	(0.057)***	(0.057)***	(0.057)***	(0.059)***
Father	-0.029	-0.029	-0.029	-0.020	-0.029	-0.030	-0.030	-0.030	-0.030	-0.024
	(0.037)	(0.037)	(0.037)	(0.036)	(0.037)	(0.036)	(0.036)	(0.036)	(0.036)	(0.034)
Brother(s)	-0.092	-0.092	-0.092	-0.085	-0.092	-0.088	-0.088	-0.088	-0.088	-0.087
	(0.045)**	(0.045)**	(0.045)**	(0.042)**	(0.045)**	(0.043)**	(0.043)**	(0.043)**	(0.043)**	(0.039)**
Ethnicity -Kinh	0.026	0.026	0.026	0.018	0.026	-0.005	-0.005	-0.005	-0.005	-0.010
	(0.070)	(0.070)	(0.070)	(0.066)	(0.070)	(0.067)	(0.067)	(0.067)	(0.067)	(0.062)
Lived away	-0.068	-0.068	-0.068	-0.060	-0.068	-0.058	-0.058	-0.058	-0.058	-0.049
	(0.041)*	(0.041)*	(0.041)*	(0.039)	(0.041)*	(0.040)	(0.040)	(0.040)	(0.040)	(0.037)
Read/write	-0.092	-0.092	-0.092	-0.095	-0.092	-0.066	-0.066	-0.066	-0.066	-0.077
	(0.100)	(0.100)	(0.100)	(0.093)	(0.100)	(0.096)	(0.096)	(0.096)	(0.096)	(0.089)
Paid work	-0.150	-0.150	-0.150	-0.151	-0.150	-0.137	-0.137	-0.137	-0.137	-0.130
	(0.042)***	(0.042)***	(0.042)***	(0.040)***	(0.042)***	(0.041)***	(0.041)***	(0.041)***	(0.041)***	(0.039)***
Friends	-0.187	-0.187	-0.187	-0.201	-0.187	-0.169	-0.169	-0.169	-0.169	-0.190
	(0.087)**	(0.087)**	(0.087)**	(0.086)**	(0.087)**	(0.084)**	(0.084)**	(0.084)**	(0.084)**	(0.081)**
Clubs	0.110	0.110	0.110	0.097	0.110	0.108	0.108	0.108	0.108	0.097
	(0.041)***	(0.041)***	(0.041)***	(0.039)***	(0.041)***	(0.040)***	(0.040)***	(0.040)***	(0.040)***	(0.037)***
Born < 1986	0.450	0.450	0.450	0.442	0.450	0.484	0.484	0.484	0.484	0.475
	(0.043)***	(0.043)***	(0.043)***	(0.042)***	(0.043)***	(0.043)***	(0.043)***	(0.043)***	(0.043)***	(0.041)***
Region										
North East	-0.012	-0.012	-0.012	0.000	-0.012	-0.121	-0.121	-0.121	-0.121	-0.123
	(0.074)	(0.074)	(0.074)	(0.069)	(0.074)	(0.073)*	(0.073)*	(0.073)*	(0.073)*	(0.067)*
North Central Coast	0.026	0.026	0.026	0.018	0.026	0.016	0.016	0.016	0.016	0.014
	(0.068)	(0.068)	(0.068)	(0.064)	(0.068)	(0.067)	(0.067)	(0.067)	(0.067)	(0.063)
South Central Coast	0.056	0.056	0.056	0.047	0.056	0.016	0.016	0.016	0.016	0.007
	(0.078)	(0.078)	(0.078)	(0.073)	(0.078)	(0.075)	(0.075)	(0.075)	(0.075)	(0.069)
Central Highlands	0.094	0.094	0.094	0.072	0.094	-0.060	-0.060	-0.060	-0.060	-0.095
	(0.086)	(0.086)	(0.086)	(0.083)	(0.086)	(0.090)	(0.090)	(0.090)	(0.090)	(0.085)
South East	0.092	0.092	0.092	0.081	0.092	-0.092	-0.092	-0.092	-0.092	-0.137
	(0.053)*	(0.053)*	(0.053)*	(0.051)	(0.053)*	(0.073)	(0.073)	(0.073)	(0.073)	(0.068)**
Mekong River Delta	0.251	0.251	0.251	0.244	0.251	0.168	0.168	0.168	0.168	0.156
	(0.082)***	(0.082)***	(0.082)***	(0.076)***	(0.082)***	(0.076)**	(0.076)**	(0.076)**	(0.076)**	(0.071)**
