Distribution Agreement

Signature:

______________________________ _______________

Deliana Kostova Date

Essays on Global Youth Tobacco Use: The Role of Cigarette Prices and Regulation

By

Deliana Kostova
Doctor of Philosophy

Economics

____
Sara Markowitz
Advisor

____
David Frisvold
Committee Member

____
David Howard
Committee Member

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____________________ Date

Essays on Global Youth Tobacco Use: The Role of Cigarette Prices and Regulation

By

Deliana Kostova

M.A., Emory University, 2003
B.A., Berea College, 2000

Advisor: Sara Markowitz, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2010

Abstract

Essays on Global Youth Tobacco Use: The Role of Cigarette Prices and Regulation

By Deliana Kostova

The first chapter of this dissertation estimates the impact of cigarette prices on youth smoking in lower-income countries using data from the Global Youth Tobacco Survey (GYTS). Country-level heterogeneity is addressed with fixed effects and by directly controlling for confounding environmental factors such as local anti-smoking sentiment, cigarette advertising, anti-smoking media messages, and compliance with youth access restrictions. I find that cigarette price is an important determinant of both smoking participation and conditional demand. The estimated price elasticity of participation is -0.63. The likelihood of participation decreases with anti-smoking sentiment and increases with exposure to cigarette advertising. The estimated price elasticity of conditional cigarette demand is approximately -1.2. Neither anti-smoking sentiment, cigarette advertising, nor access restrictions have an impact on the intensity of smoking among current smokers, but exposure to anti-smoking media may reduce the number of cigarettes smoked.

The second chapter investigates the impact of cigarette prices on smoking initiation and cessation among youth in developing countries using data from the Global Youth Tobacco Survey (GYTS). The effect of price is identified by country fixed effects which control for unobserved environmental characteristics such as anti-smoking sentiment. Three types of duration analysis are used to examine the sensitivity of the results with respect to empirical specification. These are the discrete-time logit model, the Cox hazard model, and the split-population duration model. Unlike the unsplit logit and Cox models which assume that all subjects have positive hazards of initiation (cessation), the split-population model allows for the possibility that for some individuals the hazard is zero. A statistically significant impact of cigarette price on the initiation (cessation) hazards is identified in the split-population analysis but not by the logit and Cox models. The conclusion is that individuals who are intrinsic non-smokers may not be as responsive to cigarette prices, so including them in the sample along with the potential smokers will attenuate the overall price effect. After accounting for the probability that some people will never smoke and some smokers will never quit, the price elasticity of the hazard of starting smoking is estimated at -0.165. The price elasticity of the hazard of quitting is estimated at 0.27.

Essays on Global Youth Tobacco Use: The Role of Cigarette Prices and Regulation

By

Deliana Kostova

M.A., Emory University, 2003
B.A., Berea College, 2000

Advisor: Sara Markowitz, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Economics
2010

Acknowledgements

This dissertation would not have happened without my advisor, Professor Sara Markowitz. Her knowledge and direction have been instrumental in the writing of this research. I will always be grateful that I have been able to learn from her, work with her, and benefit from her guidance.

I am very thankful to Prof. David Howard, who introduced me to the world of public health and its connection to economics. His perspective has always been truly valuable and appreciated. I am also grateful to Prof. David Frisvold for the knowledge and advice he has shared with me when I needed it. And last but certainly not least, thank you to Prof. Maria Arbatskaya, who was one of my first teachers when I started my graduate studies, and who very patiently supported me through the years to their completion.

This dissertation is dedicated to my mother, Dr. Nina Kostova, who enabled me to write it by helping me and my family when I couldn't be there.

Table of Contents Page

Table of Contents continued Page

Table of Contents continued | Page

List of Tables Page

Chapter 1

List of Tables continued Page

List of Figures | Page

Chapter 1

1. Introduction

Background

Tobacco consumption has long been established as a leading cause of preventable death, with 100 million deaths attributed to it during the 20th century, and nearly one billion deaths projected for the 21st century (World Health Organization, 2008). Today, the public health damage from tobacco is roughly similar in developed and developing countries, but the geography of smoking is shifting away from industrialized nations. Based on current trends, developing countries are expected to carry 78% of the world's tobacco-related mortality by 2020 (Tobacco Control Country Profiles, 2003).

The decline of smoking in developed countries has been accompanied by an increase in the presence of the tobacco industry in less restrictive markets like Asia, Africa, and Latin America. In these emerging markets, restrictions on tobacco advertising or youth access are not always properly enforced, increasing the susceptibility of children and youth to the attractions of smoking. It is estimated that, worldwide, one in seven teenagers smokes, and a quarter of them have tried their first cigarette before the age of 10 (The Tobacco Atlas, 2006). Among lower-income countries, smoking among children ages 13 to 15 is particularly common in Eastern Europe and Latin America, where 2005 estimates of smoking prevalence in this age group are 14 and 12 percent, respectively (Global Youth Tobacco Survey (GYTS) 1999-2006).

Adolescents are a group of special interest to the global anti-tobacco effort because smoking habits are primarily established in youth. The regional variation in youth smoking patterns worldwide is substantial and corresponds to variations in market characteristics, media influences, and cultural perceptions of smoking. According to recent GYTS data, the lowest smoking rates occur in predominantly Muslim nations (mostly below 7%), where, coincidentally, the religion-based view of tobacco is unfavorable. Similarly low rates are observed in most African countries, especially those with very low income per capita and a correspondingly poor market base. On the other hand, regions where youth smoking is common (Eastern Europe and Latin America), have neither extremely low domestic income nor cultural prohibitions against smoking.

Besides local market and cultural traits, other dimensions of the global youth smoking environment include the influence of the media and the accessibility of tobacco. These can vary depending on local legislation and the actual compliance with tobacco-related restrictions. In Poland, for example, where about 1 in 4 teens were reported to smoke in 2003 (GYTS 2003), cigarette sales to minors as well as tobacco advertising are fully banned. Despite such strict legislation on paper, 45% of Polish GYTS respondents reported seeing cigarette advertisements and 62% reported no difficulty buying cigarettes in shops. Clearly, there are multiple and often conflicting factors that come together in shaping global youth smoking patterns. Separating and evaluating their individual effects is important in determining the best way to target this public health challenge.

Given the importance of cigarette smoking as a leading cause of preventable death, much attention has been paid to the question of how to reduce smoking prevalence. The primary policy tools used for smoking deterrence are prices (in the form of taxation), tobacco advertising restrictions, smoking restrictions, and various types of anti-smoking campaigns. The effectiveness of these tools has been under evolving scrutiny ever since the harm from smoking was established, and research on the subject is extensive. Unfortunately, much of the literature has been affected by data limitations (omitted variables, econometric endogeneity) that may preclude consistent parameter estimation or affirmation of causal effects. Data unavailability has also confined the existing literature to focus almost completely on the United States and, occasionally, other industrialized countries. From a policy perspective this is a particularly weak point since developing nations, unlike First World countries, have increasing smoking prevalence - and may not have the same pattern of responsiveness to anti-smoking policies as the U.S. population.

Although more research is needed to draw conclusions about the impact of tobacco control mechanisms in lower-income countries, such research has been hindered by the logistical difficulty of obtaining smoking data from developing regions. Only recently has data on youth smoking in lower income countries become available through CDC/WHO's GYTS. GYTS data have considerable advantages as they are produced by standardized questionnaires for multiple countries and multiple years and provide rich information on youth smoking prevalences and environment. However, these data have been mostly used in descriptive applications rather than vigorous policy evaluation due to the difficulty of obtaining cigarette prices for most GYTS countries and years. Since policy evaluations are produced by models of cigarette consumption, which by standard economic definition require data on cigarette prices, unavailability of price data strongly interferes with the estimation of policy impacts. This paper overcomes this limitation by using private cigarette price data from the Economist Intelligence Unit's (EIU) World

Cost of Living Survey. These price data are underutilized because they are costly and not publicly available, but they are an excellent source for local prices in many of the countries and years covered by GYTS.

Goal and contributions

The main goal of this research is to establish the level of price responsiveness of cigarette consumption among youth in low-to-mid-income countries. A major econometric concern in this analysis is the probable presence of unobserved environmental characteristics that may influence both local prices and smoking patterns. If such unobserved characteristics are left unadressed, their effects would be picked up by the measure of price, resulting in overestimation of price elasticities, and/or would remain in the error term, leading to inconsistent estimates. In either case, inference from such a model would be misleading with regards to policy decisions.

The contributions of the paper are as follows. First, this is the first study to investigate the price elasticity of youth cigarette demand in developing countries and thus to provide a global policy-relevant perspective on youth price responsiveness. Second, it addresses the problem of unobserved heterogeneity very thoroughly and extends the identification methods beyond what has been done so far in the literature on US cigarette demand. More specifically, the effect of price on cigarette demand is identified by 1) introducing area fixed effects, 2) including a measure of the local sentiment against smoking, and 3) including controls for major confounding environmental factors such as the prevalence of cigarette advertising, the prevalence of anti-smoking media messages, and local compliance with minimum-age sale restrictions.

To the best of our knowledge, this is the first study in the literature on smoking to discuss the responsiveness of youth smoking to policy-related factors like cigarette advertising, anti-smoking media campaigns, and youth access restrictions - while also controlling for the observed effectiveness of these policies. For example, we control not just for whether or not cigarette advertising is permitted, but for how effective such advertising is in reaching an audience. Similarly, we control not just for the nominal presence of bans on cigarette sales to minors, but we control for how effective such bans are in preventing youth from buying cigarettes. We control not just for whether or not anti-tobacco campaigns are currently being run, but we control for the actual level of exposure to such campaigns. Accounting for the effectiveness of smoking-related policies provides better information on policy effects than simply controlling for nominal policy presence, since presence alone does not necessarily reflect equal levels of enforcement or compliance across countries.

A number of different specifications and estimation techniques are used as robustness checks. All of the estimators seek to accommodate the atypical nature of the data where the vast majority of outcomes have zero value. This is necessary since most of the surveyed individuals in the sample report zero cigarette consumption. We compare results from the two-part model and one-part Tobit and zero-inflated Poisson (ZIP) models. The preferred methodology is the two-part model, which is shown to be more appropriate using both statistical testing and economic theory. The main conclusions of this paper are as follows.

We find that higher cigarette prices are effective in reducing both smoking participation and conditional cigarette demand (aka smoking intensity among current smokers). The price elasticity of smoking participation ranges from -0.56 to -0.88 depending on the specification. The price elasticity of conditional demand is estimated to be approximately -1.2. Price has a significant effect on participation and intensity even after anti-smoking sentiment, media exposures, access restrictions, and fixed country-level unobservables are factored in. Macro variables like anti-smoking sentiment, the local prevalence of cigarette advertising, and youth access restrictions are also shown to lower the likelihood of participation. We estimate that perfect, as opposed to the currently sporadic, compliance with youth access restrictions in developing countries may cut participation rates by more than half. Regarding conditional demand, we find that besides prices, few macro variables can influence smoking intensity among smokers. Sentiment, advertising, and access restrictions do not seem to affect smoking behavior once smoking is established. However, there is evidence that increased prevalence of anti-smoking media messages may lead to slight reductions in both participation rates and smoking intensity.

2. Literature review

A defining characteristic of the existing research on youth smoking is that it is almost exclusively confined to the United States. Very few studies investigate the price sensitivity of smoking for adolescents in lower-income countries, although there is one study (Lance et al 2004) which evaluates micro-level data from Russia and China. Due to the lack of international studies, most of this section will discuss research based on U.S. data.

An important issue shared by most of the existing literature on U.S. youth smoking has been the difficulty of accounting for unobserved state-level heterogeneity, particularly state-level anti-smoking sentiment, that may be correlated with both prices and smoking patterns. If the influence of such unobserved characteristics is ignored, their effects would be picked up by the measure of price, resulting in overestimated and inconsistent price elasticity estimates. Only recently has work emerged that controls for regional variations in the public attitude toward smoking, either through fixed effects (DeCicca et al 2002, Carpenter and Cook 2008), or through direct inclusion of a state anti-smoking sentiment variable (DeCicca et al 2008, Carpenter and Cook 2008). The conclusions on the effect of prices in these studies have been contradictory, with DeCicca et al (2002, 2008) finding insignificant price effects and Carpenter and Cook (2008) finding repeatedly significant price effects.

Among the earliest works to examine the determinants of smoking among youths and young adults are Lewit et al (1981) and Lewit and Coate (1982). Using cross-sectional analysis of micro-level data from national health surveys, both studies find that price is more likely to affect the decision to smoke than the quantity of cigarettes smoked. Lewit et al (1981) look at a younger age group than Lewit and Coate (1982) (12-17 year olds vs 20-25 year olds) and estimate a significantly higher price elasticity of cigarette demand for teenagers than young adults (-1.44 vs -0.89). Both studies find that smoking in the younger age groups is more sensitive to prices as compared to older age groups.

Different results are obtained by Wasserman et al (1991) who find that the effect of prices on youth smoking participation and consumption is not significant regardless of age group. A key feature of their analysis is the addition of a control for state anti-smoking regulations, which they find to be highly effective in reducing smoking among teenagers. However, the anti-smoking regulation variable in this study has been blamed for causing insignificant price elasticity estimates due to its high correlation with the price level (Chaloupka and Grossman 1996, Chaloupka and Wechsler 1995, Wasserman et al 1991). It has also been criticized for being irrelevant to the teenage population since it gives much weight to smoking restrictions in private worksites (Chaloupka and Grossman 1996, Chaloupka and Wechsler 1995).

Following up on their critique of the Wasserman et al (1991) study, Chaloupka and Wechsler (1997) and Chaloupka and Grossman (1996) provide additional estimates of the effect of prices and smoking restrictions on youth smoking. Both studies use similar methodology, the two-part model of smoking participation and demand, but their samples differ in age - Chaloupka and Wechsler (1997) look at college students from the 1993 Harvard Alcohol Study while Chaloupka and Grossman (1996) use high-school students from the Monitoring the Future project. Unlike Wasserman et al (1991), both of these studies find large and significant price elasticities of cigarette demand even after controlling for smoking restrictions: -1.11 (Chaloupka and Wechsler 1997) and -1.31 (Chaloupka and Grossman 1996). These results confirm Lewit's findings from the early 1980s that youth smoking is very sensitive to prices and more so for youths than adults. In addition, both studies conclude that while restrictions on smoking in public places and schools discourage smoking among youths, they do not dominate or eliminate the expected impact of prices.

Other relatively recent studies that provide estimates of the price elasticity of cigarette demand include Tauras, Markowitz and Cawley (2005), Harris and Chan (1999), Tauras and Chaloupka (1999), Ross and Chaloupka (2003, 2004), and Czart et al (2001). With the exception of Tauras, Markowitz and Cawley (2005) and Tauras and Chaloupka (1999), who are among the first to use individual fixed effects in this context, all of these studies employ the two-part model of smoking demand in a cross-sectional framework. Although the empirical methods are similar, each of these studies contributes to the literature by offering price elasticity estimates from different datasets and by using different measures of price and/or public policies. All of them agree that price has a negative and significant effect on youth smoking. Czart et al (2001) look at a sample of college students from the 1997 Harvard Alcohol Study and examine the impact of campus-level tobacco restrictions in addition to prices. They calculate a total price elasticity of -0.88 but find that the impact of campus-level anti-tobacco policies is inconclusive. Harris and Chan (1999) use the 2002-03 Current Population Survey to provide price elasticity estimates by age group. Their estimates agree with previous findings that the price sensitivity of cigarette demand steadily decreases with age. They calculate price elasticities as large as -0.83 for teenagers and as small as -0.20 for young adults. Tauras and Chaloupka (1999) utilize the panel nature of their young adult dataset obtained from the Monitoring the Future program. After controlling for individual fixed effects, they estimate a price elasticity of -0.79. Ross and Chaloupka (2003, 2004) contribute to the literature by introducing a new measure of cigarette prices that reflects the teenagers' perception of the price. This teen-specific price is constructed from self-reported cigarette prices available from a 1996 high school survey, and represents what the respondents think cigarettes cost or what they would pay for the cigarettes of their choice. In Ross and Chaloupka (2003), the estimate of the perception-adjusted price elasticity is higher than the list-price elasticity and ranges from -0.67 to -1.02. However, the authors recognize that self-reported prices may suffer from endogeneity which could lead to over-estimation of the price effect as smokers choose cigarettes with lower prices. Ross and Chaloupka (2004) extend their 2003 study by controlling for public smoking restrictions and youth access restrictions, where the latter are constructed from actual compliance rates. The perception-adjusted price elasticity estimates are similar to those produced in the 2003 study and range from -0.7 to -1.0. The authors also confirm that compliance with youth access restrictions reduces smoking prevalence.

Although all of the papers discussed so far provide valuable contributions to the economics of youth smoking, they do not account for the possibility that unobserved state heterogeneity in general and anti-smoking sentiment in particular may correlate with state-level taxes or tobacco regulations, leading to a spurious negative relationship between prices and smoking. This drawback has not passed unrecognized by researchers, and most recent work addresses the omitted variables bias by either employing state fixed effects or explicitly including controls for state anti-smoking sentiment. DeCicca et al (2002) use the 1998 National Education Longitudinal Survey to investigate the impact of prices on youth smoking participation in a cross-sectional framework, where state anti-smoking sentiment is proxied by a number of state anti-tobacco policy measures. In these cross sectional models, they compute price elasticities of participation of -2.03 for the youngest sample, eighth graders, -1.31 for the tenth graders, and -0.72 for the twelfth graders. In the same study, the authors also evaluate the price elasticity of smoking initiation using state fixed effects in a longitudinal framework. In this case, they find that price has no significant effect on smoking initiation.

In a subsequent study, DeCicca et al (2008) construct an explicit measure of state anti-smoking sentiment during the 1990s from the Tobacco Use Supplements of the Current Population Survey. They include this new measure in cross-sectional models of youth smoking demand using data from the 1992 and 2000 National Education Longitudinal Studies. The main conclusion is that the price effect on smoking participation disappears once state anti-smoking sentiment is directly controlled for. However, state anti-smoking sentiment is found to play a smaller role in the case of conditional demand, and the impact of prices on conditional demand can be significant depending on which survey cohort is evaluated. In one of their cohorts, the price effect is not significant while the other cohort produces a price elasticity of conditional demand of -0.52. The authors conclude that prices may indeed reduce smoking among youth who already smoke - but that the high price sensitivity of youth smoking participation and initiation reported by previous literature is overestimated and is a consequence of failing to account for the significant role of state anti-smoking sentiment.

In an attempt to reproduce the DeCicca et al (2008) results, Carpenter and Cook (2008) evaluate the effect of prices on youth smoking participation using 1993-2005 data from the Youth Risk Behavior Surveys. While their data source is different from DeCicca et al (2008), Carpenter and Cook (2008) use the same measure of anti-smoking sentiment as a control in a similar cross-sectional model of smoking participation. Their results contradict DeCicca et al (2008) by showing a statistically significant, albeit small, price elasticity of -0.14. In the same study, the authors also present alternative estimates of the price elasticity of participation produced by state fixed-effects methods. Again, they find a significant price effect and elasticities in the range of -0.23 to -0.56.

The majority of the domestic literature on youth smoking has been limited by a failure to control for unobserved state heterogeneity. The few studies that address this issue generate conflicting results regarding the impact of cigarette prices on youth smoking patterns. The literature on youth smoking in developing countries, however, has even bigger limitations. One example is reliance on aggregate data (Chapman and Richardson 1990). Aggregate data such as average cigarette consumption per capita has major simultaneity issues when estimated as a function of price. A recent study by Lance et al (2004) avoids this problem by using micro-level longitudinal surveys of individuals in a number of communities in Russia and China. Lance et al (2004) estimate price elasticities for China and Russia using fixed effects methods to allow for unobserved community-level heterogeneity. Their estimated price elasticity of teenage smoking is smaller than corresponding estimates for the U.S., ranging between 0 for China and -0.2 for Russia. However, their samples are restricted to relatively small numbers of males only, which may prevent extending the results to the general population.

This research advances the literature on youth smoking in developing countries by introducing the first worldwide model of smoking demand. It uses rich micro-level data from multiple countries over multiple years while building on empirical and methodological insights from an extensive U.S.-based literature on youth smoking. Drawing on previous domestic conclusions about the importance of unobserved geographic heterogeneity, the study controls for unobserved country-level heterogeneity through fixed effects, and addresses the omitted variable bias by including controls for anti-smoking sentiment, media influences, and access restrictions. To the best of our knowledge, it provides the most comprehensive picture of youth smoking demand and price sensitivity in the developing world.

3. Data and variables

The dataset is a combination of two main sources. Micro-level data on individual characteristics and smoking behavior are obtained from the Global Youth Tobacco Survey (GYTS). These are merged with country-level data on cigarette prices from the Economist Intelligence Unit's World Cost of Living Survey (EIU). This is the fist study to utilize GYTS data in combination with cigarette prices and is therefore an original analysis of youth's smoking decisions as a function of price.

The GYTS is a survey developed by the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) to track tobacco use of young people across countries with a common methodology. It has been conducted in 135 low-to-mid-income countries from the six major world regions (Africa, Europe, Americas, Southeast Asia, Middle East, and Western Pacific) in various years from 1999 to 2006. It captures prevalence, access, media exposure and attitudes related to tobacco use among individuals in school grades corresponding to ages 13 to 15, although in practice the age range of the survey is wider and covers individuals between the ages of 11 and 19.

The final datasets used in this analysis come in two sizes depending on the type of model applied to them, specifically depending on the type of fixed effects (FE). Since price data are available for 47 GYTS countries, the maximum size of the sample for this study includes the individuals from the 47 countries with available cigarette prices. All countries correspond to six world regions, and can be combined to form repeated region cross-sections to which region fixed effects can be applied. The final region-level dataset contains data on 491,660 individuals from 6 regions, corresponding to 47 countries and 159 local sites (i.e. cities/provinces). Since only 20 of these 47 countries are surveyed in multiple years, repeated country-level cross-sections which would allow the use of country fixed affects are available from 20 countries only. Therefore, the country-level dataset contains data on 349,930 individuals from 20 countries corresponding to 118 local sites. A list of the final set of countries, regions, and survey years is shown in Table 1. Based on the geographic location of each individual, it is possible to match the current smoking status and cigarette demand of each individual to the cigarette price that he/she is facing at that point in time, allowing us to model demand as a function of price and other relevant environmental and individual characteristics.

Cigarette Prices

Data on the price of cigarettes over time is obtained from the EIU World Cost of Living Survey. This is a privately developed survey by the publishers of The Economist magazine. It collects retail price data of a wide range of consumer products on a bi-annual basis from multiple cities worldwide, including many developing countries.

Cigarette price data are available on two different brands, a locally popular brand and an imported brand, usually Marlboro. Prices are collected from one or more cities in each country. If for a particular country cigarette price data come from multiple cities, the averaged national price is used in this study. Prices are in U.S. dollars based on the relevant exchange rate and are converted into real terms using the 2000 U.S. GDP deflator. They are also adjusted using purchasing power parities (PPP) obtained from the World Bank's World Development Indicators database. The PPP adjusts prices for the local standard of living and allows for better price comparison between countries. The lowest prices, on average, are found in Africa ($0.75 per pack), and the highest are in the Western Pacific (Singapore, $3.40). In the primary analysis of smoking initiation and cessation, I use local-brand cigarette prices, but a sensitivity analysis using Imported-brand prices is performed as well.

The final dataset used in this research excludes many of the original GYTS countries due to unavailability of matching price data. However, the geographic variation of price is increased by the fact that in some countries GYTS surveys were conducted in multiple local sites like cities or provinces. Where the GYTS city survey site matches the EIU city survey site, local city prices are used instead of the nationally averaged price. This produces geographic variation of price within country for some countries.

Individual-level variables

Variables that have individual-level variation include the outcome variables (i.e., smoking participation and conditional cigarette demand) as well as some explanatory variables that describe personal characteristics. Smoking participation is a binary variable equal to 1 if the individual describes himself as a smoker and has smoked at least one cigarette in the past month. Smoking participation varies across regions and countries. The highest smoking prevalence rates are observed in Eastern Europe (16%) and Latin America (12%), although smoking intensity among smokers is considerably lower in Latin America than in Europe. Some of the lowest prevalence rates are in countries in Africa and the Middle East (7%). Conditional cigarette demand is defined as the average number of cigarettes consumed in the previous month for each current smoker. Although the GYTS does not provide cigarette demand data in the form of a precise number of cigarettes smoked, demand can be approximately calculated from survey questions as follows. The GYTS provides information, in categorical ranges, on the number of days that smoking occurred in the past month as well as the average number of cigarettes smoked daily. With respect to smoking days, each respondent indicates if he has smoked on days in the past month, 1 to 2 days, 3 to 5 days, 6 to 9 days, 10 to 19 days, 20 to 29 days, or all 30 days. With respect to daily smoking intensity, each respondent indicates if on each of his smoking days he has smoked, on average, less than 1 cigarette, 1 cigarette, 2 to 5 cigarettes, 6 to 10 cigarettes, 11 to 20 cigarettes, or more than 20 cigarettes. Conditional cigarette demand is calculated as the midpoint of each person's smoking days category multiplied by the midpoint of his average daily cigarettes category, providing an approximation for the intensity of smoking. For the smokers in this sample, the range of average monthly cigarette consumption is 1.5 to 630 cigarettes.

Besides the outcome variables, other variables that vary by individual are personal attributes such as age (*Age*), gender (*Male*), parental smoking status (*Parental Smoking*), and availability of pocket money or personal income (*Pocket Money*). *Male* is a binary variable equal to 1 if the subject is male. The samples are relatively evenly represented by males and females, with the exception of Saudi Arabia where all survey participants are male. The average age in all countries is about 14 years. *Pocket Money* is a binary indicator equal to 1 if the subjects receives pocket money or personal income at the time of the interview. Except for Africa, more than half of the sampled teens worldwide and almost three-quarters of the teens in Europe and the Western Pacific region reported receiving pocket money. *Parental Smoking* is a binary indicator equal to 1 if one or both parents smoke at the time of the interview. Parental smoking is most common in Europe, where 60% of the surveyed teens have a parent who smokes, followed closely by the Western Pacific. It is perhaps not a coincidence but rather an example of the income effect that parents who are more likely to provide pocket money to their children (as those in Europe and the Western Pacific) are also more likely to smoke.

Environmental characteristics

Variables that describe the local environment of each subject can vary by survey site and over time. These include the level of anti-smoking sentiment (*Sentiment*), the prevalence of cigarette advertising (*Cigarette Advertising*), the prevalence of anti-tobacco media messages (*Anti-tobacco Media*), and the observed effectiveness of minimum-age tobacco purchase policies (*Youth Access Restrictions*). All of these are constructed from individual survey responses which are then used to produce aggregate measures on the site level.

Anti-smoking sentiment has been recognized in the domestic literature as an important predictor of local cultural attitudes toward smoking and of the smoking pattern itself. Omitting anti-smoking sentiment from a model of smoking demand can be problematic when its effect remains in the error term, causing the error term to be correlated with both smoking status and cigarette prices. In this paper *Anti-Smoking Sentiment* is defined as the percentage of non-smokers in the survey who favor bans on smoking in public places. We base this measure on non-smokers only (as opposed to all survey participants including smokers) in order to eliminate the potential for endogeneity bias when smokers' attitudes are included. In the case of smokers, it is not clear if sentiment affects smoking or smoking affects sentiment, so sentiment would be endogenous to smoking. Excluding the attitudes of the smokers from the measure of anti-smoking sentiment helps ensure that the relationship between sentiment and smoking is one-directional and that the sentiment variable is properly exogenous.

The most smoking-friendly attitude is observed in the African region where only 61 percent of the survey participants think smoking should not be allowed in public, followed by the Western Pacific region, where, for instance, only 39 percent of Phillippino participants have the same opinion. Smoking is viewed much more negatively in the Middle East, where for example 95 percent of surveyed youth in Pakistan are against public smoking.

Although excluding smokers from the calculation of *Anti-Smoking Sentiment* is expected to minimize the danger of endogeneity of the sentiment variable, we further address this concern by employing an alternative proxy for sentiment which is not derived from self-reported attitudes. This alternative proxy is the strength of the domestic tobacco industry (*Tobacco Production*). The presumption is that countries with higher tobacco production may view smoking more favorably. In this paper, the sentiment proxy is the annual production of tobacco leaf in tons per capita for each country, obtained from the Food and Agriculture Organization FAOSTAT database. Surprisingly, the simple correlation between the original *Anti-Smoking Sentiment* variable derived from GYTS and its proxy *Tobacco Production* has a positive sign, indicating that youth in countries with larger tobacco industries have a less permissive attitude toward smoking. This is generally not the case in the US where states with heavier tobacco production have lower cigarette taxation, pointing to lower anti-smoking sentiment. One possible explanation is that developing countries with more productive tobacco industries are more productive in general, have a higher income, a more educated population, and are more enlightened about the dangers of smoking. Regardless of the actual sign of the relationship between *Tobacco Production* and *Anti-Smoking Sentiment*, we will employ *Tobacco Production* as a proxy for sentiment with the assumption that the two are sufficiently correlated.

Another important site-specific variable in this analysis is the prevalence of cigarette advertising. *Cigarette Advertising* is determined by the proportion of survey participants who have been recently exposed to cigarette ads on billboards, newspapers or magazines. It provides an estimate of the likelihood of exposure to print media advertising and contains information on how effective local advertising is in reaching an audience and encouraging smoking. The heaviest exposure to cigarette advertising is observed in Poland, Indonesia, and Argentina, where almost all participants (96%) had recently seen print media cigarette promotions. The high advertising exposure in Poland is surprising given the existence of a complete ban on cigarette advertising there, and illustrates the disparity between policy presence and policy compliance in some countries. Least exposed to cigarette advertising are teens in Turkey (46%) but the average advertising exposure rate for the whole sample is fairly high at 86 percent.

The prevalence of anti-tobacco media messages can be interpreted as a proxy for the enthusiasm of local efforts to reduce smoking. *Anti-Tobacco Media* is determined by the proportion of responders who have been recently exposed to anti-smoking messages in broadcast and print media. Anti-smoking messages reach the least number of teens in Africa (56% in Cote D'Ivoire, 72% average for the region), and the highest number of teens in Europe (100% in Greece, Hungary, Ukraine, and Kazakhstan, 94% average for the region). It must be noted that a possible disadvantage associated with this variable is its potential correlation with both anti-smoking sentiment and price, as countries with lower tolerance for smoking may be more active in anti-tobacco campaigns and may impose more aggressive cigarette taxes. This concern is addressed in greater detail in Section 5, where we look for evidence of multicollinearity. Although *Anti-Tobacco Media* passes the conventional rule of thumb for sufficiently low collinearity, we also provide comparison estimates from specifications that exclude this measure.

Finally, the observed effectiveness of policies against cigarette sales to minors (*Youth Access Restrictions*) controls for ease of access to cigarettes, and is calculated as the proportion of survey participants who recently tried to buy cigarettes but were turned away by vendors due to age. Refusals to sell to minors occurred least frequently in Greece, where sales to minors are not regulated and only 10 percent of teen buyers reported difficulty buying cigarettes. Another country with relatively unhindered youth access to cigarettes is Turkey where only 15% of attempted underage buyers report being rejected by cigarette vendors. However Turkey, unlike Greece, does have a minors sales ban and is therefore yet another example of an ineffective policy with low compliance. The highest level of compliance with the minors sales ban is in South Korea where 72% of attempted underage buyers report difficulty buying cigarettes. For the sample as a whole, the average proportion of minors unable to buy cigarettes is 35%, which indicates that youth access restrictions have relatively weak enforcement in lower-income countries.

It is important to highlight the fact that the last three variables discussed here (*Cigarette Advertising*, *Anti-Tobacco Media*, and *Youth Access Restrictions*) are constructed from individual response data and consequently represent levels of policy effectiveness or policy compliance as opposed to simply indicating the existence of a related policy. This is an important distinction since the nominal presence of smoking policies like advertising bans or minors sale bans does not provide information on how well these policies are enforced in different countries. Using variables that describe levels of policy effectiveness is a considerable methodological improvement over the usual binary policy indicator variables.

Since different countries provide data from different years, it is necessary to account for a secular time effect that may influence smoking. In particular, attitudes toward smoking may change independently over time as more health information becomes available and/or more schools in developing countries implement anti-tobacco education. The independent effect of different time periods is accounted by a dummy variable for each year in the dataset.

As it frequently happens with individual-level data obtained from surveys, multiple observations are missing due to non-response or absent questionnaire parts. This is a nontrivial problem since missing observations from four major individually descriptive variables – *Age*, *Male*, *Parental Smoking*, and *Pocket Money* – add up to 20 percent of the total number of observations. Out of these, *Pocket Money* is missing most frequently due to absence of a related question in the survey questionnaires for some countries and years. Since we cannot assume that these observations are "missing completely at random", excluding them may lead to estimation bias. We may assume, however, that the missing observations can be classified as "missing at random", meaning that they can be explained by available data and therefore imputed. This is especially obvious in the case of the pocket money variable, where most missing values can be explained by country and year. We use the method of iterative imputation to fill in missing observations for *Age*, *Male*, *Parental Smoking*, and *Pocket Money*. This method has been recognized to have advantage over alternatives such as substitution of missing values by the sample mean or regression methods, both of which can lead to underestimation of the standard errors and erroneously significant results (Schafer & Olsen 1998).

4. Methods

Since roughly 90% of the survey participants in the dataset are current nonsmokers, a defining characteristic of the dataset is the prevalence of zero outcomes. I consider three types of econometric models, all of which reflect the skewness of the data toward zero. These are the two-part model, the Tobit model, and the zero-inflated Poisson model. This section discusses each of these models, concluding that the two-part model is the preferred choice.

4.1. Tobit

When a large proportion of the observed outcomes is zero, linear models are inappropriate due to the possibility of obtaining negative predictions and due to the assumption of constant partial effects over the whole range of outcomes. Tobit improves the linear model by allowing the marginal effects to differ by outcome and by precluding the prediction of negative outcomes.

The Tobit model expresses the outcome y as

$$y=\max(0, X\beta+u)$$

where X is a vector of explanatory variables, β is the parameter vector, and the error $u \sim N(0,\sigma^2)$.

If y takes on a strictly positive value with probability P, then $P(y>0) = P(X\beta+u) = \Phi(X\beta/\sigma)$ where Φ is the standard normal cdf. The strictly positive values of y come from a normal distribution with density ϕ.

If d_i is an indicator variable denoting observation i as positive vs. zero, the parameters β and σ in the above model can be estimated by maximum likelihood using the following log-likelihood function:

$$L = \sum_i \left[(1-d_i)\ln P(y_i = 0) + d_i \ln \phi(y_i \mid y_i > 0)\right] =$$

$$= \sum_i \left[(1-d_i)\ln\left(1-\Phi\left(\frac{X_i\beta}{\sigma}\right)\right) + d_i \ln\left(\frac{1}{\Phi\left(\frac{X_i\beta}{\sigma}\right)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - X_i\beta)^2}{2\sigma^2}\right)\right)\right] \quad (1)$$

While the Tobit model is an improvement over ordinary linear estimation, it has a significant drawback in assuming that both the positive outcomes (how much to smoke) and the zero outcomes (whether to smoke) come from the same decision making process (the normal distribution). This forces the parameters to have the same sign and the same relative magnitude in the data generating processes of both the zero outcomes and the strictly positive outcomes. To illustrate how this could be unnecessarily restrictive, consider the relationship between cigarette smoking and age. All things equal, older individuals are less likely to participate in or initiate smoking. However, among those who already participate in smoking, older persons may be the heavier smokers, perhaps due to increased habit or addiction. This possibility is not allowed in a Tobit model, where any variable which increases the probability of a positive outcome $P(y>0)$ must also increase the mean of the positive outcomes $E(y|y>0)$. In addition, the Tobit model does not allow variation in the relative strengths of the parameters between the two decision processes. In other words, the marginal effect of age relative to the effect of, say, price, would have to be the same in the estimations of both the decision to smoke and the subsequent decision of how much to smoke.

Due to the restrictive nature of Tobit, it is not entirely appropriate for modeling some zero-dominant datasets. A better alternative is the two-part (hurdle) model, described below.

4.2. Two-part model

The two-part model can be viewed as an extension of the Tobit model which relaxes the assumption that the same mechanism governs both the decision to smoke and the decision how much to smoke. In this model, these decision processes are independent (conditional on the explanatory variables) and can be determined by different factors if desired.

In the two-part model, the zeros can come from a standard normal or logistic density while the positives can come from a separate density function which can be, for example, normal, log-normal, or Poisson. Let us first consider the case where the probability of a zero outcome is generated by a standard normal, $P(y=0) = 1 - \Phi(X\beta)$, and the positive outcomes are generated by a normal distribution with pdf $\phi(y|y>0)$. The log-likelihood function looks very similar to Tobit, but notice that the parameters β in the density for the positives are allowed to be different from the parameters γ in the density for the zeros:

$$
\begin{aligned}
L &= \sum_i \left[(1-d_i)\ln P(y_i = 0) + d_i \ln \phi(y_i \mid y_i > 0)\right] = \\
&= \sum_i \left[(1-d_i)\ln\left(1-\Phi\left(X_i\beta\right)\right) + d_i \ln\left(\frac{1}{\Phi\left(X_i\beta\right)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - X_i\gamma)^2}{2\sigma^2} \right)\right)\right]
\end{aligned}
\qquad (2)
$$

If the parameter vector β from the zero-generating process equals the parameter vector γ/σ from the positives-generating process, the above log-likelihood function reduces to Tobit. It follows that Tobit can be described as a version of the two-part model under certain parameter restrictions and normal distributional assumptions.