Constant	-2.044	-2.043	-2.043	-1.304	-2.044	-6.786	-6.786	-6.786	-6.786	-7.161
	(1.461)	(1.461)	(1.462)	(1.410)	(1.461)	(2.240)***	(2.240)***	(2.240)***	(2.240)***	(2.043)***
Gamma (ρ)	0.332	0.332	0.332	0.362	0.332	0.322	0.322	0.322	0.322	0.358
	(0.012)***	(0.012)***	(0.012)***	(0.017)***	(0.012)***	(0.012)***	(0.012)***	(0.012)***	(0.012)***	(0.017)***
Frailty (θ)	–	0.000	0.000	1.876	0.000	–	0.000	0.000	0.000	2.602
		(0.000)	(0.000)	(2.414)	(0.000)		(0.000)	(0.000)	(0.000)	(3.225)
LR test, θ=0, *p-value*	–	1.000	1.000	0.009	1.000	–	1.000	1.000	1.000	0.012
Log likelihood	-968.6	-968.6	-968.6	-965.8	-968.6	-964.9	-964.9	-964.9	-964.9	-962.3

Note: Standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%. Red River Delta is the reference region.
Number of observations = 6721; Number of subjects = 1809; Number of failures = 582.

Table 4A. Discrete-time cloglog duration results for starting smoking

Models	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
				Frailty (Gamma)	*Frailty (Gamma)*	*Split Population*	*Split Population*
Duration dependency	None	Natural log	Non parametric	Natural log	Non parametric	Natural log	Non parametric
Prices (in logs)							
CPI tobacco	0.700	-1.646	-1.527	-1.704	-1.593	-1.689	-1.608
	(0.752)	(0.725)**	(0.733)**	(0.790)**	(0.777)**	(0.799)**	(0.774)**
Student	-0.743	-0.319	-0.323	-0.358	-0.334	-0.352	-0.332
	(0.101)***	(0.105)***	(0.105)***	(0.116)***	(0.112)***	(0.114)***	(0.110)***
Urban	0.129	0.141	0.149	0.181	0.164	0.219	0.175
	(0.103)	(0.104)	(0.104)	(0.118)	(0.115)	(0.118)*	(0.119)
Wealth							
Quintile 2	0.200	0.194	0.174	0.215	0.185	0.215	0.185
	(0.170)	(0.170)	(0.170)	(0.188)	(0.179)	(0.187)	(0.177)
Quintile 3	0.251	0.215	0.199	0.237	0.209	0.216	0.202
	(0.162)	(0.162)	(0.162)	(0.180)	(0.171)	(0.178)	(0.168)
Quintile 4	0.596	0.559	0.541	0.634	0.574	0.613	0.565
	(0.163)***	(0.162)***	(0.162)***	(0.186)***	(0.189)***	(0.178)***	(0.172)***
Quintile 5 (high)	0.429	0.296	0.297	0.356	0.321	0.342	0.307
	(0.196)**	(0.195)	(0.195)	(0.220)	(0.213)	(0.218)	(0.204)
Smoking behaviour							
Friends	1.253	1.225	1.228	1.299	1.260	1.291	1.251
	(0.154)***	(0.154)***	(0.154)***	(0.167)***	(0.176)***	(0.160)***	(0.161)***
Father	0.028	0.046	0.040	0.052	0.044	0.042	0.042
	(0.086)	(0.086)	(0.086)	(0.097)	(0.091)	(0.097)	(0.090)
Brother(s)	0.210	0.236	0.237	0.259	0.246	0.235	0.241
	(0.099)**	(0.099)**	(0.099)**	(0.113)**	(0.108)**	(0.112)**	(0.105)**
Ethnicity -Kinh	-0.046	-0.075	-0.066	-0.075	-0.065	-0.074	-0.065
	(0.160)	(0.161)	(0.160)	(0.179)	(0.168)	(0.179)	(0.167)