We do not necessarily need to assign normal density to the positives-generating process. In fact, other distributions may be more suitable for positive-only values. One such distribution is the log-normal, where the positive outcomes are assigned normal density after a logarithmic transformation. Using log-normal density in the second part of the log-likelihood function, we can estimate the parameters β and γ from maximizing the following:

$$L = \sum_i \left[(1-d_i)\ln P(y_i = 0) + d_i \ln \phi(y_i \mid y_i > 0)\right] = \\ = \sum_i \left[(1-d_i)\ln\left(1-\Phi\left(X_i\beta\right)\right) + d_i \ln\left(\frac{1}{\Phi\left(X_i\beta\right)} \frac{1}{y_i\sqrt{2\pi}\sigma} \exp\left(-\frac{(\ln(y_i) - X_i\gamma)^2}{2\sigma^2} \right)\right)\right] \quad (3)$$

Yet another possible distribution for the strictly positive outcomes is Poisson with conditional pdf $f(y|y>0)$, which more aptly represents count values. Substituting Poisson density in the log-likelihood function, we get

$$L = \sum_i \left[(1-d_i)\ln P(y_i = 0) + d_i \ln f(y_i \mid y_i > 0)\right] = \\ = \sum_i \left[(1-d_i)\ln\left(1-\Phi\left(X_i\beta\right)\right) + d_i \ln\left(\frac{1}{\Phi\left(X_i\beta\right)} \frac{\exp(-e^{X_i\gamma})e^{X_i\gamma y_i}}{y_i!} \right)\right] \quad (4)$$

The contribution to the log-likelihood of an individual who is not observed to smoke is the probability of not smoking. The contribution of an individual who is a current smoker is the probability density of smoking a positive amount. In other words, the log-likelihood function of the two-part model has two distinct parts – one determining whether the outcome is zero, the other determining the distribution of the positive outcomes. Each of these parts can be estimated with separate equations since each comes from a separate distribution. In this paper, the first part of the two-part model estimates the probability of smoking participation with a logit model. The second part estimates conditional cigarette demand as a generalized linear model with a log link.

4.3. Zero-inflated Poisson (ZIP)

An alternative to the two-part model is the ZIP model, which is similar but slightly different than the two-part Poisson-based model discussed earlier. Although ZIP can have similar distributional assumptions to the two-part model (standard normal for the binary process and Poisson for the count process), it is different in that the count/Poisson part of the model is not restricted to generating positive values only, but can also generate zero outcomes. In other words, a person defined as a current smoker is not constrained to smoking strictly positive amounts of cigarettes but can also potentially choose to smoke zero cigarettes. This specification can provide interesting insights if we would like to know whether fluctuations in prices could induce a smoker into zero consumption. The drawback is that even though ZIP allows a smoker to choose zero consumption, it does not allow an independent process to generate this choice separately from the choice of simply smoking less. This disadvantage is similar to that of Tobit because smokers are restricted to the same decision-making mechanism when choosing to smoke both zero and positive amounts.

In ZIP, as in the two-part model, the certain zero outcomes (the true non-smokers) occur with probability $P(y=0)$. However, even when the observed outcome is positive and the individual is currently a smoker, ZIP allows for the possibility that zero consumption could occur with probability 1- $P(y=0)$. Smokers' zero outcomes are generated together with their positive outcomes from the Poisson density. After substituting the expressions for normal cdf in the binary probability process and Poisson pdf for the count generation process, the following log-likelihood function is obtained:

$$
\begin{aligned}
L &= \sum_i \left[(1-d_i)\ln\left(P(y_i=0)+P(y_i>0)f(y_i=0)\right)+d_i\ln\left(P(y_i>0)f(y_i)\right)\right]= \\
&= \sum_i \left[(1-d_i)\ln\left(\left(1-\Phi(X_i\beta)\right)+\Phi(X_i\beta)\exp(-e^{X_i\beta})\right)+d_i\ln\left(\Phi(X_i\beta)\frac{\exp(-e^{X_i\gamma})e^{X_i\gamma y_i}}{y_i!}\right)\right]
\end{aligned} \quad (5)
$$

If inducing a smoker into zero consumption comes from a different mechanism than inducing him to merely decrease consumption, ZIP is not appropriate. It is therefore more suitable to model smoking participation as a separate process from smoking intensity using the two-part model. This allows the estimation of price responsiveness among smokers to take place without possible interference from the effects of smokers switching smoking status, while the "switching" effect is still being properly accounted for by the first part of the model.

5. Empirical application and identification concerns

Since roughly 90% of the survey participants in our sample are current nonsmokers, a defining characteristic of the dataset is the prevalence of zero outcomes. The type of model that best reflects the skewness of the data toward zero in this particular application is the two-part (hurdle) model. This model relaxes the assumption that the same mechanism must govern both the decision to smoke and the decision of how much to smoke. These decision processes are assumed to be independent (conditional on the explanatory variables) and can be determined by different factors if desired. In this paper, the first part of the two-part model estimates the probability of smoking participation with a logit model. Following Tauras (2006), the second part of the two-part model estimates the amount of cigarettes smoked by smokers through a generalized linear model where the outcome has normal distribution and the link function is logarithmic. In general notation, the second-part GLM model can be expressed as $g(E(y)) = x\beta$ where the link function $g(.) = \ln(.)$ and $y \sim$ Normal.

Specifying a normal distribution in a log-link GLM is similar to but not equivalent to an ordinary least squares (OLS) regression on $\ln(y)$ because it produces more consistent and less biased elasticity estimates in the presence of heteroskedasticity (Manning and Mullahy 2001, Mullahy 1998, Tauras 2005, 2006).[1] Tauras (2005) estimates that the bias from using OLS instead of GLM in the estimation of conditional cigarette demand for U.S. adults can be substantial and can result in more-than-double overestimation of price elasticity. We use the GLM framework in an effort to avoid overestimating the price effect.

[1] For additional discussion on GLM versus OLS see Appendix 1.

In the first part of the two-part model of cigarette demand, smoking participation or the probability of consuming a positive number of cigarettes Y is modeled as a function of cigarette price, individual characteristics X_1, observed environmental characteristics X_2, year fixed effects *Year* and country fixed effects *Country*:

$$\Pr(Y_{ijt} > 0) = f(\alpha_0 + \alpha_1 Price_{jt} + \alpha_2 X_{1ijt} + \alpha_3 X_{2jt} + \alpha_4 Year_t + \alpha_5 Country_j) \quad (1)$$

In the second part of the two-part model, cigarette demand conditional on participation is expressed as

$$(Y_{ijt} | Y_{ijt} > 0) = f(\beta_0 + \beta_1 Price_{jt} + \beta_2 X_{1ijt} + \beta_3 X_{2jt} + \beta_4 Year_t + \beta_5 Country_j) \quad (2)$$

Alternatively, one-part models like the tobit or zero-inflated poisson models which do not distinguish between the decision to smoke and the decision how much to smoke can be expressed as

$$Y_{ijt} = f(\beta_0 + \beta_1 Price_{jt} + \beta_2 X_{1ijt} + \beta_3 X_{2jt} + \beta_4 Year_t + \beta_5 Country_j) \quad (3)$$

where i denotes individual, j denotes country/geographic location, and t denotes year. X_1 is a vector of individual-level variables which include *Age*, *Male*, *Parental Smoking*, and *Pocket Money*. X_2 is a vector of location-specific characteristics which include the *Cigarette Advertising*, *Anti-Tobacco Media*, *Anti-Smoking Sentiment*, and *Youth Access Restrictions*. Summary statistics and descriptions of all variables are listed in Tables 2-4.

The final results from the estimation of Equations 1, 2, and 3 are presented in Section 6. Potential econometric issues are discussed below.

5.1 Multicollinearity

One concern is that the five policy-relevant variables in this analysis, namely *Price*, *Anti-Smoking Sentiment*, *Cigarette Advertising*, *Anti-Tobacco Media*, and *Youth Access Restrictions*, may be correlated. The presence of multicollinearity may cause us to misjudge the separate effects of these variables and underestimate their statistical significance. To check whether multicollinearity among a number of variables is a problem, a simple but effective diagnostic is the variance inflation factor (VIF). VIF is calculated for each suspect variable by regressing it on the remaining suspect variables and taking the inverse of one minus the R-squared from the regression. A common rule of thumb is that if the R-squared is close to 1 and VIF exceeds 10, the suspect variable is well explained by a linear function of the remaining suspect variables, and collinearity between them is highly likely (Chatterjee et al 2000). Table 5 contains the VIFs and the corresponding R-squared statistics for all five suspect variables. Although the VIF estimates in Table 5a are all much smaller than 10 and therefore pass the conventional rule of thumb by a wide margin, it is noticeable that the anti-smoking sentiment variable has a relatively higher VIF. When the VIFs are recalculated after *Sentiment* is removed from the set of suspect variables (Table 5b), the remaining VIFs shrink even further. The conclusion is that there is evidence of some correlation between *Sentiment* and the rest of the policy-related variables but that this correlation may be small enough not to interfere with the estimation. However, in order to remove any doubt about validity problems from potential multicollinearity, we present results from specifications where *Sentiment* is both included and excluded. In either case, the results are not much different, providing evidence that the correlation between anti-smoking sentiment and policy-related country characteristics is minor.

5.2 Endogeneity of Price

Another concern about the identification of the price effect is that *Price* may be econometrically endogenous. Endogeneity can cause serious issues for the estimation process in terms of consistency and must be addressed before any results can be interpreted as causal. In this analysis, *Price* endogeneity may be expected to arise from a couple of sources, namely unobserved heterogeneity and simultaneity. Both of these are addressed as follows.

Unobserved heterogeneity bias

Unobserved heterogeneity refers to environmental characteristics that may influence both the cigarette price level and individual cigarette consumption. One example is the local cultural attitude toward smoking. A country with a predominantly unfavorable perception of smoking may impose more aggressive cigarette taxes and higher prices. At the same time, such a culture may also discourage young people from smoking. If the local cultural attitude is left out of the analysis, its effect may be absorbed by the price variable, leading to biased estimate. We address this possibility by including a direct control for local anti-smoking sentiment. It must be recognized, however, that *Anti-Smoking Sentiment* itself may suffer from a similar endogeneity problem if external factors, like media influences which could affect both smoking behavior and sentiment, are not accounted for. We address this possibility by including additional site-level controls for confounding media factors like exposure to cigarette advertising and anti-tobacco campaigns. While these controls are expected to reduce bias from unobserved heterogeneity, country fixed effects are added in order to deal with any remaining unobservables that do not vary over time.

Simultaneity bias

An additional concern about endogeneity of the *Price* variable may arise from the expectation that cigarette prices and cigarette demand are simultaneously determined. The use of micro-level data in this study considerably reduces the danger of such endogeneity because the smoking decision of a single individual could not affect market demand enough to change the price level. Certain characteristics of the local market demand, however, *can* influence the individual smoking decision by affecting the price level. For example, a weak market demand for cigarettes corresponds to higher cigarette prices, which in turn discourages individual smoking. Since market demand can affect both individual smoking decisions and prices, it can present another source of *Price* endogeneity in the form of unobserved market characteristics. Note that this is a different source of bias than simultaneity, and we can account for it by including country fixed effects, which can be interpreted as market fixed effects because cigarette prices have country-level variation.

Although using micro-level sufficiently reduces the simultaneity bias in the estimate of *Price*, we go a step further in investigating this possibility by substituting the price of Imported cigarettes instead of local brand cigarettes. The advantage of local-brand prices is that they are more likely to be considered when the average individual decides to consume cigarettes because local brand cigarettes are typically less expensive. Even though local-brand prices may be more relevant for individual purchasing decisions, imported-brand prices are likely to be more exogenous to cigarette demand because they are imported. This is because the price of imported cigarettes contains a larger exogenous (not determined by market demand) component such as transportation costs and import duties. The larger exogenous component makes imported-brand prices stickier and less vulnerable to changes in market demand than local-brand prices. We supply results from specifications using imported-brand prices as robustness checks and find that neither the significance nor the size of the price effect is reduced by switching to imported-brand prices. This provides evidence that the risk of simultaneity bias in the *Price* estimate is low.

5.3. Country vs. Region Fixed Effects

The use of fixed effects in combination with multiple relevant controls in the empirical estimation of smoking patterns is an effective method of reducing endogeneity bias and helps provide plausible evidence for the presence (or absence) of causal effects.

Two types of fixed effects are alternatively examined in this research: country fixed effects and region fixed effects. The advantage of using country-level as opposed to region-level fixed effects is narrower geographic definition and hence, the possibility of capturing more information on local unobservables. The disadvantage is that the dataset needs to be restricted only to countries which have survey data from multiple years (because the price effect is identified from variability in prices within country over time). Since GYTS was conducted only once in many countries, the country fixed-effects dataset would have to be reduced to 21 countries from the original 47. Using region fixed effects for each of the six world regions avoids this data loss problem in return to assuming that countries within the same region have similar unobserved characteristics. There is evidence that supports this assumption and suggests that the use of region fixed effects instead of country fixed effects may be acceptable due to lack of very strong variation between countries in the same region. To illustrate this point, Table 6a compares the within-country correlation of individual cigarette demand (0.066) to the within-region correlation (0.045). Both of these estimates are very similar to each other, indicating that individual demand patterns are similarly correlated across individuals within each region as they are across individuals within separate countries. Table 6b provides additional macro-level comparisons of within-region cluster correlations of smoking patterns and their major predictors. On a macro level, the within-region correlations are high, and especially so for defining country characteristics such as smoking prevalence (32%), conditional cigarette demand (46%), anti-smoking sentiment (46%), and the prevalence of cigarette advertising (52%). This points to substantial similarities among countries in the same region in terms of smoking patterns, cultural attitudes, and media exposure, and suggests that other relevant unobserved country-level characteristics are also likely to be shared within a region. In such case, region fixed effects would capture a similar amount of information as country fixed effects but with a smaller loss of degrees of freedom.

One particular example of within-region similarities is the level of cigarette consumption among smokers. In Europe, smokers consume around 100 cigarettes per month regardless of country. In contrast, Southeast Asian smokers consume twice as few cigarettes, and even the highest consumption in Southeast Asia (68 cigarettes per month in Bangladesh) remains below the lowest consumption in Europe (81 cigarettes per month in Ukraine). Regional differences in anti-smoking sentiment provide another example of a case where similarities between countries in the same region are strong while differences between the regions themselves can be considerable. Consider the Middle East region where anti-smoking sentiment in the sample ranges from 71% to 95% with an average of 84%. In comparison, anti-smoking sentiment in the African region is much lower on average at 61%. Table 6b shows that while the observable characteristics of countries within the same region may not be perfectly correlated, they can be shared to a considerable extent. This supports the assumption that unobservable county characteristics may be similarly shared, validating the use of region fixed effects as a satisfactory substitute for country fixed effects in the effort to account for unobserved country-level heterogeneity. Although our preferred specification employs country fixed effect, region fixed effects are used for robustness checks.

5.4 Clustering

Unobserved heterogeneity can present a major econometric problem even if it is not correlated with any of the observed predictors. This can happen when the unobserved disturbances are correlated within groups or clusters of observations. Survey data from different geographic locations is particularly vulnerable to such clustering because individuals within each country are likely to be correlated in some unknown way even if they are uncorrelated between countries. Uncontrolled clustering produces severely underestimated standard errors and spurious findings of statistical significance (Pepper 2002, Cameron et al 2006, Wooldridge 2003, Wooldridge 2002). Bertrand et al (2004) have shown that clustering can remain even after including state and year fixed effects and will lead to invalid inference if not controlled for. Moulton (1990) points out that the clustering issue is especially aggravated in cases where the groupings are used to merge macro variables with micro data in order to explain micro outcomes. This is particularly relevant for this application where country-level characteristics such as cigarette prices are used to predict individual-level smoking behaviors.

All models in this paper adjust the standard errors for clustering by survey site. The adjustment is similar to the Huber-White treatment of heteroskedastic errors but the variance scaling factor is the sum of the squared products of residuals and regressors within cluster (as opposed to only the squared products of residuals and regressors). This corrects for error correlations of unknown form within clusters.

To summarize, the main econometric issues in this research that may interfere with identification are potential multicollinearity and unobserved heterogeneity where the latter can be both correlated and uncorrelated with the predictors. We find no substantial evidence for multicollinearity so attention is mostly focused on the latter issue. Unobserved heterogeneity is treated by including country fixed effects and by reducing the omitted variable problem with a number of relevant controls. Any remaining site-level heterogeneity is addressed by allowing for cluster correlation of observations within geographic sites. Additional precision of the estimates is sought by the use of generalized linear modeling instead of OLS in the second part of the two-part model.

6. Results

6.1. Two-part model

Smoking participation

Table 7a contains results from the first part of the two-part model which estimates the probability of smoking participation by logit with country fixed effects using local-brand cigarette prices. The main conclusion from Table 7a is that cigarette price is a significant determinant of smoking participation along with cigarette advertising and youth access restrictions. There are eight specifications depending on which predictors are included. The baseline specification (Equation 1) looks at the effect of *Price* on participation without controlling for either anti-smoking sentiment or media effects. It provides a statistically significant estimate of price elasticity of -0.72, indicating that an increase in cigarette price of 10% would correspond to a 7.2% reduction in smoking prevalence. Equations 2, 3, and 4 contain different combinations of the media exposure variables but still exclude anti-smoking sentiment. Equations 5 through 8 correspond to Equations 1 through 4 but in addition include *Sentiment*.

In all specifications, *Price* has a significant negative effect on smoking participation. The estimated price elasticity of participation ranges from -0.56 to -0.88. *Sentiment* is shown to be a significant predictor of participation as well and has the expected negative sign, confirming that higher anti-smoking sentiment is indeed associated with lower participation. However, unlike DeCicca et al (2002, 2008) and more in line with Carpenter and Cook (2008) we find that *Sentiment* is not the most influential factor determining smoking participation. Although it is statistically significant, controlling for *Sentiment* does not reduce either the magnitude or the significance of the effect of *Price* on participation.

We find that a major determinant of smoking participation is *Cigarette Advertising*. The local prevalence of cigarette advertising increases the probability of participation, most likely through higher advertising exposure. We estimate that if cigarette advertising succeeded in reaching every single individual (so that the proportion of youth exposed to advertising approached 100% from the current mean of 86%), then the average smoking prevalence rate would increase by up to 1.8 percentage points, from 10% to almost 12%. In terms of elasticity, we estimate that the advertising elasticity of participation ranges from 1.1 to 1.9, implying that a 10% increase in the proportion of people who observe cigarette advertising is associated with up to 19% increase in the prevalence of smoking.

We also find that *Youth Access Restrictions* have a sizeable and robustly significant effect on smoking participation. If bans against selling cigarettes to youth were fully enforced (i.e., if the proportion of underage youth unable to buy cigarettes increased from the observed mean of 35% to 100%), then the probability of participation would go down by 6.5 percentage points based on the estimate from Specification 8. Given that the mean smoking prevalence in our sample is 10%, full enforcement of youth access restrictions would cut participation rates by more than half. This finding has direct policy relevance because it illustrates the importance of compliance with anti-tobacco policies and highlights the difference in outcomes between actual and desired policy effectiveness.

Most of the specifications where *Anti-Tobacco Media* is included provide evidence that increased likelihood of exposure to anti-smoking messages reduces smoking participation. This result is not entirely robust to different specifications but may indicate that anti-tobacco campaigns may have some effect in reducing the occurrence of smoking. Based on the estimate from Specification 8, if anti-tobacco campaigns had perfect outreach and the proportion of youth witnessing them increased to 100% from the current mean of 83%, smoking prevalence may decline by about 1.5 percentage points. It is interesting to note that this would almost wash out the effect of cigarette advertising. If cigarette advertising reached everyone, participation would grow by 1.8 points; if anti-tobacco messages reached everyone, participation would decline by 1.5 points, and the two media effects would almost neutralize each other at the mean.

Conditional cigarette demand

Results from the second part of the two-part model which estimates conditional cigarette demand are presented in Table 7b. The result that stands out from Table 7b is that *Price* is a significant predictor of conditional cigarette demand. The price elasticities of conditional demand are estimated in the range of -1.14 to -1.46. The price elasticity in Specification 8 is -1.2, indicating that a 10% increase in *Price* corresponds to a 12% decrease in the intensity of cigarette consumption.

There is no evidence that *Anti-Smoking Sentiment*, *Cigarette Advertising*, or *Youth Access Restrictions* can influence cigarette demand among current smokers. This leads us to believe that once the decision to smoke is made, not many factors besides cigarette prices can help explain how many cigarettes are smoked. One notable exception is *Anti-Tobacco Media*, which is shown to be a significant albeit a very small determinant of smoking intensity. *Anti-Tobacco Media* has a sample mean of 0.83, meaning that anti-tobacco messages reach 83% of the current smokers. If instead all smokers were exposed to anti-tobacco media, the conditional demand for cigarettes would drop by 0.22% or by less than half a cigarette per month per smoker.

Ordered logit estimates

To see how prices may affect different types of smokers, we use an ordered logit model of conditional cigarette demand with four smoker categories: very light smokers (1 – 15 cigarettes per month), light to medium smokers (15 to 100 cigarettes per month), medium smokers (100 to 300 cigarettes per month) and heavy smokers (over 300 cigarettes per month). Table 7c lists the price responsiveness of the probability of being in each smoker category. The results are similar and significant across all eight specifications. Taking Specification 8 as an example, the estimates imply that increasing price by 10% decreases the probability of being a heavy smoker by 8.7%, decreases the probability of being a medium smoker by about 6.9%, decreases the probability of being a light to medium smoker by 3.4%, and increases the probability of being a very light smoker by 4%. These estimates show that higher prices progressively reduce the intensity of smoking for all but the lightest smokers and increase the likelihood of smokers switching down to a lighter smoker status.

The results from the two-part model with country fixed effects and local-brand cigarette prices can be summarized as follows. Price is a major determinant of both smoking participation and conditional cigarette demand and in addition seems to be the only major predictor of conditional demand. Smoking participation is responsive to more factors besides prices and can be influenced by anti-smoking sentiment, youth access restrictions, anti-tobacco media, and cigarette advertising.

6.2. Sensitivity checks on the two-part model

6.2.1. Using imported-brand prices instead of local-brand cigarette prices

Local-brand cigarette prices are preferred over imported-brand prices because they are cheaper and therefore more relevant to the consumption decision of the average individual. The concern that *Price* may be simultaneously determined with consumption is alleviated by using micro-level data. To provide further reassurance that the *Price* estimate is not likely to contain a simultaneity bias we provide specifications where *Price* is defined as the price of imported-brand cigarettes. Imported-brand prices are more exogenous to cigarette demand than local-brand prices because they are more likely to contain a larger market-unrelated component like transportation costs and import duties. Results from specifications using imported prices are listed in Table 8. The results are very similar to the original models with local-brand cigarette prices. In fact, the elasticity of participation and demand with respect to imported-brand prices are even larger in magnitude. If a simultaneity bias is present in the original local-brand price elasticity estimate, using a presumably more exogenous price variable would lead to a reduction in the price elasticity. Since no such reduction is evident, we can safely assume that the risk of simultaneity bias in the original local-brand price estimate is minor.

6.2.2. Using Tobacco production per capita as a proxy for anti-smoking sentiment

The purpose of employing *Tobacco Production* as a proxy for sentiment is to check the robustness of the results to using a sentiment variable that is not derived from self-reported attitudes. Although endogeneity of the original *Anti-Smoking Sentiment* variable is minimized by excluding smokers from the sentiment calculation, *Tobacco Production* can be substituted for a sensitivity check. The resulting estimates are reported in Table 8. In discussing the results, it must be noted that *Tobacco Production* is not an ideal proxy for sentiment. Although the correlation between *Tobacco Production* and *Anti-Smoking Sentiment* is not trivial (0.08), it does not have the expected negative sign. *Tobacco Production* may also be subject to a sizeable measurement error because some of the countries assumed to have zero tobacco production may have positive but unreported production. With these limitations in mind, we find that the results mostly are not sensitive to how sentiment is defined, with a few exceptions. First, unlike *Anti-Smoking Sentiment* in the original setup, *Tobacco Production* is not a significant factor in determining smoking participation. Second, unlike *Anti-Smoking Sentiment*, controlling for *Tobacco Production* eliminates the statistical significance of the price effect on participation in some (but not all) specifications. Although the price elasticity of participation is somewhat sensitive to how sentiment is defined, the price elasticity of conditional demand is robust in terms of magnitude as well as statistical significance.

6.2.3. Using region fixed effects instead of country fixed effects

Under the assumption that region FE can provide similar information about unobserved area characteristics as country FE, region FE have the advantage of retaining a larger number of countries from the GYTS sample. Tables 8ab contains results from specifications which are the same as the original framework in Tables 7ab except for employing region FE instead of country FE. We find that the results are similar but not entirely robust to the type of fixed effects. First, controlling for *Cigarette Advertising* reduces the statistical significance of the price effect on participation in models with region FE but not in models with country FE. Second, *Youth Access Restrictions* lose their effect on participation but gain a significant effect on conditional demand when region FE are employed. Third, the price elasticity of conditional demand shrinks in magnitude when region FE are used even though it remains statistically significant. It is unclear why region FE, which are not nearly as geographically specific as country FE, can lead to 1) insignificant price effects where significant price effects have already been identified with country FE (in the modeling of participation) or 2) reduce the magnitude of the price effects where larger price effects have already been identified with country FE (in the modeling of conditional demand). This is contrary to expectation because in the case of unobserved heterogeneity bias, region FE are less precise in capturing area heterogeneity and are therefore less likely to reduce bias than country FE. Since region FE do not behave as expected in identifying the price effects when compared to country FE, this may be taken as a sign that region FE are a questionable substitute for country FE.

6.3. One-part models: Tobit and Zero-inflated Poisson

Estimates from the Tobit model are presented in Table 11, estimates from the ZIP model are presented in Table 12. Similar to the original two-part framework in Tables 7ab, both define *Price* as local-brand cigarette price and control for country FE. Unlike the two-part model, the Tobit and ZIP models estimate the effect of the covariates on the total demand for cigarettes, not just the conditional (positive only) demand. Because of the fact that a two-part model consisting of a probit first part and a linear second part can be reduced to a Tobit if the parameters from the two parts were restricted to equal each other, we are able to apply a Hausman-type test of specification to the Tobit. This is the Ruud specification test (Ruud 1986) which is a regular likelihood-ratio test where the restricted model is the Tobit and the unrestricted model is the two-part model. Under the null hypothesis, the Tobit restrictions hold and the Tobit is correctly specified. The Ruud test statistics are listed at the bottom of Table 11. In all specifications, the null is strongly rejected, leading to the conclusion that a Tobit framework is not appropriate for this analysis.

Since the Tobit specification is incorrect, it is perhaps not surprising that it fails to detect significant effects for any of the covariates. This is true not only for all of the macro variables but also for important individual characteristics like age and sex. Even in the absence of a specific test for model validity, the inability of the Tobit to identify any covariate impacts may cause us to be suspicious about the framework.

Unlike the Tobit model, the ZIP model detects statistically significant effects for all covariates. In magnitude, the Tobit and ZIP marginal effects are similar and, taking Specification 8 as an example, suggest that a 10% increase in the price of cigarettes would correspond to a decrease in cigarette consumption by 0.44 cigarettes per month at the mean. In terms of size, this is a non-trivial marginal effect when we consider that the predicted average cigarette demand for the whole population (counting both smokers and non-smokers) is 2.9 cigarettes per month in ZIP and 4.5 cigarettes per month in Tobit. The total price elasticity of demand in specification 8 is -1.53 for ZIP and -0.97 for Tobit. These estimates are similar but somewhat lower than the total price elasticity of demand from the two-part model in Table 7, which is -1.83 as calculated by the sum of the elasticities of participation and conditional demand.

7. Conclusion

The contribution of this research is to provide insight into the factors that shape cigarette consumption among youth in developing countries. Besides estimating price elasticities of demand, we are the first to offer a thorough examination of multiple environmental aspects that may affect smoking, including cigarette advertising, anti-tobacco media campaigns, and the observed effectiveness of youth access restrictions. Although other papers have looked at the effect of smoking and advertising bans, we are able to extend our analysis beyond the nominal presence of smoking-related policies and are able to control for the observed effectiveness of such policies.

This research has multiple policy implications. It confirms the importance of cigarette prices in determining both smoking participation and conditional cigarette demand. We estimate that the price elasticity of smoking participation is -0.6 while the price elasticity of conditional demand is -1.2. We find that anti-smoking sentiment, cigarette advertising, and youth access restrictions influence the decision to participate in smoking but not the intensity of cigarette consumption among current smokers. We estimate that perfect compliance with youth access restrictions may cut smoking participation by more than half. We also show that anti-tobacco media campaigns may be effective in reducing both participation and intensity.

Appendix

The difference between a log-link GLM with normal distribution and an OLS model of ln(y) is that the former estimates ln(E(y)) while the latter estimates E(ln(y)). In other words, when linearizing the relationship between outcome y and covariates x, the log-link GLM model takes the log of the linear predictor xb while the OLS model takes the log of the outcome y.

In the second part of the two-part model, conditional cigarette demand is expressed as

$$y = e^{x\beta}e^{u} \tag{1}$$

Traditionally, OLS is applied to the log transformation of the outcome y, so that

$$\ln(y)=x\beta + u, \text{ and} \tag{2}$$

$$E(\ln(y))=xb \text{ under standard assumptions.} \tag{3}$$

However, recall that the formula for price elasticity is $\frac{\partial \ln E(y)}{\partial \ln p}$, not $\frac{\partial E(\ln(y))}{\partial \ln p}$.

Therefore for the calculation of price elasticity we need to know ln(E(y)), not E(ln(y)) as provided by OLS in (3). To obtain ln(E(y)), we first need to recover E(y). From (1),

$$\begin{aligned} E(y) &= E(e^{x\beta}e^{u}) \\ &= e^{xb}E(e^{u}) \end{aligned} \tag{4}$$

where

$$E(e^{u}) = \begin{cases} \rho & \text{if the error } u \text{ is homoskedastic} \\ \rho(p) & \text{if the error } u \text{ is heteroskedastic in price } p \end{cases}$$

Taking the log of both sides in (4):

$$\ln(E(y)) = \begin{cases} xb + \rho & \text{under homoskedasticity} \\ xb + \rho(p) & \text{under heteroskedasticity in price } p \end{cases} \quad (5)$$

Substituting (5) into the formula for price elasticity:

$$\text{Price elasticity} = \frac{\partial \ln E(y)}{\partial \ln p} = \begin{cases} \dfrac{\partial xb}{\partial p} & \text{under homoskedasticity} \\ \dfrac{\partial xb}{\partial p} + \dfrac{\partial \rho(p)}{\partial p} & \text{under heteroskedasticity in price } p \end{cases} \quad (6)$$

Equation (6) illustrates two points. First, if the error term is unrelated to price (is homoskedastic), price elasticity can be estimated consistently by OLS. Second, if heteroskedasticity in price is present, the OLS-estimated price elasticity will be biased by the amount $\frac{\partial \rho(p)}{\partial p}$. Estimation of *y* by log-link GLM avoids the possibility of bias by directly estimating ln(*E*(*y*)). It precludes the need for transforming the OLS-estimated *E*(ln(*y*)) into *E*(*y*) and then into ln(*E*(*y*)) when estimating price elasticities. By avoiding this transformation, it also avoids incurring the heteroskedasity-related bias described here.

Bibliography

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics* 119: 249-275.

Cameron, Colin A., Jonah B. Gelbach, and Douglas L. Miller. 2006. “Robust Inference with Multi-Way Clustering.” NBER Technical Working Papers: 327.

Carpenter, Christopher, and Philip J. Cook. 2008. “Cigarette Taxes and Youth Smoking: New Evidence from National, State, and Local Youth Risk Behavior Surveys.” *Journal of Health Economics* 27: 287-299.

Chaloupka, Frank J., and Michael Grossman. 1996. "Price, Tobacco Control Policies and Youth Smoking." NBER Working Papers: 5740.

Chaloupka, Frank J., and Henry Wechsler. 1995. "Price, Tobacco Control Policies and Smoking among Young Adults." NBER Working Papers: 5012.

Chaloupka, Frank J., and Henry Wechsler. 1997. "Price, Tobacco Control Policies and Smoking among Young Adults." *Journal of Health Economic* 16: 359-373.

Chapman, S., and J. Richardson. 1990. “Tobacco Excise and Declining Consumption: The Case of Papua New Guinea.” *American Journal of Public Health* 80: 537-540.

Chatterjee, S., A. S. Hadi, and B. Price. 2000. Regression Analysis by Example. 3rd ed. New York: Wiley.

Czart, Christina, Rosalie Ricardo Pacula, Frank J. Chaloupka, and Henry Wechsler. 2001. "The Impact of Prices and Control Policies on Cigarette Smoking among College Students." *Contemporary Economic Policy* 19(2): 135-149.

DeCicca, Philip, Don Kenkel, and Alan Mathios. 2008. "Cigarette Taxes and the Transition from Youth to Adult Smoking: Smoking Initiation, Cessation, and Participation." *Journal of Health Economics* 27(4): 904-917.

DeCicca, Philip, Donald Kenkel, and Alan Mathios. 2002. "Putting Out the Fires: Will Higher Taxes Reduce the Onset of Youth Smoking?." *Journal of Political Economy* 110(1): 144-169.

Decicca, Philip, Donald Kenkel, Alan Mathios, Yoon-Jeong Shin, and Jae-Young Lim. 2008. "Youth Smoking, Cigarette Prices, and Anti-smoking Sentiment." *Health Economics* 17(6): 733-749.

DeCicca, Philip, and Logan McLeod. 2008. "Cigarette Taxes and Older Adult Smoking: Evidence from Recent Large Tax Increases." *Journal of Health Economics* 27(4): 918-929.

Emery, Sherry, Martha M. White, and John P. Pierce. 2001. "Does Cigarette Price Influence Adolescent Experimentation?" *Journal of Health Economics* 20(2): 261-270.

Farrelly, Matthew C., Terry F. Pechacek, and Frank J. Chaloupka. 2003. "The Impact of Tobacco Control Program Expenditures on Aggregate Cigarette Sales: 1981-2000." *Journal of Health Economics* 22(5): 843-859.

Hammar, Henrik, and Fredrik Carlsson. 2005. "Smokers' Expectations to Quit Smoking." *Health Economics* 14(3): 257-267.

Harris, Jeffrey E., and Sandra W. Chan. 1999. "The Continuum-of-Addiction: Cigarette Smoking in Relation to Price among Americans Aged 15-29." *Health Economics* 8(1): 81-86.

Keeler, Theodore E. et al. 2001. "State Tobacco Taxation, Education and Smoking: Controlling for the Effects of Omitted Variables." *National Tax Journal* 54(1): 83-102.

Kooreman, Peter. 1986. "Alternative Specification Tests for Tobit and Related Models." *Economics Letters* 21: 31-33.

Lance, Peter M._et al. 2004. "Is Cigarette Smoking in Poorer Nations Highly Sensitive to Price? Evidence from Russia and China." *Journal of Health Economics* 23(1): 173-189.

Lewit, Eugene M. and Douglas Coate. 1982. "The Potential for Using Excise Taxes to Reduce Smoking." *Journal of Health Economics* 1: 121-145.

Lewit, Eugene M., Douglas Coate, and Michael Grossman. 1981. "The Effects of Government Regulation on Teenage Smoking." *Journal of Law and Economics* 24(3): 545-569.

Manning, W. G., and J. Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform." *Journal of Health Economics* 20(4): 461-94.

Moulton, B.R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* 72: 334-38.

Mullahy, J. 1998. "Much Ado about Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics." *Journal of Health Economics* 17(3): 247-81.

Nelson, Jon P. 2003. "Youth Smoking Prevalence in Developing Countries: Effect of Advertising Bans." *Applied Economics Letters* 10(13): 805-811.

Nicolas, Angel Lopez. 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 Economics* 11(6): 521-535.

Pepper, John V. 2002. "Robust Inferences from Random Clustered Samples: An Application Using Data from the Panel Study of Income Dynamics." *Economics Letters* 75: 341-345.

Powell, Lisa M., John A. Tauras, and Hana Ross. 2005. "The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior." *Journal of Health Economics* 24(5): 950-968.

Ross, Hana, and Frank J. Chaloupka. 2004. "The Effect of Public Policies and Prices on Youth Smoking." *Southern Economic Journal* 70(4): 796-815.

Ross, Hana, and Frank J. Chaloupka. 2003. "The Effect of Cigarette Prices on Youth Smoking." *Health Economics* 12(3): 217-230.

Ross, Hana, Frank J. Chaloupka, and Melanie Wakefield. 2006. "Youth Smoking Uptake Progress: Price and Public Policy Effects." *Eastern Economic Journal* 32(2): 355-367.

Ross, Hana, Lisa M. Powell, John A. Tauras, and Frank J. Chaloupka. 2005. "New Evidence on Youth Smoking Behavior Based on Experimental Price Increases." *Contemporary Economic Policy* 23(2): 195-210.

Ruud, Paul. 1984. "Tests of Specification in Econometrics." *Econometric Reviews* 3: 211-276.

Schafer J, Olsen. 1998. M. Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research* 33(4); 545-571.

Tauras, John A. 2006. "Smoke-Free Air Laws, Cigarette Prices, and Adult Cigarette Demand." *Economic Inquiry* 44(2): 333-342.

Tauras, John, Sara Markowitz, and John Cawley. 2005. "Tobacco Control Policies and Youth Smoking: Evidence from a New Era" in Advances in Health Economics and Health Services Research, Vol. 16-Substance Use: Individual Behavior, Social Interactions, Markets and Politics, edited by Bjorn Lindgren and Michael Grossman, Elsevier, Oxford, England.

Tauras, John A. 2005. "An Empirical Analysis of Adult Cigarette Demand." *Eastern Economic Journal* 31(3): 361-375.

Tauras, John A., and Frank J. Chaloupka. 1999. "Determinants of Smoking Cessation: An Analysis of Young Adult Men and Women." NBER Working Papers: 7262.

Tauras, John A., and Frank J. Chaloupka. 1999. "Price, Clean Indoor Air, and Cigarette Smoking: Evidence from the Longitudinal Data for Young Adults." NBER Working Papers: 6937.

Tauras, John A., Patrick M. O'Malley, and Lloyd D. Johnston. 2001. "Effects of Price and Access Laws on Teenage Smoking Initiation: A National Longitudinal Analysis." NBER Working Papers: 8331.

Wasserman, Jeffrey_et al. 1991. "The Effects of Excise Taxes and Regulations on Cigarette Smoking." *Journal of Health Economics* 10(1): 43-64.

Wooldridge, Jeffrey M. 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press: Cambridge, MA.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review* 93(2): 133-138.