Lived away	0.179	0.170	0.172	0.191	0.181	0.169	0.171
	(0.093)*	(0.093)*	(0.093)*	(0.106)*	(0.101)*	(0.105)	(0.098)*
Read/write	0.496	0.374	0.361	0.393	0.371	0.389	0.373
	(0.232)**	(0.232)	(0.232)	(0.257)	(0.244)	(0.255)	(0.241)
Paid work	0.319	0.359	0.376	0.400	0.395	0.411	0.394
	(0.100)***	(0.099)***	(0.099)***	(0.112)***	(0.113)***	(0.110)***	(0.108)***
Friends	0.468	0.530	0.527	0.564	0.545	0.586	0.553
	(0.208)**	(0.209)**	(0.209)**	(0.230)**	(0.221)**	(0.229)**	(0.220)**
Clubs	-0.078	-0.182	-0.190	-0.210	-0.204	-0.228	-0.206
	(0.095)***	(0.095)*	(0.095)**	(0.108)*	(0.106)*	(0.108)**	(0.105)**
Born < 1986	-0.160	-0.743	-0.813	-0.822	-0.851	-0.817	-0.842
	(0.105)	(0.116)***	(0.120)***	(0.135)***	(0.157)***	(0.128)***	(0.134)***
Region							
North East	0.083	0.014	0.022	0.025	0.027	-0.043	0.002
	(0.168)	(0.168)	(0.167)	(0.190)	(0.177)	(0.186)	(0.178)
North Central Coast	-0.059	-0.067	-0.075	-0.078	-0.077	-0.080	-0.077
	(0.161)	(0.161)	(0.161)	(0.178)	(0.168)	(0.176)	(0.167)
South Central Coast	0.137	-0.040	-0.016	-0.048	-0.022	-0.035	-0.021
	(0.174)	(0.176)	(0.176)	(0.198)	(0.186)	(0.201)	(0.187)
Central Highlands	-0.174	-0.176	-0.183	-0.221	-0.198	-0.239	-0.204
	(0.205)	(0.205)	(0.206)	(0.228)	(0.218)	(0.228)	(0.217)
South East	-0.125	-0.171	-0.165	-0.204	-0.178	-0.222	-0.184
	(0.123)	(0.121)	(0.121)	(0.138)	(0.132)	(0.139)	(0.133)
Mekong River Delta	-0.303	-0.526	-0.517	-0.593	-0.549	-0.625	-0.560
	(0.189)	(0.189)***	(0.190)***	(0.213)***	(0.213)***	(0.207)***	(0.210)***
Duration dependency							
ln *t*	–	1.135	–	1.314	–	1.309	–
		(0.091)***		(0.146)***		(0.119)***	
Temporal dummies							
d_2	–	–	0.708	–	0.718	–	0.713
			(0.175)***		(0.176)***		(0.175)***
d_3	–	–	1.128	–	1.153	–	1.142
			(0.171)***		(0.183)***		(0.173)***
d_4	–	–	1.610	–	1.659	–	1.639
			(0.174)***		(0.214)***		(0.182)***
d_5	–	–	2.082	–	2.160	–	2.132
			(0.185)***		(0.271)***		(0.206)***
d_6	–	–	2.075	–	2.190	–	2.162
			(0.205)***		(0.355)***		(0.258)***
d_7	–	–	2.103	–	2.254	–	2.237
			(0.232)***		(0.446)***		(0.339)***
d_8	–	–	1.081	–	1.263	–	1.275
			(0.478)**	–	(0.660)*		(0.599)**
Constant	-7.529	2.417	1.936		2.185	2.546	2.317
	(3.537)***	(3.392)	(3.437)	–	(3.613)	(3.729)	(3.622)
Participation							
Constant	–	–	–	–	–	-1.845	-2.782
						(0.405)***	(1.748)
Frailty (θ)	–	–	–	0.354	0.146	–	–
				(0.228)	(0.370)		
LR test, θ=0, *p-value*	–	–	–	0.045	0.342	–	–
Log likelihood	-1851.7	-1762.3	-1751.4	-1760.9	-1751.4	-1759.3	-1751.3

Note: Standard erros in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%. Red River Delta is the reference region.