Table 1. List of countries, regions, and survey years

Region	Country	Years
Africa	Cote D'Ivoire	2003
	Kenya	2001
	Nigeria	2000
	Senegal	2002
	South Africa	1999, 2002
Africa		1999, 2000, 2001, 2002, 2003
Middle East	Bahrain	2002
	Egypt	2001, 2005
	Iran	2003
	Jordan	1999, 2003
	Kuwait	2001, 2005
	Morocco	2001, 2006
	Pakistan	2003, 2004
	Saudi Arabia	2001
	Tunisia	2001
	UAE	2002, 2005
Middle East		1999, 2001, 2002, 2003, 2004, 2005, 2006
Europe	Czech Republic	2002
	Greece	2005
	Hungary	2003
	Kazakhstan	2004
	Poland	1999, 2003
	Romania	2004
	Russia	2002, 2004
	Turkey	2003
	Ukraine	2005
Europe		1999, 2002, 2003, 2004, 2005
Americas	Argentina	2000
	Brazil	2002, 2004, 2005, 2006
	Chile	2000, 2003
	Colombia	2001
	Costa Rica	1999, 2002
	Guatemala	2002
	Mexico	2000, 2005, 2006
	Panama	2002
	Paraguay	2003
	Peru	2000, 2002, 2003
	Uruguay	2000
	Venezuela	1999, 2001, 2003
Americas		1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006
Southeast Asia	Bangladesh	2004
	India	2000, 2001, 2002, 2003, 2004, 2006
	Indonesia	2000, 2004, 2005, 2006

	Nepal	2004
	Sri Lanka	1999, 2003
	Thailand	2005
Southeast Asia		1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006
Western Pacific	China	1999, 2001, 2005
	Phillippines	2000, 2004
	Singapore	2000
	South Korea	2005
	Vietnam	2003
Western Pacific		1999, 2000, 2001, 2003, 2004, 2005

Table 2a. Descriptive statistics for country-level dataset.

Variable type	Variable name	Variable description	Full sample (N=349130)				Smokers only (N=33187)			
			mean	sd	min	max	mean	sd	min	max
Individual-level	Current Smoker	1 if smoked at least one cigarette in past month, 0 otherwise	0.1	0.29	0	1	1	0	1	1
	Cigarette Demand	Number of cigarettes smoked in past month	6.9	46.6	0	630	71.9	135	1.5	630
	Age	Age in years	14	1.4	8.6	19	14.5	1.5	9.7	19
	Male	1 if male, 0 otherwise	0.5	0.5	0	1	0.63	0.48	0	1
	Pocket Money	1 if receives pocket money/income, 0 otherwise	0.62	0.49	0	1	0.85	0.36	0	1
	Parental Smoking	1 if at least one parent smokes, 0 otherwise	0.46	0.5	0	1	0.63	0.48	0	1
		* 1 if supports public smoking bans, 0 otherwise	0.76	0.43	0	1	0.56	0.5	0	1
		* 1 if recently exposed to cigarette advertising in print media, 0 otherwise	0.86	0.35	0	1	0.9	0.3	0	1
		* 1 if recently exposed to anti-smoking media messages, 0 otherwise	0.81	0.39	0	1	0.81	0.39	0	1
		* 1 if tried to buy cigarettes but was turned away due to age	0.05	0.22	0	1	0.23	0.42	0	1
Site-level	Anti-Smoking Sentiment	% nonsmokers who support public smoking bans	0.83	0.11	0.4	0.96				

	Cigarette Advertising	% survey participants who report recent exposure to cigarette advertising in print media	0.88	0.09	0.44	0.99
	Anti-Tobacco Media	% survey participants who report recent exposure to anti-smoking media messages	0.83	0.07	0.61	1
	Youth Access Restrictions	% survey participants who report being unable to buy cigarettes due to age	0.37	0.18	0.05	0.87
Country-level		* Nominal price of local-brand cigarettes	1.03	0.34	0.54	1.53
		* Nominal price of Imported-brand cigarettes	1.43	0.58	0.67	2.92
	Price (local brand)	Real price of local brand cigarettes, PPP-adjusted, constant 2000 USD	2.4	0.84	1.12	4.68
	Price (Imported-brand)	Real price of imported cigarettes, PPP-adjusted, constant 2000 USD	3.39	1.64	1.45	8.94
	Tobacco Production	Tobacco leaf production, tons per capita	560	896	0	3740

* These variables are not used in any of the models but are displayed here for better sample description

Table 2b. Distribution of conditional cigarette demand for country-level dataset

	Number of cigarettes per month
Mean	71.9
Min	1.5
10th percentile	1.5
25th percentile	3.8
Median	14
75th percentile	85.8
90th percentile	240
Max	630
N	33187

Table 3a. Descriptive statistics for region-level dataset.

Variable type	Variable name	Variable description	Full sample (N=491660)				Smokers only (N=48097)			
			mean	sd	min	max	mean	sd	min	max
Individual-level	Current Smoker	1 if smoked at least one cigarette in past month, 0 otherwise	0.1	0.3	0	1	1	0	1	1
	Cigarette Demand	Number of cigarettes smoked in past month	7.6	49.2	0	630	77.1	139.3	1.5	630
	Age	Age in years	14	1.4	9	19	14.5	1.5	10	19
	Male	1 if male, 0 otherwise	0.5	0.5	0	1	0.63	0.48	0	1
	Pocket Money	1 if receives pocket money/income, 0 otherwise	0.63	0.48	0	1	0.86	0.35	0	1
	Parental Smoking	1 if at least one parent smokes, 0 otherwise	0.46	0.5	0	1	0.62	0.48	0	1
		* 1 if supports public smoking bans, 0 otherwise	0.78	0.41	0	1	0.56	0.5	0	1
		* 1 if recently exposed to cigarette advertising in print media, 0 otherwise	0.84	0.37	0	1	0.89	0.31	0	1
		* 1 if recently exposed to anti-smoking media messages, 0 otherwise	0.83	0.38	0	1	0.83	0.38	0	1
		* 1 if tried to buy cigarettes but was turned away due to age	0.05	0.22	0	1	0.24	0.43	0	1
Site-level	Anti-Smoking Sentiment	% nonsmokers who support public smoking bans	0.83	0.11	0.4	0.96				

	Cigarette Advertising	% survey participants who report recent exposure to cigarette advertising in print media	0.86	0.1	0.44	0.99
	Anti-Tobacco Media	% survey participants who report recent exposure to anti-smoking media messages	0.83	0.09	0.56	1
	Youth Access Restrictions	% survey participants who report being unable to buy cigarettes due to age	0.35	0.17	0.04	0.87
Country-level		* Nominal price of local-brand cigarettes	1.13	0.58	0.33	3.4
		* Nominal price of Imported-brand cigarettes	1.61	0.69	0.72	3.63
	Price (local brand)	Real price of local brand cigarettes, PPP-adjusted, constant 2000 USD	2.31	0.87	0.8	4.88
	Price (Imported-brand)	Real price of imported cigarettes, PPP-adjusted, constant 2000 USD	3.37	1.3	1.45	8.94
	Tobacco Production	Tobacco leaf production, tons per capita	823	1830	0	11344

* These variables are not used in any of the models but are displayed here for better sample description

Table 3b. Distribution of conditional cigarette demand for region-level dataset

	Number of cigarettes per month
Mean	77.08
Min	1.5
10th percentile	1.5
25th percentile	4
Median	14
75th percentile	85.75
90th percentile	240
Max	630
N	48097

Table 4. Sample means of variables by country and region

Region	Country	Smoking prevalence	Conditional cig demand	Age	Male	Pocket Money	Parenal Smoking
Africa	Cote D'Ivoire	0.07	34.25	14.27	0.54	0.36	0.18
	Kenya	0.06	76.42	14.06	0.48	0.4	0.19
	Nigeria	0.07	70.73	14.62	0.52	0.31	0.13
	Senegal	0.09	43.7	14.82	0.58	0.28	0.22
	S Africa	0.17	96.86	15.35	0.46	0.44	0.45
Africa		0.12	81.28	14.83	0.5	0.39	0.31
Mid East	Bahrain	0.11	103.61	14.27	0.46	0.83	0.32
	Egypt	0.03	72.7	13.59	0.67	0.65	0.52
	Iran	0.02	111.12	14.02	0.47	0.87	0.34
	Jordan	0.12	90.03	14.1	0.47	0.71	0.52
	Kuwait	0.13	150.18	14.24	0.46	0.75	0.39
	Morocco	0.04	96.08	14.25	0.52	0.4	0.27
	Pakistan	0.01	82.66	14.39	0.62	0.67	0.32
	Saudi Arabia	0.08	132.98	14.41	1	0.88	0.18
	Tunisia	0.09	111.17	14.22	0.49	0.65	0.52
	UAE	0.05	69.33	13.77	0.48	0.57	0.3
Mid East		0.07	104.14	14.03	0.53	0.66	0.38
Europe	Czech Rep	0.31	101.65	13.9	0.5	0.8	0.54
	Greece	0.1	178.1	13.87	0.51	0.82	0.68
	Hungary	0.28	129.68	14.39	0.45	0.85	0.58
	Kazakhstan	0.08	85.1	13.8	0.47	0.46	0.52
	Poland	0.2	130.46	14.47	0.46	0.83	0.63
	Romania	0.19	118.91	15.06	0.45	0.81	0.61
	Russia	0.23	123.26	13.45	0.48	0.79	0.63
	Turkey	0.07	97.8	13.7	0.54	0.73	0.59
	Ukraine	0.18	80.9	13.65	0.48	0.83	0.61
Europe		0.16	115.51	13.87	0.49	0.74	0.6
Americas	Argentina	0.26	127.4	14.55	0.52	0.8	0.58
	Brazil	0.1	89.41	14.73	0.45	0.58	0.37
	Chile	0.24	45.24	13.51	0.5	0.75	0.63
	Colombia	0.18	30.21	13.36	0.5	0.68	0.44
	Costa Rica	0.15	56.02	13.97	0.52	0.84	0.31
	Guatemala	0.09	35.45	14.08	0.44	0.57	0.26
	Mexico	0.12	40.91	13.5	0.47	0.64	0.4
	Panama	0.08	45.89	13.43	0.49	0.27	0.25
	Paraguay	0.1	44.11	13.72	0.47	0.48	0.35
	Peru	0.12	24.45	14.2	0.48	0.63	0.41
	Uruguay	0.17	100.74	13.6	0.49	0.72	0.53
	Venezuela	0.04	35.97	13.04	0.44	0.57	0.37
Americas		0.12	54.53	13.82	0.47	0.63	0.41
SE Asia	Bangladesh	0.02	68.42	13.97	0.72	0.46	0.35
	India	0.05	59.41	14.04	0.57	0.47	0.45
	Indonesia	0.12	35.44	13.6	0.47	0.93	0.57
	Nepal	0.06	10.47	14.78	0.54	0.15	0.56

	Sri Lanka	0.01	34.08	14.03	0.5	0.84	0.48
	Thailand	0.1	50.66	13.89	0.5	0.58	0.48
SE Asia		0.07	50.81	13.99	0.55	0.54	0.47
W Pacific	China	0.05	88.84	14.1	0.5	0.76	0.64
	Phillippines	0.12	58.48	14.91	0.38	0.58	0.58
	Singapore	0.1	108.64	14.52	0.49	0.93	0.37
	S Korea	0.05	100.82	13.47	0.46	0.85	0.56
	Vietnam	0.04	111.79	15.18	0.45	0.59	0.56
W Pacific		0.07	83.93	14.41	0.46	0.74	0.57

Table 4 continued. Sample means of variables by country and region

Region	Country	Anti-Smoking Sentiment	Cig Advertising	Anti-Tobacco Media	Youth Access Restrictions
Africa	Cote D'Ivoire	0.87	0.73	0.56	0.28
	Kenya	0.52	0.87	0.82	0.3
	Nigeria	0.62	0.72	0.7	0.41
	Senegal	0.91	0.77	0.8	0.28
	S Africa	0.55	0.86	0.79	0.34
Africa		0.72	0.78	0.71	0.31
Mid East	Bahrain	0.87	0.83	0.7	0.34
	Egypt	0.89	0.83	0.79	0.35
	Iran	0.91	0.67	0.84	0.25
	Jordan	0.8	0.75	0.81	0.33
	Kuwait	0.86	0.94	0.68	0.26
	Morocco	0.79	0.66	0.68	0.42
	Pakistan	0.96	0.81	0.78	0.58
	Saudi Arabia	0.75	0.77	0.69	0.41
	Tunisia	0.88	0.74	0.68	0.18
	UAE	0.73	0.86	0.74	0.37
Mid East		0.86	0.79	0.75	0.39
Europe	Czech Rep	0.83	0.93	0.77	0.33
	Greece	0.91	0.86	1	0.1
	Hungary	0.85	0.85	1	0.32
	Kazakhstan	0.92	0.87	1	0.32
	Poland	0.88	0.96	0.92	0.3
	Romania	0.93	0.88	1	0.27
	Russia	0.91	0.81	0.87	0.47
	Turkey	0.94	0.46	0.99	0.15
	Ukraine	0.92	0.87	1	0.37
Europe		0.9	0.84	0.94	0.31
Americas	Argentina	0.84	0.96	0.62	0.11
	Brazil	0.89	0.87	0.89	0.18
	Chile	0.88	0.91	0.8	0.17
	Colombia	0.88	0.91	0.83	0.3
	Costa Rica	0.91	0.95	0.74	0.38
	Guatemala	0.87	0.78	0.7	0.33
	Mexico	0.89	0.91	0.85	0.49
	Panama	0.85	0.66	0.77	0.28
	Paraguay	0.87	0.94	0.85	0.32
	Peru	0.91	0.88	0.9	0.28
	Uruguay	0.87	0.95	0.86	0.22
	Venezuela	0.88	0.86	0.81	0.31
Americas		0.89	0.89	0.85	0.31
SE Asia	Bangladesh	0.96	0.89	0.88	0.19
	India	0.76	0.92	0.79	0.49
	Indonesia	0.91	0.96	0.91	0.37
	Nepal	0.61	0.95	0.97	0.14

	Sri Lanka	0.92	0.89	0.9	0.63
	Thailand	0.93	.	0.91	0.68
SE Asia		0.78	0.92	0.82	0.46
W Pacific	China	0.61	0.65	0.79	0.17
	Phillippines	0.4	0.91	0.84	0.5
	Singapore	.	.	0.92	0.48
	S Korea	0.85	0.68	0.87	0.72
	Vietnam	0.87	0.68	0.9	0.23
W Pacific		0.69	0.68	0.84	0.26

Table 4 continued. Sample means of variables by country and region

Region	Country	Cig Price, local brand	Cig Price, imported	Tobacco Production
Africa	Cote D'Ivoire	1.58	4.73	0
	Kenya	2.16	4.32	608
	Nigeria	2.68	4.02	176
	Senegal	1.15	2.3	0
	S Africa	2.87	2.87	0
Africa		2.09	3.65	157
Mid East	Bahrain	2.29	2.29	0
	Egypt	2.78	2.78	0
	Iran	1.81	5.05	0
	Jordan	1.41	3.45	119
	Kuwait	2.06	2.06	0
	Morocco	2.73	5.26	222
	Pakistan	2.18	3.4	555
	Saudi Arabia	1.85	2.64	0
	Tunisia	4.53	4.53	323
	UAE	2.54	3.38	0
Mid East		2.42	3.49	122
Europe	Czech Rep	2.89	3.97	0
	Greece	2.8	3.11	11344
	Hungary	2.15	3.73	1128
	Kazakhstan	0.8	1.81	948
	Poland	2.22	3.21	575
	Romania	1.76	2.84	344
	Russia	1.88	3.55	0
	Turkey	3.42	4.34	1578
	Ukraine	1.84	2.15	10
Europe		2.19	3.19	1770
Americas	Argentina	1.6	2.13	3104
	Brazil	1.57	1.81	4998
	Chile	3.11	3.89	484
	Colombia	2.08	2.71	618
	Costa Rica	1.12	1.45	0
	Guatemala	2.57	2.82	1742
	Mexico	1.93	2.53	184
	Panama	2.14	2.58	702
	Paraguay	2.01	2.61	2288
	Peru	2.26	3.43	477
	Uruguay	1.81	4.03	844
	Venezuela	2.6	2.92	0
Americas		2.07	2.74	1287
SE Asia	Bangladesh	1.62	3.24	259
	India	2.8	4.57	492
	Indonesia	1.64	2.11	740
	Nepal	2.19	3.6	125

	Sri Lanka	4.68	8.94	0
	Thailand	2.07	2.99	1111
SE Asia		2.5	4.24	454
W Pacific	China	3.8	5.3	1956
	Philippines	1.26	1.7	649
	Singapore	4.88	5.21	0
	S Korea	2.24	2.8	0
	Vietnam	2.35	3.44	0
W Pacific		2.9	3.69	521

Table 5. Variance inflation factors for variables suspected of multicollinearity. R-squared from OLS of suspect variable on all other suspect variables.

Variable	With Sentiment VIF	With Sentiment R-squared	Without Sentiment VIF	Without Sentiment R-squared
Price, local brand	1.32	0.24	1.14	0.12
Anti-Smoking Sentiment	1.34	0.26		
Cigarette Advertising	1.12	0.11	1.11	0.10
Anti-Tobacco Media	1.16	0.14	1.10	0.09
Youth Access Restrictions	1.11	0.10	1.11	0.10
Mean VIF	1.21		1.11	

Table 6a. Cluster correlations of individual cigarette demand

Within-country	0.066
Within-region	0.045

Table 6b. Within-region cluster correlations of aggregate macro variables

Variable	Within-region cluster correlation
Smoking prevalence	0.32
Conditional cig demand	0.46
Price, local brand	0.24
Price, Imported-brand	0.42
Anti-Smoking Sentiment	0.46
Anti-Tobacco Media	0.36
Cigarette Advertising	0.52
Youth Access Restrictions	0.19

Tables 7abc. Two-part model of cigarette demand as a function of local-brand cigarette prices

Table 7a. Logit model of smoking participation
Coefficients represent marginal effects on the probability of smoking participation

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local	-0.043**	-0.047**	-0.034**	-0.033**	-0.050***	-0.052***	-0.038***	-0.037**
brand)	(0.019)	(0.021)	(0.014)	(0.015)	(0.018)	(0.020)	(0.014)	(0.014)
Anti-Tobacco		-0.077*		-0.103**		-0.062		-0.090*
Media		(0.045)		(0.046)		(0.047)		(0.047)
Cigarette			0.075*	0.111**			0.100**	0.130***
Advertising			(0.044)	(0.048)			(0.041)	(0.045)
Anti-Smoking					-0.090***	-0.083***	-0.101***	-0.094***
Sentiment					(0.030)	(0.029)	(0.029)	(0.026)
Youth Access	-0.094***	-0.098***	-0.087***	-0.089***	-0.111***	-0.112***	-0.105***	-0.104***
Restrictions	(0.032)	(0.028)	(0.032)	(0.027)	(0.035)	(0.030)	(0.034)	(0.028)
Age	-0.051***	-0.052***	-0.047***	-0.048***	-0.050***	-0.051***	-0.047***	-0.047***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Age^2	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Parental Smoking	0.036***	0.036***	0.036***	0.035***	0.036***	0.036***	0.036***	0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pocket Money	0.061***	0.060***	0.060***	0.059***	0.061***	0.061***	0.060***	0.060***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Obs	349,130	349,130	345,847	345,847	349,130	349,130	345,847	345,847
Price elasticity	-0.724**	-0.786**	-0.576**	-0.557**	-0.837***	-0.883***	-0.648***	-0.631**
Advertising elasticity			1.096*	1.638**			1.474**	1.916***

Table 7b. Generalized linear model of conditional cigarette demand
Coefficients represent marginal effects on log cigarettes per month.