Number of observations = 6721; Number of subjects = 1809; Number of failures = 582.

Table 4B. Dscrete-time cloglog duration results for starting smoking

<u>Models</u>	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Frailty (Gamma)	*Frailty (Gamma)*	*Split Population*	*Split Population*
Duration dependency	None	Natural log	Non parametric	Natural log	Non parametric	Natural log	Non parametric
Prices (in logs)							
Vinataba	0.972	0.021	-0.028	0.020	-0.038	-0.071	-0.076
	(0.411)**	(0.421)	(0.426)	(0.449)	(0.437)	(0.452)	(0.450)
BAT 555	0.828	-1.283	-1.057	-1.250	-1.068	-1.200	-1.081
	(0.525)	(0.522)**	(0.533)**	(0.550)**	(0.543)**	(0.552)**	(0.545)**
Waterpipe	-0.448	-0.155	-0.133	-0.137	-0.128	-0.109	-0.116
	(0.170)***	(0.175)	(0.176)	(0.189)	(0.181)	(0.187)	(0.184)
Student	-0.699	-0.324	-0.325	-0.358	-0.330	-0.357	-0.333
	(0.102)***	(0.105)***	(0.106)***	(0.115)***	(0.111)***	(0.114)***	(0.110)***
Urban	0.127	0.152	0.158	0.188	0.166	0.227	0.183
	(0.103)	(0.104)	(0.104)	(0.117)	(0.114)	(0.117)*	(0.121)
Wealth							
Quintile 2	0.168	0.206	0.187	0.223	0.192	0.223	0.197
	(0.170)	(0.170)	(0.170)	(0.186)	(0.176)	(0.186)	(0.177)
Quintile 3	0.221	0.234	0.216	0.249	0.221	0.231	0.219
	(0.162)	(0.162)	(0.162)	(0.178)	(0.167)	(0.178)	(0.167)
Quintile 4	0.562	0.590	0.569	0.650	0.585	0.638	0.591
	(0.163)***	(0.163)***	(0.163)***	(0.184)***	(0.188)***	(0.178)***	(0.175)***
Quintile 5 (high)	0.423	0.334	0.330	0.379	0.340	0.370	0.338
	(0.197)**	(0.197)*	(0.197)*	(0.218)*	(0.209)	(0.219)*	(0.205)*
Smoking behaviour							
Friends	1.266	1.224	1.228	1.289	1.243	1.285	1.248
	(0.155)***	(0.154)***	(0.154)***	(0.166)***	(0.176)***	(0.160)***	(0.161)***
Father	0.030	0.048	0.043	0.054	0.045	0.046	0.045
	(0.086)	(0.086)	(0.086)	(0.095)	(0.089)	(0.096)	(0.090)
Brother(s)	0.219	0.243	0.243	0.261	0.247	0.240	0.246
	(0.099)**	(0.099)**	(0.099)**	(0.111)**	(0.105)**	(0.112)**	(0.104)**
Ethnicity -Kinh	-0.089	-0.019	-0.015	-0.018	-0.014	-0.013	-0.011
	(0.158)	(0.159)	(0.159)	(0.176)	(0.163)	(0.176)	(0.165)
Lived away	0.181	0.160	0.163	0.177	0.167	0.158	0.161
	(0.093)*	(0.094)*	(0.094)*	(0.105)*	(0.099)*	(0.105)	(0.098)*
Read/write	0.522	0.340	0.330	0.354	0.334	0.348	0.338
	(0.232)**	(0.232)	(0.232)	(0.254)	(0.238)	(0.255)	(0.240)
Paid work	0.346	0.364	0.380	0.399	0.388	0.413	0.395