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local brand)	-1.259***	-1.462***	-1.146***	-1.205***	-1.263***	-1.461***	-1.138***	-1.198***
	(0.311)	(0.322)	(0.298)	(0.322)	(0.310)	(0.320)	(0.297)	(0.327)
Anti-Tobacco Media		-1.278**		-1.288**		-1.284**		-1.274**
		(0.502)		(0.527)		(0.513)		(0.537)
Cigarette Advertising			0.924	1.068			1.009	1.112
			(0.803)	(0.696)			(0.819)	(0.723)
Anti-Smoking Sentiment					-0.115	0.044	-0.255	-0.130
					(0.438)	(0.305)	(0.436)	(0.310)
Youth Access Restrictions	0.516	0.181	0.543*	0.150	0.507	0.184	0.518*	0.138
	(0.317)	(0.325)	(0.314)	(0.327)	(0.319)	(0.327)	(0.315)	(0.334)
Age	-0.959***	-0.994***	-0.927***	-0.962***	-0.960***	-0.994***	-0.927***	-0.962***
	(0.261)	(0.261)	(0.265)	(0.265)	(0.261)	(0.261)	(0.265)	(0.265)
Age^2	0.037***	0.038***	0.035***	0.037***	0.037***	0.038***	0.035***	0.037***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Male	0.212***	0.210***	0.221***	0.218***	0.212***	0.210***	0.221***	0.218***
	(0.036)	(0.036)	(0.037)	(0.037)	(0.036)	(0.036)	(0.037)	(0.037)
Parental Smoking	0.133***	0.130***	0.136***	0.133***	0.133***	0.130***	0.136***	0.133***
	(0.022)	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.023)
Pocket Money	0.205***	0.203***	0.202***	0.200***	0.205***	0.203***	0.202***	0.200***
	(0.070)	(0.068)	(0.073)	(0.071)	(0.070)	(0.068)	(0.073)	(0.071)
Obs	33,187	33,187	32,532	32,532	33,187	33,187	32,532	32,532
Price elasticity	-1.259***	-1.462***	-1.146***	-1.205***	-1.263***	-1.461***	-1.138***	-1.198***
Advertising elasticity			0.811	0.938			0.886	0.976

Table 7c. Ordered logit estimates of the price elasticity of the probability of being in a smoker category

	Without Sentiment				With Sentiment			
Cigarettes per month	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 to 15	0.515***	0.527***	0.420***	0.396***	0.513***	0.520***	0.420***	0.396***
15 to 100	-0.426***	-0.436***	-0.359***	-0.339***	-0.424***	-0.430***	-0.360***	-0.339***
100 to 300	-0.878***	-0.899***	-0.728***	-0.688***	-0.874***	-0.887***	-0.729***	-0.687***
>300	-1.116***	-1.142***	-0.920***	-0.869***	-1.110***	-1.127***	-0.922***	-0.868***

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* p<.1, ** p<.05, *** p<.01

Tables 8ab. Two-part model of cigarette demand as a function of Imported-brand cigarette prices

Table 8a. Logit model of smoking participation
Coefficients represent marginal effects on the probability of smoking participation

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price	-0.065***	-0.069***	-0.054***	-0.051**	-0.072***	-0.074***	-0.056***	-0.054***
(Imported brand)	(0.021)	(0.024)	(0.019)	(0.021)	(0.019)	(0.022)	(0.018)	(0.019)
Anti-Tobacco		-0.084*		-0.105**		-0.069		-0.091*
Media		(0.046)		(0.048)		(0.049)		(0.051)
Cigarette			0.060	0.097**			0.085**	0.115**
Advertising			(0.044)	(0.049)			(0.040)	(0.046)
Anti-Smoking					-0.090***	-0.082***	-0.099***	-0.091***
Sentiment					(0.030)	(0.029)	(0.030)	(0.027)
Youth Access	-0.096***	-0.100***	-0.091***	-0.093***	-0.114***	-0.115***	-0.109***	-0.108***
Restrictions	(0.032)	(0.028)	(0.032)	(0.027)	(0.036)	(0.030)	(0.035)	(0.028)
Age	-0.048***	-0.049***	-0.045**	-0.046**	-0.048***	-0.048***	-0.044**	-0.045**
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)
Age^2	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Parental Smoking	0.036***	0.036***	0.035***	0.035***	0.036***	0.036***	0.035***	0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pocket Money	0.057***	0.056***	0.056***	0.055***	0.057***	0.057***	0.057***	0.056***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Obs	326,597	326,597	323,314	323,314	326,597	326,597	323,314	323,314
Price elasticity	-1.124***	-1.197***	-0.938***	-0.904**	-1.242***	-1.293***	-0.986***	-0.950***
Advertising elasticity			0.898	1.456*			1.291**	1.749**

Table 8b. Generalized linear model of conditional cigarette demand
Coefficients represent marginal effects on log cigarettes per month.

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price	-1.395***	-1.601***	-1.147***	-1.239***	-1.396***	-1.602***	-1.135***	-1.227***
(Imported brand)	(0.368)	(0.418)	(0.347)	(0.384)	(0.366)	(0.420)	(0.342)	(0.384)
Anti-Tobacco		-1.219**		-1.337**		-1.226**		-1.320**
Media		(0.485)		(0.536)		(0.495)		(0.542)
Cigarette			1.244	1.365*			1.331	1.417*
Advertising			(0.914)	(0.802)			(0.923)	(0.820)
Anti-Smoking					-0.100	0.047	-0.269	-0.161
Sentiment					(0.485)	(0.351)	(0.426)	(0.303)
Youth Access	0.533	0.213	0.500	0.104	0.525	0.217	0.471	0.087
Restrictions	(0.339)	(0.335)	(0.317)	(0.330)	(0.341)	(0.336)	(0.314)	(0.335)
Age	-0.912***	-0.942***	-0.879***	-0.915***	-0.913***	-0.942***	-0.880***	-0.915***
	(0.269)	(0.267)	(0.271)	(0.269)	(0.269)	(0.267)	(0.271)	(0.270)
Age^2	0.035***	0.036***	0.034***	0.035***	0.035***	0.036***	0.034***	0.035***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Male	0.217***	0.215***	0.225***	0.223***	0.217***	0.215***	0.226***	0.223***
	(0.036)	(0.036)	(0.038)	(0.038)	(0.036)	(0.036)	(0.037)	(0.038)
Parental Smoking	0.135***	0.132***	0.138***	0.135***	0.135***	0.132***	0.138***	0.135***
	(0.022)	(0.022)	(0.023)	(0.023)	(0.022)	(0.022)	(0.023)	(0.023)
Pocket Money	0.204***	0.203***	0.199***	0.198***	0.205***	0.203***	0.200***	0.198***
	(0.072)	(0.071)	(0.075)	(0.073)	(0.072)	(0.071)	(0.075)	(0.074)
Obs	30,534	30,534	29,879	29,879	30,534	30,534	29,879	29,879
Price elasticity	-1.395***	-1.601***	-1.147***	-1.239***	-1.396***	-1.602***	-1.135***	-1.227***
Advertising elasticity			1.090	1.196*			1.166	1.242*

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* $p<.1$, ** $p<.05$, *** $p<.01$

Tables 9ab. Two-part model of cigarette demand as a function of local-brand cigarette prices and tobacco production as proxy for sentiment

Table 9a. Logit model of smoking participation
Coefficients represent marginal effects on the probability of smoking participation

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local brand)	-0.043**	-0.047**	-0.034**	-0.033**	-0.044*	-0.044	-0.020	-0.010
	(0.019)	(0.021)	(0.014)	(0.015)	(0.024)	(0.027)	(0.024)	(0.022)
Anti-Tobacco Media		-0.077*		-0.103**		-0.079		-0.121**
		(0.045)		(0.046)		(0.048)		(0.051)
Cigarette Advertising			0.075*	0.111**			0.093**	0.148***
			(0.044)	(0.048)			(0.045)	(0.053)
Tobacco Production					0.763	-5.085	-18.934	-30.912**
					(16.155)	(15.704)	(19.241)	(15.513)
Youth Access Restrictions	-0.094***	-0.098***	-0.087***	-0.089***	-0.094***	-0.096***	-0.081***	-0.078***
	(0.032)	(0.028)	(0.032)	(0.027)	(0.030)	(0.028)	(0.031)	(0.028)
Age	-0.051***	-0.052***	-0.047***	-0.048***	-0.051***	-0.052***	-0.047***	-0.048***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Age^2	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***	0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Parental Smoking	0.036***	0.036***	0.036***	0.035***	0.036***	0.036***	0.036***	0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pocket Money	0.061***	0.060***	0.060***	0.059***	0.061***	0.060***	0.060***	0.059***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Obs	349,130	349,130	345,847	345,847	349,130	349,130	345,847	345,847
Price elasticity	-0.724**	-0.786**	-0.576**	-0.557**	-0.732*	-0.734	-0.347	-0.179
Advertising elasticity			1.096*	1.638**			1.363**	2.175***

Table 9b. Generalized linear model of conditional cigarette demand
Coefficients represent marginal effects on log cigarettes per month.

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local brand)	-1.259***	-1.462***	-1.146***	-1.205***	-1.381***	-1.442***	-1.119***	-0.947**
	(0.311)	(0.322)	(0.298)	(0.322)	(0.306)	(0.320)	(0.387)	(0.416)
Anti-Tobacco Media		-1.278**		-1.288**		-1.349**		-1.663***
		(0.502)		(0.527)		(0.573)		(0.599)
Cigarette Advertising			0.924	1.068			0.965	1.563*
			(0.803)	(0.696)			(0.965)	(0.878)
Tobacco Production					168.096	-50.944	-32.681	-352.451
					(181.126)	(195.709)	(264.098)	(307.515)
Youth Access Restrictions	0.516	0.181	0.543*	0.150	0.313	0.230	0.583	0.520
	(0.317)	(0.325)	(0.314)	(0.327)	(0.386)	(0.377)	(0.415)	(0.412)
Age	-0.959***	-0.994***	-0.927***	-0.962***	-0.952***	-0.998***	-0.927***	-0.976***
	(0.261)	(0.261)	(0.265)	(0.265)	(0.259)	(0.259)	(0.265)	(0.263)
Age^2	0.037***	0.038***	0.035***	0.037***	0.036***	0.038***	0.035***	0.037***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Male	0.212***	0.210***	0.221***	0.218***	0.213***	0.210***	0.221***	0.219***
	(0.036)	(0.036)	(0.037)	(0.037)	(0.036)	(0.036)	(0.037)	(0.037)
Parental Smoking	0.133***	0.130***	0.136***	0.133***	0.134***	0.130***	0.136***	0.131***
	(0.022)	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)
Pocket Money	0.205***	0.203***	0.202***	0.200***	0.203***	0.204***	0.202***	0.200***
	(0.070)	(0.068)	(0.073)	(0.071)	(0.069)	(0.068)	(0.073)	(0.071)
Obs	33,187	33,187	32,532	32,532	33,187	33,187	32,532	32,532
Price elasticity	-1.259***	-1.462***	-1.146***	-1.205***	-1.381***	-1.442***	-1.119***	-0.947**
Advertising elasticity			0.811	0.938			0.848	1.372*

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* $p<.1$, ** $p<.05$, *** $p<.01$

Tables 10ab. Two-part model of cigarette demand as a function of local-brand cigarette prices and region fixed effects

Table 10a. Logit model of smoking participation
Coefficients represent marginal effects on the probability of smoking participation

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local	-0.030***	-0.029***	0.004	0.006	-0.035***	-0.036***	0.003	0.006
brand)	(0.010)	(0.010)	(0.013)	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)
Anti-Tobacco		-0.045		-0.119***		-0.049		-0.112***
Media		(0.044)		(0.036)		(0.044)		(0.035)
Cigarette			0.193***	0.209***			0.193***	0.209***
Advertising			(0.045)	(0.041)			(0.045)	(0.041)
Anti-Smoking					-0.047	-0.041	-0.060**	-0.048*
Sentiment					(0.031)	(0.033)	(0.028)	(0.027)
Youth Access	0.004	0.007	-0.014	-0.012	-0.001	0.001	-0.019	-0.016
Restrictions	(0.022)	(0.023)	(0.022)	(0.020)	(0.022)	(0.022)	(0.022)	(0.020)
Age	-0.074***	-0.075***	-0.061***	-0.065***	-0.073***	-0.074***	-0.060***	-0.064***
	(0.018)	(0.018)	(0.016)	(0.016)	(0.018)	(0.018)	(0.016)	(0.016)
Age^2	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.040***	0.040***	0.038***	0.038***	0.040***	0.040***	0.038***	0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Parental Smoking	0.044***	0.044***	0.041***	0.041***	0.044***	0.044***	0.041***	0.041***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pocket Money	0.077***	0.077***	0.073***	0.073***	0.078***	0.078***	0.074***	0.074***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
Obs	494,515	494,515	459,035	459,035	482,521	482,521	459,035	459,035
Price elasticity	-0.430***	-0.426***	0.053	0.095	-0.512***	-0.525***	0.053	0.087
Advertising elasticity			2.454***	2.677***			2.466***	2.671***

Table 10b. Generalized linear model of conditional cigarette demand
Coefficients represent marginal effects on log cigarettes per month.

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local	-0.067	-0.067	-0.235*	-0.256*	-0.245**	-0.258**	-0.235*	-0.257**
brand)	(0.151)	(0.161)	(0.124)	(0.132)	(0.104)	(0.109)	(0.123)	(0.129)
Anti-Tobacco		-0.374		-0.670		-0.662		-0.670
Media		(0.535)		(0.500)		(0.469)		(0.499)
Cigarette			0.125	0.047			0.124	0.046
Advertising			(0.471)	(0.449)			(0.471)	(0.447)
Anti-Smoking					0.360	0.335	-0.058	-0.049
Sentiment					(0.415)	(0.442)	(0.539)	(0.557)
Youth Access	-0.830**	-0.828**	-1.320***	-1.335***	-1.056***	-1.081***	-1.331***	-1.345***
Restrictions	(0.373)	(0.377)	(0.425)	(0.381)	(0.355)	(0.336)	(0.485)	(0.430)
Age	-1.361***	-1.377***	-1.308***	-1.338***	-1.361***	-1.391***	-1.306***	-1.336***
	(0.234)	(0.235)	(0.250)	(0.251)	(0.243)	(0.244)	(0.258)	(0.260)
Age^2	0.051***	0.052***	0.049***	0.050***	0.051***	0.052***	0.049***	0.050***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
Male	0.135***	0.135***	0.148***	0.147***	0.138***	0.138***	0.148***	0.147***
	(0.031)	(0.031)	(0.034)	(0.034)	(0.032)	(0.033)	(0.033)	(0.033)
Parental Smoking	0.137***	0.135***	0.146***	0.146***	0.151***	0.150***	0.146***	0.146***
	(0.023)	(0.024)	(0.023)	(0.023)	(0.021)	(0.022)	(0.023)	(0.023)
Pocket Money	0.254***	0.255***	0.263***	0.266***	0.262***	0.265***	0.262***	0.266***
	(0.074)	(0.075)	(0.081)	(0.082)	(0.075)	(0.076)	(0.082)	(0.083)
Obs	48,317	48,317	44,468	44,468	47,132	47,132	44,468	44,468
Price elasticity	-0.067	-0.067	-0.235*	-0.256*	-0.245**	-0.258**	-0.235*	-0.257**
Advertising elasticity			0.108	0.041			0.107	0.040

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* $p<.1$, ** $p<.05$, *** $p<.01$

Table 11. Tobit model of cigarette demand as a function of local-brand cigarette prices
Coefficients represent marginal effects on total cigarette demand

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local	-5.420	-5.821	-4.282	-4.161	-5.948	-6.236	-4.563	-4.434
brand)	(11.551)	(9.470)	(20.535)	(11.938)	(7.180)	(6.913)	(12.383)	(8.401)
Anti-Tobacco		-7.988		-11.424		-6.665		-10.231
Media		(13.621)		(38.220)		(7.074)		(22.874)
Cigarette			7.310	11.525			9.442	13.006
Advertising			(43.037)	(41.298)			(32.767)	(31.909)
Anti-Smoking					-8.227	-7.453	-9.229	-8.415
Sentiment					(12.473)	(10.762)	(30.198)	(19.831)
Youth Access	-8.675	-9.027	-7.984	-8.148	-10.094	-10.260	-9.432	-9.458
Restrictions	(22.941)	(18.675)	(44.735)	(28.168)	(15.061)	(14.806)	(29.921)	(21.867)
Age	-7.732	-7.783	-7.287	-7.370	-7.643	-7.692	-7.170	-7.252
	(16.477)	(12.662)	(34.949)	(21.145)	(9.226)	(8.528)	(19.458)	(13.741)
Age^2	0.324	0.325	0.307	0.309	0.320	0.322	0.303	0.305
	(0.690)	(0.529)	(1.472)	(0.888)	(0.386)	(0.356)	(0.821)	(0.578)
Male	3.858	3.852	3.800	3.795	3.847	3.842	3.785	3.781
	(9.972)	(7.570)	(21.098)	(12.806)	(5.872)	(5.241)	(12.179)	(8.625)
Parental Smoking	3.622	3.581	3.520	3.471	3.620	3.586	3.519	3.474
	(9.572)	(7.274)	(19.749)	(11.959)	(5.679)	(5.077)	(11.489)	(8.132)
Pocket Money	5.515	5.479	5.392	5.329	5.565	5.530	5.436	5.376
	(14.793)	(11.260)	(30.769)	(18.662)	(8.880)	(7.932)	(18.103)	(12.828)
Obs	349,130	349,130	345,847	345,847	349,130	349,130	345,847	345,847
Price elasticity	-1.145	-1.235	-0.929	-0.909	-1.263	-1.328	-0.997	-0.973
Advertising elasticity			1.364	2.162			1.772	2.454

Ruud test of Tobit specification	51,946 (0.000)	51,942 (0.000)	51,294 (0.000)	51,319 (0.000)	51,907 (0.000)	51,901 (0.000)	51,253 (0.000)	51,277 (0.000)

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* p<.1, ** p<.05, *** p<.01

Table 12. Zero-inflated Poisson model of cigarette demand as a function of local-brand cigarette prices
Coefficients represent marginal effects on total cigarette demand

	Without Sentiment				With Sentiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Price (local	-5.352***	-5.663***	-4.445***	-4.248***	-5.699***	-5.941***	-4.630***	-4.441***
brand)	(1.529)	(1.783)	(1.172)	(1.388)	(1.483)	(1.682)	(1.106)	(1.284)
Anti-Tobacco		-7.466***		-9.138***		-6.713**		-8.403***
Media		(2.665)		(2.513)		(2.805)		(2.585)
Cigarette			6.050*	8.744***			7.557***	9.803***
Advertising			(3.095)	(3.274)			(2.931)	(3.111)
Anti-Smoking					-5.071**	-4.309**	-5.927***	-5.256***
Sentiment					(2.166)	(1.802)	(2.136)	(1.629)
Youth Access	-3.970*	-4.937***	-3.450*	-4.437***	-4.899**	-5.693***	-4.418**	-5.258***
Restrictions	(2.097)	(1.717)	(2.013)	(1.590)	(2.334)	(1.888)	(2.213)	(1.709)
Age	-6.055***	-6.056***	-5.701***	-5.731***	-6.007***	-6.008***	-5.630***	-5.660***
	(1.541)	(1.522)	(1.496)	(1.477)	(1.524)	(1.505)	(1.476)	(1.456)
Age^2	0.248***	0.247***	0.235***	0.235***	0.246***	0.245***	0.232***	0.232***
	(0.053)	(0.052)	(0.051)	(0.051)	(0.053)	(0.052)	(0.051)	(0.050)
Male	2.542***	2.504***	2.507***	2.472***	2.533***	2.496***	2.495***	2.461***
	(0.308)	(0.296)	(0.315)	(0.306)	(0.305)	(0.294)	(0.314)	(0.305)
Parental Smoking	2.265***	2.208***	2.204***	2.144***	2.263***	2.210***	2.201***	2.144***
	(0.206)	(0.208)	(0.211)	(0.214)	(0.199)	(0.204)	(0.205)	(0.210)
Pocket Money	3.702***	3.622***	3.606***	3.522***	3.734***	3.655***	3.634***	3.552***
	(0.250)	(0.241)	(0.243)	(0.238)	(0.253)	(0.245)	(0.246)	(0.241)
Obs	349,130	349,130	345,847	345,847	349,130	349,130	345,847	345,847
Price elasticity	-1.754***	-1.885***	-1.498***	-1.454***	-1.877***	-1.985***	-1.570***	-1.528***
Advertising elasticity			1.752*	2.572**			2.202**	2.899***

All specifications include year and country dummies
Standard errors clustered by survey site
Standard errors in parentheses
* $p<.1$, ** $p<.05$, *** $p<.01$

Chapter 2

1. Introduction

This paper investigates the impact of cigarette prices on smoking initiation and cessation among youth in low- and mid-income countries. Unlike industrialized nations, cigarette smoking is on the rise in developing countries, which are expected to carry 78% of the world's tobacco-related mortality by 2020 (Tobacco Control Country Profiles, 2003). In these countries smoking also occurs earlier in life - the starting age of smoking in the developing world can be as low as 7, and over 18 percent of adolescent smokers start smoking before the age of 10 (Global Youth Tobacco Survey (GYTS) 1999-2006).

The contribution of this paper to the existing literature on tobacco use is twofold. First, this is the first paper to examine initiation and cessation decisions among youth outside of the United States. In producing an estimate of the responsiveness of these decisions to cigarette prices, this research not only provides guidance for global anti-smoking policy but also provides an interesting comparison for the disagreeing domestic literature on the topic.

The second contribution is in the methodological application and in particular in discussing the sensitivity of the results to different types of empirical analysis. This research is the first to compare in a single study three different types of duration methods previously used in the literature with often conflicting conclusions about the role of prices on the decisions to start or quit smoking. These methods are the discrete-time logit hazard model, the Cox hazard model, and the split-population duration model. The split-population model extends the other two by relaxing one of the underlying assumptions of standard duration analysis, namely, the assumption that all subjects who do not initiate smoking over the period of observation will eventually and certainly take up smoking if given enough time. This model is particularly appropriate in the modeling of smoking decisions where some people will never smoke or some smokers will never quit. While there are several studies utilizing the split-population model in the context of smoking initiation of cessation, this is the first to compare the sensitivity of the results between the split-population approach and the Cox and discrete-time Logit models.

A statistically significant impact of cigarette price on the initiation (cessation) hazards is identified by the split-population analysis but not by either of the unsplit Logit and Cox models. The split-population model may provide more valid price elasticity estimates by taking into account the influence of those individuals who never smoke or who, having started smoking, will never quit. These individuals may not be as responsive to price and can attenuate the effect of price on the hazards of starting and quitting for the whole population. Once their influence is accounted for in the split-population model, cigarette price are be shown to be effective in delaying the onset of smoking as well as in shortening the duration of the smoking spell. The price elasticity of initiation is estimated at -0.165, so that a 10% increase in the price of cigarettes corresponds to about 1.7% decrease in the hazard of starting smoking. The price elasticity of cessation is estimated at 0.27, so that a 10% increase in the price of cigarettes corresponds to about 2.7% increase in the hazard of quitting.

A key identification concern in the analysis is the presence of country-specific unobserved characteristics that can be correlated with both cigarette prices and the frequency of smoking initiation. If these are not accounted for, they may lead to biased and inconsistent estimates of the price effect. For example, local anti-smoking sentiment may translate into both higher cigarette prices (through higher taxation) and lower smoking. If sentiment is left out of the analysis, its impact on smoking will be picked up by cigarette prices, leading to a biased price effect estimate. This paper addresses this issue with country fixed effects which remove the unobserved time-invariant influences of the local environment.

Although the usual way of evaluating smoking patterns and their determinants is by modeling smoking participation at a point in time, looking at the decision to start and quit smoking over a period of time has several advantages over looking at smoking participation rates alone. Smoking participation describes a static condition – it is a stock variable reflecting the inflow of starters and the outflow of quitters at each point in time. On an individual level, smoking participation at a certain age reflects the cumulative *past* decisions to start or quit smoking at all previous ages. These past decisions to smoke or not to smoke hold a wealth of information about the determinants of smoking which would be missed by looking exclusively at current smoking participation. Another problem with modeling smoking participation as a stock variable is that it ignores the addictiveness of smoking by not conditioning on past smoking status (DeCicca et al 2008). In contrast, past smoking status enters implicitly in the modeling of initiation and cessation decisions because the probabilities of starting or quitting smoking in each time period are conditional on past smoking status.

In investigating the drivers of negative behaviors such as smoking, it can be a good idea to focus on populations that are particularly vulnerable to such behaviors. In this sense, this research targets the right sample by considering youth from developing countries. Since smoking is initiated mainly in youth, adolescents in general are at higher risk for starting to smoke. Adolescents in lower-income countries are at an even higher risk due to the higher popularity of tobacco use there. I use GYTS survey data on nearly 420,000 school-age individuals from 44 lower-income countries to construct a dataset where each individual's smoking behavior is followed over time in an effort to determine if this behavior can be affected by cigarette prices.

Smoking behavior in youth has different patterns in different parts of the world. The relative hazards of initiation by region are shown in Figure 1a, and Figure 1b has the corresponding hazards of cessation. The hazard of initiation among adolescents is highest in Europe and Latin America and is lowest in Southeast Asia and the Middle East. Although youth from Latin America are at high risk for starting to smoke, they are less likely to stick with the habit and have one of the highest rates of quitting. This is not the case with Europe where youth not only have one of the highest risks for starting to smoke but are also among the least likely to quit. Youth in Southeast Asia have a low hazard of starting but once having started, they maintain the habit longer and have a very low quitting rate. The same holds for countries in the African region. In most regions, the hazard of initiation peaks between the ages of 12 and 15. The hazards of cessation do not have a recognizable peak for most regions, and are relatively constant over time. One exception is the Western Pacific region, where the highest rate of quitting occurs about 9 years after smoking has started.

2. Literature review

Although evaluating smoking initiation and cessation has some advantages over evaluating smoking participation, the literature on initiation and cessation is not nearly as extensive as that on participation. Studies have generally focused on analyzing data from the United States (Cawley, Markowitz and Tauras (2006, 2004)), (DeCicca et al (2008, 2002), Tauras and Chaloupka (1999), Douglas (1998), Douglas and Hariharan (1994)), with some looking at British data (Forster and Jones (2001), Madden (2007)) and one study considering Spanish data (Nicolas (2002)). There seems to be no consensus regarding the effect of price or tax on smoking initiation. Some studies find a significant price effect while others do not, and the effects may differ according to gender. The price responsiveness of quitting however seems to be less controversial than that of starting; most studies find that higher cigarette prices lead to increased quitting.

Among the studies providing mixed evidence on the impact of cigarette price on adolescent smoking initiation are Cawley, Markowitz and Tauras (2006, 2004). These studies use data on youth from different cohorts of the National Longitudinal Survey of Youth and estimate the effect of price while also taking into account the effect that body weight could have on the smoking uptake decision. They find evidence that price affects the smoking behavior of males but not females.

Studies that show lack of tax/price responsiveness of smoking initiation include DeCicca et al (2008, 2002), Douglas (1998), and Douglas and Hariharan (1994). DeCicca et al (2002) use US data on high-school students and find that cigarette taxes have no effect on the onset of smoking between the 8^{th} and 12^{th} grades once state fixed effects are included. In a follow-up study from 2008, the authors confirm these findings using a slightly older sample of individuals up to age 26. In this study, DeCicca et al

(2008) use a direct measure of state anti-smoking sentiment as well as state fixed effects to identify the tax responsiveness of smoking initiation and cessation in the presence of state-level unobservables. Employing a discrete-time hazard probit model, they find that the young people in their sample are not responsive to cigarette taxes when initiation is concerned but that higher taxes may promote quitting.

Douglas and Hariharan (1994) are the first to use the split-population duration model for estimating smoking initiation as a function of cigarette price. This model allows them to control for the possibility that a certain portion of the population will never smoke. However, their specification does not account for state or time fixed effects. Ignoring state fixed effects can interfere considerably with the identification of the price effect if unobserved state characteristics like anti-smoking sentiment are correlated with both cigarette prices and smoking prevalence. An exogenous time trend is also necessary to reflect the change in attitudes toward smoking that may have occurred independently over time. Douglas (1998) extends the analysis in Douglas and Hariharan (1994) by introducing controls for state-level regulation but still fails to control for state and time effects. Both of these two studies find no evidence that cigarette prices influence smoking initiation.

Studies finding that price/tax lowers initiation or increases cessation include Tauras and Chaloupka (1999), Forster and Jones (2001), and Nicolas (2002). Tauras and Chaloupka (1999) follow a sample of U.S. high-school cigarette-smoking seniors through early adulthood and examine their decision to quit smoking during this period. They use a Cox hazard model to estimate the effect of cigarette prices and find that it is a significant determinant of the quitting hazard with a price elasticity of quitting slightly above unity. Although this study includes fixed effects by Census region, it does not include state fixed effects which may bias the results if there is unobserved state-level heterogeneity. Forster and Jones (2001) use British data to estimate the tax elasticities of starting and quitting smoking. Like Douglas (1998) and Douglas and Hariharan (1994), they allow for the probability of never smoking by modeling initiation with a split-population duration model. In modeling the prevalence of initiation, they do not include tax as an explanatory variable so no inference can be made about the tax elasticity of the probability of initiation. Tax, however, is included in the model of time-until-initiation, resulting in a 0.16 and 0.08 tax elasticity of age of initiation for men and women, respectively. They do not use a split-population model to model cessation since they assume that every smoker will eventually quit. Applying a non-split generalized gamma model on the sample of smokers, they estimate the tax elasticity of the length of the smoking spell as -0.6 for men and -0.46 for women.

Nicolas (2002) uses Spanish data to perform a study of initiation and cessation that mirrors the Forster and Jones (2001) analysis of British data. Using a split-population duration model for initiation, he finds that a 10% increase in price corresponds to a modest month-long delay in the onset of smoking. No estimate of the price elasticity of the probability of initiation is provided because price is not used to predict it. Quitting is estimated with a non-split Weibull model, which shows that a 10% increase in price can shorten the smoking habit by up to 18 months.

Most recently, Madden (2007) uses data on Irish women to determine the effect of cigarette taxes and education on starting and quitting smoking. The analysis is similar to Forster and Jones (2001) in employing split-population model to initiation and standard non-split parametric model to cessation. For both initiation and cessation, the results on the tax effect are inconclusive and its significance is not robust to different specifications. Given the limited nature of the sample and the relatively weak tax results, this study does not provide a definitive conclusion regarding the impact of tax.

In summary, the literature on smoking initiation and cessation has produced contradictory conclusions regarding the effect of cigarette prices (taxes), and more research is needed to shed light on the question of people's responsiveness to cigarette prices. This is especially true for developing countries which have had to rely on evidence from developed countries for insight on their own policies. In taking a global view at this question, this article provides a unique contribution to a yet unsettled topic of research.

3. Data and variables

The dataset is a combination of two main sources. Micro-level data on individual characteristics and smoking behavior are obtained from the Global Youth Tobacco Survey (GYTS). These are merged with country-level data on cigarette prices from the Economist Intelligence Unit's World Cost of Living Survey (EIU). Besides Kostova et al (2009), this is the fist study to utilize GYTS data in combination with cigarette prices and is therefore an original analysis of youth's smoking decisions as a function of price.

The GYTS is a survey developed by the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) to track tobacco use of young people across countries with a common methodology. It has been conducted in 135 low-to-mid-income countries from the six major world regions (Africa, Europe, Americas, Southeast Asia, Middle East, and Western Pacific) in various years from 1999 to 2006. It captures prevalence, access, media exposure and attitudes related to tobacco use among individuals in school grades corresponding to ages 13 to 15, although in practice the age range of the survey is wider and covers individuals between the ages of 11 and 19.

Since the goal of this research is to analyze the rates of smoking initiation and cessation that occur over time during the years of adolescence, the dataset needs to be of longitudinal nature. The GYTS is not a longitudinal survey that follows and re-interviews individuals over time. Instead, it collects information on current smoking status and other behaviors at a fixed point in time. However, it is possible to construct the necessary longitudinal dataset by inferring the past from some of the survey questions. This is possible because the GYTS contains retrospective information on smoking such as age of initiation and time of quitting, from which the length of the smoking habit can be inferred. Specifically, the answer to the question "How long ago did you stop smoking?" allows me to distinguish between never-smokers, current smokers, and former smokers who have quit within the previous year, 2 years earlier, or 3 or more years earlier. This produces information on the year of cessation[2]. For those individuals classified as current or former smokers, the year of initiation is approximated from the question "How old were you when you first tried a cigarette?" [3] Knowing the

[2] For those who indicate having quit 3 or more years earlier, I assume quitting occurred 3 years earlier. Although this would produce a measurement error for those quitting 4 or more years earlier, I assume that this error is negligible. The average age of the survey participants is fairly young, so it is safe to assume that not many would have had time to start, form a habit, and quit more than 3 years before they were interviewed.

[3] The answers to this question do not provide the exact age of starting smoking but only provide ranges spanning over 2 years, such as: 7 years old or younger, 8-9 years old, 10-11 years old, 12-13 years old, 14-15 years old, and 16 years old or older. I take the midpoint of each range as the age of starting smoking which effectively means that initiations in this sample are assumed to have occurred only at the following

years of initiation and cessation permits matching of the time of the smoking decision to the price of cigarettes at the time of the decision. Over the period of observation, 16 percent of subject initiate smoking, and 46% of smokers quit. The average length of the smoking spell among quitters is 3 years.

Cigarette Prices

Data on the price of cigarettes over time is obtained from the EIU World Cost of Living Survey. This is a privately developed survey by the publishers of The Economist magazine. It collects retail price data of a wide range of consumer products on a bi-annual basis from multiple cities worldwide, including many developing countries. Cigarette price data are available on two different brands, a locally popular brand and an imported brand, usually Marlboro. Prices are collected from one or more cities in each country. If for a particular country cigarette price data come from multiple cities, the averaged national price is used in this study. Prices are in U.S. dollars based on the relevant exchange rate and are converted into real terms using the 2000 U.S. GDP deflator. They are also adjusted using purchasing power parities (PPP) obtained from the World Bank's World Development Indicators database. The PPP adjusts prices for the local standard of living and allows for better price comparison between countries. In the primary analysis of smoking initiation and cessation, I use local-brand cigarette prices, but a sensitivity analysis using Marlboro prices is performed as well.

---

ages: 7, 9, 11, 13, 15, or 16. This certainly produces a measurement error in the estimated year of initiation. However, the year of initiation as well as the year of cessation form the dependent variable in this analysis, so this measurement error should not cause estimation problems as it will be reflected in the error term.

The final dataset used in this research excludes many of the original GYTS countries due to unavailability of matching price data. However, the geographic variation of price is increased by the fact that in some countries GYTS surveys were conducted in multiple local sites like cities or provinces. Where the GYTS city survey site matches the EIU city survey site, local city prices are used instead of the nationally averaged price. This produces geographic variation of price within country for some countries.

Other independent variables

Besides prices, other determinants of smoking behavior in this analysis include age (*Age*), sex (*Sex*), parental smoking status (*Parental Smoking*), a variable indicating whether the person receives pocket money at the time of the survey (*Pocket Money*), and per capita GDP (*GDP*). *Sex* is a binary variable equal to 1 if the subject is male. *GDP* controls for the relative wealth of each country and puts cigarette prices in better perspective in terms of affordability. It is also necessary in case people's smoking behavior changes in times of economic growth. *Parental Smoking* is a binary indicator equal to 1 if one or both parents smoke at the time of the interview. *Pocket Money* is a binary indicator equal to 1 if the subjects receives pocket money or personal income at the time of the interview. Although *Parental Smoking* and *Pocket Money* are not inherently fixed over time, the only information available on them is from the year of the interview. There are no questions in the survey from which it is possible to infer past parental smoking or pocket income in previous years. Therefore these variables are merely proxies for some unobserved fixed family characteristics that may influence smoking. For example, if an individual reports that at least one of his parents smokes at the time of the interview when he is 16, this does not necessarily mean they also smoked when he was 10 (although it is fairly likely), but it may indicate a more permissive parental attitude toward smoking that is constant over time. As another example, if an individual reports receiving pocket money on the date of the interview, then this may indicate a higher-income family and more access to money prior to the interview as well.

In addition to controlling for individual and family characteristics, it is necessary to account for a secular time trend that may influence smoking. In particular, attitudes toward smoking may change independently over time as more health information becomes available and/or more schools in developing countries implement anti-tobacco education. This is done by adding a time trend *Year* to the model. *Year* is measured in years since 1989. The time trend reduces noise but is also costly in terms of identification of the price effect. Since all variation in cigarette prices can be attributed to variation across calendar years, it is difficult to separate the time trend and the price effect. What has customarily been done in the literature to counteract this problem is adding a higher-order polynomial in the time trend. In this analysis a quartic polynomial in *Year* is included.

As it frequently happens with individual-level data obtained from surveys, multiple observations are missing due to non-response or absent questionnaire parts. This is a nontrivial problem since missing observations from four major individually descriptive variables – age, sex, parental smoking status, and receiving of pocket money – add up to 20 percent of the total number of observations. Out of these, the pocket money variable is missing most frequently due to absence of a related question in the survey questionnaires for some countries and years. Since we cannot assume that these observations are "missing completely at random", excluding them may lead to estimation bias. We may assume, however, that the missing observations can be classified as "missing at random", meaning that they can be explained by available data and therefore imputed. This is especially obvious in the case of the pocket money variable, where most missing values can be explained by country and year. We use the method of iterative imputation to fill in missing observations for age, sex, parental smoking status, and pocket money. This method has been recognized to have advantage over alternatives such as substitution of missing values by the sample mean or regression methods, both of which can lead to underestimation of the standard errors and erroneously significant results (Schafer & Olsen 1998).

4. Methods

The goal is to estimate the hazard of starting and quitting smoking over a period of time – in this case, during the years of adolescence. This presumes use of duration (survival) analysis. Duration analysis is concerned with analyzing time to the occurrence of an event, also known as failure. "Failure" in this article refers to either starting smoking in the modeling of initiation, or quitting smoking in the modeling of cessation. In estimating the length of time to failure, duration analysis calculates the hazard rate of failure (the hazard) – the probability that failure will occur at each time t, conditional on its not having occurred yet. Longer times-to-failure translate into lower hazard rates and vice versa.