	(0.100)***	(0.100)***	(0.100)***	(0.111)***	(0.112)***	(0.110)***	(0.108)***
Friends	0.483	0.523	0.520	0.550	0.528	0.579	0.544
	(0.208)**	(0.209)**	(0.209)**	(0.227)**	(0.217)**	(0.227)**	(0.220)**
Clubs	-0.091	-0.185	-0.192	-0.210	-0.199	-0.232	-0.209
	(0.095)	(0.095)*	(0.096)**	(0.107)**	(0.105)*	(0.108)**	(0.106)**
Born < 1986	-0.090	-0.783	-0.843	-0.850	-0.862	-0.853	-0.873
	(0.107)	(0.123)***	(0.127)***	(0.140)***	(0.166)***	(0.134)***	(0.145)***
Region							
North East	-0.060	0.187	0.173	0.193	0.176	0.133	0.161
	(0.170)	(0.172)	(0.172)	(0.191)	(0.178)	(0.190)	(0.181)
North Central Coast	-0.162	-0.082	-0.083	-0.090	-0.083	-0.083	-0.080
	(0.165)	(0.164)	(0.164)	(0.179)	(0.168)	(0.178)	(0.170)
South Central Coast	0.194	0.124	0.128	0.121	0.126	0.115	0.122
	(0.168)	(0.171)	(0.172)	(0.192)	(0.177)	(0.200)	(0.184)
Central Highlands	-0.378	0.063	0.016	0.019	0.012	-0.011	0.003
	(0.221)*	(0.221)	(0.222)	(0.242)	(0.228)	(0.243)	(0.231)
South East	-0.209	0.207	0.166	0.162	0.161	0.129	0.153
	(0.174)	(0.174)	(0.176)	(0.194)	(0.182)	(0.194)	(0.186)
Mekong River Delta	-0.311	-0.345	-0.342	-0.403	-0.355	-0.445	-0.377
	(0.178)*	(0.180)***	(0.180)**	(0.202)**	(0.197)*	(0.199)**	(0.200)*
Duration dependency							
ln *t*	–	1.167	–	1.317	–	1.328	–
		(0.098)***		(0.148)***		(0.123)***	
Temporal dummies							
d_2	–	–	0.710	–	0.715	–	0.715
			(0.175)***		(0.185)***		(0.175)***
d_3	–	–	1.141	–	1.153	–	1.155
			(0.172)***		(0.219)***		(0.175)***
d_4	–	–	1.631	–	1.654	–	1.658
			(0.177)***		(0.279)***		(0.187)***
d_5	–	–	2.103	–	2.140	–	2.151
			(0.190)***		(0.368)***		(0.216)***
d_6	–	–	2.104	–	2.159	–	2.185
			(0.211)***		(0.464)***		(0.275)***
d_7	–	–	2.155	–	2.227	–	2.280
			(0.242)***		(0.681)***		(0.366)***
d_8	–	–	1.157	–	1.244	–	1.337
			(0.486)**		(5.660)*		(0.630)**
Constant	-16.013	8.270	6.429	7.624	6.540	7.757	6.903
	(5.150)***	(5.403)	(5.560)	(5.714)	(5.660)	(5.723)	(5.753)
Participation							
Constant	–	–	–	–	–	-1.922	-2.904
						(0.436)***	(2.095)**
Frailty (θ)	–	–	–	0.300	0.069	–	–
				(0.227)	(0.376)		
LR test, θ=0, *p-value*	–	–	–	0.076	0.427	–	–
Log likelihood	-1844.9	-1761.3	-1751.1	-1760.3	-1751.1	-1758.7	-1751.0

Note: Standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%. Red River Delta is the reference region.
Number of observations = 6721; Number of subjects = 1809; Number of failures = 582.