The events of interest in this analysis are smoking initiation and cessation between the ages 9 and 19. The hazard of smoking initiation at a particular moment in time is the probability that a person will start smoking at that time, given that he has not started smoking yet. The hazard of smoking cessation at a point in time is the probability that a current smoker will quit smoking at that time, given that he is still a smoker. Although unconditional hazard rates can be calculated from the observed timings of smoking initiation or cessation in the dataset, hazard rates that are conditional on covariates such as prices, gender, etc. are not observed and must be modeled parametrically to provide inference on covariate effects. The goal of this research is to compare the appropriateness of three different duration techniques for modeling the hazards of initiation and cessation and to estimate how these hazards may respond to changes in cigarette price.

A fundamental difference between duration analysis and conventional longitudinal analyses is that in duration analysis, time does not have the same meaning as calendar time. Calendar time flows in the same way for every individual. Time in duration analysis is the number of periods elapsed since the individual started being at risk for failure, and varies for each individual. Explaining the individual variation in failure times is what duration analysis is concerned with. Given this distinction, it is necessary to start observing each individual from the time when he starts being at risk for failure (time zero). The period of observation ends for each individual when he experiences the risky event (fails). If no failure occurs by the date of the interview, observation ends on the year of the interview and the individual is considered censored. In the case of smoking initiation, time zero is assumed to be the year in which each individual is 9 years old – i.e., each individual starts being at risk for starting to smoke at age 9. Therefore, when reference is made to initiation time *t*, this means not calendar year but to the number of years elapsed between age 9 and the year of initiation (or censoring, whichever comes earlier.) The choice of age 9 as the beginning of risk for smoking is somewhat arbitrary. It is based on the information that about 20 percent of the smokers in the GYTS survey report having tried their first cigarette before the age of 10. Deciding on time zero in the analysis of smoking cessation is much clearer. In this case, time zero is the year smoking is initiated – i.e., each individual starts being at risk for quitting as soon as he starts smoking. When reference is made to cessation time *t*, this means the number of years from the year of initiation to the year of cessation or censoring (i.e., length of smoking spell).

4.1. Non-parametric estimation.

A critical assumption in parametric duration modeling is the probability distribution of failure times. Non-parametric modeling of the hazard rates can be used to provide information on the probability distribution of failure times and can therefore be helpful in finding the best distributional assumption. More specifically, the shape of the hazard function with respect to time can be used to determine what the assumed probability distribution of failure times should be.[4] Non-parametric methods such as the

[4] By definition, at each possible failure time t, the hazard rate $h(t)$ directly corresponds to the probability density of failure times $f(t)$. In notation, $h(t)=\frac{f(t)}{S(t)}=\frac{f(t\mid t>0)}{f(t\mid t=0)}$, where $S(t)$ is the survival function. In words, the hazard at time t is the ratio of the density of failing at time t to the probability of not failing, or surviving, until time t. Since the hazard is proportional to the density, the density function can be inferred from the shape of the hazard function.

Nelson-Aalen method can produce a picture of the observed hazards of initiation (cessation) in our sample without conditioning on any covariates or assuming a distributional form for the times to initiation (cessation). Based on the assumption that the hazard is the same for everyone, the Nelson-Aalen estimator approximates the hazard at each point in time as the ratio of the number of events occurring at that time to the number of individuals at risk. Figures 2a and 3a represent the nonparametric estimates of the initiation and cessation hazards, respectively. For both initiation and cessation, the hazard is non-monotonic, first increasing and then decreasing. Initiation hazard peaks around the age of 13 at about 3 percent. Cessation hazard increases until around 4 years into the smoking spell, remains fairly flat at 15 percent between 4 and 9 years and falls slightly afterward. The nonparametric shape of the hazard functions suggests that the distributions of time to initiation or cessation must be non-monotonic such as the log-normal, logistic, or log-logistic functions.

Using the nonparametric hazards in Figures 2a and 3a as a baseline for comparison, I develop three parametric hazard models of initiation and cessation. I begin by setting up a simple discrete-time model of initiation and cessation as in DeCicca et al (2008). Next I set up a semi-parametric Cox hazard model as in Tauras and Chaloupka (1999). Finally, I introduce a split-population duration model as in Forster and Jones (2001) and Douglas and Hariharan (1994), among others. Comparing the shape of the hazard rates of each parametric model to the nonparametric hazards can be used as evidence for goodness of fit.

4.2. Parametric estimation

The hazards (h_t) of initiation and cessation are modeled as functions of current cigarette price ($Price_t$), a vector of individual characteristics X_i, country-level characteristics that are both time-variant (GDP_{jt}) and fixed ($Country_j$), and a quartic calendar time trend ($Year_t$):

$$h_t(\text{initiation}) = \Pr(\text{initiate}|\text{no prior smoking}) =$$

$$f(\alpha_0+\alpha_1 Price_t+\alpha_2 X_i+\alpha_3 GDP_{jt}+\alpha_4 Year_t+\alpha_5 Country_j) \qquad (1)$$

$$h_t(\text{cessation}) = \Pr(\text{cessate}|\text{prior smoking}) =$$

$$f(\beta_0+\beta_1 Price_t+\beta_2 Age+\beta_3 X_i+\beta_4 GDP_{jt}+\beta_5 Year_t+\beta_6 Country_j) \qquad (2)$$

where t denotes analysis time[5], i denotes individual, and j denotes country. The vector of individual characteristics X_i includes *Sex*, *Parental Smoking*, and *Pocket Money*. *Age* is age at time t [6], GDP_{jt} is the log of current per-capita GDP, $Year_t$ is a quartic polynomial in calendar year since 1989, and $Country_j$ is a binary variable for each country – the country fixed effects.

4.2.1 The discrete-time logit hazard model

[5] It is important to remember that in duration analysis, analysis time t is time since beginning of risk, which is not necessarily equivalent to calendar time. For example, in the initiation model, t=1 is when age=9, which occurs in different calendar years for different individuals. In the cessation model, t=1 is the year when smoking started, which is also different for everyone.

[6] Note that the initiation hazard model in Equation 1 does not contain *age* as a covariate. This is because for each individual analysis time t is determined solely in reference to age (initiation risk begins at age 9, so t is simply the number of years since age 9). Since the hazard rate h_t^I is the distribution of initiation times t, the covariate age_t would be collinear with the outcome and would have no explanatory power. This is not the case in the cessation hazard model, where analysis time t is determined in reference to the year of smoking initiation. *Age* is therefore included in the model of cessation.

If $f(.)$ is the logistic distribution function, Equations 1 and 2 can be estimated by logit. This is known as the discrete-time logit hazard model, where the hazard function (using general notation) is

$$h_t^{LOGIT} = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}.$$

4.2.2 The Cox proportional hazards model

It is possible to estimate the hazard parameters without specifying a particular distribution for the hazard. This can be done with the Cox proportional hazards model. The Cox model does not impose a particular shape on the hazard function but it assumes that whatever the shape of the hazard may be, it is the same for everybody. The presumption of the Cox model is that there exists a baseline hazard function, h_t^{baseline}, which is not affected by individual covariates. This baseline hazard is the same for everyone and is the single underlying shape of everybody's individual hazard. The individual hazards are determined by multiplying the baseline hazard by the effect of the covariates, i.e., the individual hazards shift multiplicatively with the baseline and are proportional to it. In general notation, the hazard is

$$h_t^{COX} = h_t^{baseline}\exp(x_i\beta)$$

The baseline hazard h_t^{baseline} is estimated non-parametrically as in the Nelson-Aalen method, which means that the distribution of the number of failures over time is taken as given by the data. Once this is done, the individual probability of failure at each time is modeled as a function of the covariates, parametrizing the covariate effects on the hazard. The Cox hazard model is semi-parametric because it estimates the distribution of failure times nonparametrically while estimating the probability of failure around each failure time parametrically.

4.2.3 The split-population duration model

Both the logit hazard model and the Cox hazard model treat individuals who have not experienced failure by the date of the interview as censored observations, assuming that they would eventually fail if the observation period were long enough. This may not be a reasonable assumption in the case of smoking, where some individuals will never smoke, or some smokers will never quit. In the analysis of smoking initiation, it is possible to account for the possibility that a certain proportion of people will never start smoking by splitting the sample into never-smokers and potential/observed smokers. First, a probability of ever smoking is estimated in order to distinguish between never-smokers and potential smokers. Then, assuming that the initiation hazard for the never-smokers is zero for all times t, initiation hazard rates are estimated only for the smokers. A similar procedure can also be applied to the analysis of smoking cessation where the sample of smokers can be split between never-quitters and quitters. The type of duration model which allows for the possibility of never failing is referred to as the split-population model in the economics literature and was first introduced by Schmidt and Witte (1989) in the treatment of criminal recidivism.

The main advantage of the split-population model over non-split models is that the former weights the hazard of failure by the probability of ever experiencing failure. The contribution of individual i to the log-likelihood function of smoking initiation is

$$d_i \text{*ln}\{\text{P(ever initiate)*}f(t|t>0)\} + (1-d_i)\text{*ln}\{\text{P(never initiate)+P(ever initiate)*}f(t|t=0)\}$$

where d_i is a binary indicator for starting to smoke at some point during the period of observation, t is the time of initiation measured in number of years since age 9, and $f(t)$ is the probability density function of these times-to-initiation. Initiation time t is positive if initiation takes place between age 9 and the interview date, and zero if no initiation is observed by the interview date. If an individual is observed to starts smoking, his contribution to the likelihood function is the probability of starting to smoke multiplied by the density of starting to smoke at time t. If he is not observed to start smoking, his contribution is the probability that he will never smoke, plus the probability of starting multiplied by the density of starting after the observation period ends.

Similarly, for quitters, the individual contribution to the likelihood function is

$$d_i \text{*ln}\{\text{P(ever quit)*}f(t|t>0)\} + (1-d_i)\text{*ln}\{\text{P(never quit)+P(ever quit)*}f(t|t=0)\}$$

where d_i is a binary indicator if quitting is observed, t is the time of cessation measured in number of years since smoking began, and $f(t)$ is the probability density function of these times-to-cessation. Cessation time *t* is positive if quitting takes place by the interview date, and zero if no quitting is observed by the interview date. If an individual is observed to quit, his contribution to the likelihood function is the probability of quitting multiplied by the density of quitting at time *t*. If he is not observed to quit, his contribution is the probability that he will never quit, plus the probability of quitting multiplied by the density of quitting after the end of observation.

The choice of the distribution of failure times for the starters and quitters in the split-population model depends on the shape of the hazard functions for starting and quitting. As shown in Section 4.1, non-parametric approximations of the hazard rates of both initiation and cessation indicate that they first increase and then decrease. Therefore the functional form has to be non-monotonic, allowing for changes in the direction of the hazard over time. Following convention, I use the log-logistic distribution to model the hazards of initiation and cessation for the smokers and the quitters, respectively. The probabilities of ever smoking or quitting are estimated using logit.

5. Results

5.1. Initiation

The results from initiation hazard models using local-brand cigarette prices are listed in Table 1a. Table 1b contains the corresponding price elasticities of initiation. The first two columns of Table 1a contain the marginal effects from the Logit hazard models of initiation. The coefficients from the Cox and split-population models are shown in the middle two and last two columns, respectively. These coefficients do not have a direct interpretation as marginal effects, but their signs and statistical significance provide information about the effects of the covariates on the hazard.[7]

With the exception of *Price,* all explanatory variables have a positive and significant impact on the hazard of initiation. Male sex, availability of pocket income, having parents who smoke, and higher per capita GDP all increase the probability that a previously non-smoking individual will start smoking. In contrast, cigarette price has a negative coefficient in all specifications, and is also statistically significant as long as country fixed effects are not accounted for. Accounting for country fixed effects reduces the magnitude of the impact of *Price* and removes its statistical significance in the Logit and Cox models. However, country fixed effects do not reduce the statistical significance of *Price* in the split-population model, even though they still reduce its magnitude.

After using country fixed effects to control for unobserved environmental characteristics such as anti-smoking sentiment, it is possible to identify the presence of a price effect only by splitting the sample into never-smokers and eventual smokers. In interpreting this result, it is important to remember the key structural distinction between the split-population model and the unsplit models, namely that the split-population model removes the influence on the initiation hazard of the individuals who are never going to initiate, the never-smokers. The unsplit models, in contrast, assign a positive initiation hazard to everyone, including the never-smokers. If cigarette prices are not a significant

[7] The coefficients from the Cox and split-population models do not have a direct interpretation but can be transformed into hazard ratios by exponentiation. The hazard ratio is the ratio of the hazard for a 1-unit change in the corresponding covariate. For example, the exponentiated coefficient on *Price* in the last column of Table 1 is exp(-0.14) which equals to a hazard ratio of 0.87. This means that the hazard of initiation for someone facing a cigarette price that is higher by $1 is only 87% of the hazard for someone who does not face the higher price. In other words, a $1 cigarette price increase corresponds to a 13% decrease in the initiation hazard.

factor in determining the initiation hazards of the never-smokers (as should be expected, since the never-smokers are not truly at risk of initiation), then including the never-smokers in the sample may dilute the estimated overall price effect and lower its significance.

The implication is that if everyone was assumed to start smoking eventually, then cigarette prices do not determine the onset of smoking initiation. If, however, the analysis allows for the possibility that some people will never smoke and only considers the true eventual smokers, then higher cigarette prices reduce the hazard of initiation at each point in time and can effectively delay the onset of smoking. The price elasticity of initiation in the split-population model is -0.165, so that a 10% increase in the price of cigarettes corresponds to about 1.7% decrease in the initiation hazard rate.

Initiation results by gender

To check how the determinants of initiation may differ by gender, the logit discrete-time hazard model is applied to subsamples of males and females only. After controlling for country fixed effects, the price elasticity for boys becomes statistically equal to zero. In comparison, the price elasticity for girls is higher in terms of both magnitude and statistical significance. As shown in Table 5, it is equal to -0.317, implying that a 10% increase in the price level would correspond to a 3% drop in the initiation rate for girls. This is in contrast to recent findings from U.S. data which find that cigarette price is more likely to influence initiation among young men than women (Cawley et al. 2006, 2004). Why such difference in the gender responses may exist between developing countries and the U.S. is open to interpretation. Perhaps girls in developing countries have higher income constraints than boys, leading them to be more responsive to changes in the price of cigarettes. This is possible because even though the present models include a rough control for family income through the *Pocket Money* variable, they do not account for the actual amount of money received by each subject at each point in time. Another possible explanation is behavioral differences in money handling among the genders in developing countries. It has been shown that low income females in some developing countries show more responsibility with money allocation than males. The present finding of higher price responsiveness among females may be an indirect reflection of such traits.

5.2. Cessation

The results from cessation hazard models using local-brand cigarette prices are listed in Table 2a, and the corresponding price elasticities are presented in Table 2b. Covariates that are shown to delay quitting in youth are male sex, parents who smoke, having pocket money, and country GDP. As expected, the hazard of cessation increases with age. The sign of the *Price* coefficient has the expected positive sign in all specifications but is not statistically significant, failing to produce evidence that price has an effect on smoking cessation among youth. It is perhaps not surprising that it is difficult to pinpoint the effect of price on the quitting hazard of youth. The very young age of the sample may cast doubt on the validity of the responses to the quitting question, since very young smokers may mistakenly indicate a temporary non-smoking spell as quitting.

Cessation results by gender are shown in Table 6, and do not reveal differences by gender. Just as in the full-sample analysis, cessation rates are not responsive to changes in cigarette price for both males and females.

6. Testing for model specification and sensitivity analysis

One way to determine the goodness of fit of the parametric models is to compare the shapes of their predicted hazard functions of initiation and cessation to their nonparametrically-estimated counterparts in Figures 2a and 3a. Although a visual inspection is informative for the logit and split-sample models, a comparison between the semi-parametric Cox hazard function and the nonparametric hazard function is not useful because the two are a perfect match.[8] However, it is possible to test the appropriateness of the Cox model by testing if the proportionality assumption holds.

The proportionality assumption of the Cox model is examined using the test of Schoenfeld residuals. The idea behind this test is that if the proportionality assumption holds, then the residuals from the Cox model would not vary by failure time (Since all variation in the hazard with respect to failure time should already be explained by the specification, the residuals should be constant over time). The test of Schoenfeld residuals is essentially a test of the correlation between the Cox residuals and time.

[8] Recall that the baseline Cox hazard (which does not account for any individual covariates) is estimated nonparametrically much in the same way as the Nelson-Aalen method described in Section 4.1. Since the overall Cox hazard (which does account for individual covariates) is just a multiplicative function of the baseline, the overall Cox hazard has the same shape as both the baseline Cox hazard rate and the nonparametric hazard rate.

Under the null hypothesis of proportional hazards, the residuals are independent of time. The test statistics are listed in Tables 1a and 2a and show that the null hypothesis is rejected for both initiation and cessation. Although this result indicates that the proportionality assumption may not hold, it does not necessarily render the Cox model irrelevant. Since all Cox regressions in this analysis contain a time-varying covariate, *Price*, the proportionality assumption can be relaxed[9].

6.1. The shape of the initiation hazard

Figure 2a presents the nonparametric Nelson-Aalen approximation of the initiation hazard function without controlling for any individual or geographic differences in smoking patterns. It shows that the risk of initiation increases to almost 3 percent around the age of 13 and declines thereafter. As discussed above, a perfect fit to the shape of the non-parametric hazard function is provided by definition by the Cox model, so a graph of the Cox hazard is not separately shown. A graph of the hazard function predicted by the Logit model is shown in Figure 2b, and a graph of the hazard function predicted by the split-population model is shown in Figure 2c. In terms of fit, the semi-parametric Cox model is superior to the Logit and split-population models, both of which impose a specific distribution on the probability of initiation. Because of this

---

[9] By definition, in the Cox model the movement of each individual hazard over time (i.e., the shape of the hazard function) is determined by two factors: 1) the movement of the baseline hazard over time and 2) the change in individual covariates over time. If the covariates are fixed over time, the only factor to determine how the individual hazard changes over time is the baseline hazard. Since the baseline hazard function is the same for everybody, every individual hazard will follow the baseline hazard in exactly the same way over time and will therefore be proportional to the baseline and to each other at each point in time. If the covariates change over time, then the shape of the individual hazard with respect to time will be determined not only by the baseline but also by individual covariates. This means that the individual hazard functions would be truly individual, not merely multiples of the baseline – and therefore not necessarily proportional to the baseline or to each other.

distributional restriction, the initiation hazards predicted by the Logit and split-population models cannot resemble the nonparametric hazard as closely. As shown in Figure 2b, the Logit hazard function peaks before the age of 12 at about 2 percent and declines relatively sharply thereafter. The split-population hazard in Figure 2c has a slightly better fit to the data because it does not decline as sharply as the Logit hazard, but it still peaks earlier than the data says it should.

The conclusion regarding the relative performance of the initiation models as evidenced by the shape of the predicted hazard rates is as follows. If we believe the assumption that all people are eventual smokers, then the Cox semiparametric model is probably a better modeling option than the Logit parametric model. If we wish to relax this assumption, the split-population logistic parametric model is a very good alternative as it still provides a good approximation of the data. Since it is likely that the assumption of the unsplit models does not hold in the case of smoking initiation, the preferred model in this research is the parametric split-population model.

6.2. The shape of the cessation hazard

The nonparametric depiction of the cessation hazard is shown in Figure 3a. The observed hazard of cessation is relatively flat. It slightly peaks around 4 years into the smoking spell and stays flat at around 14 percent for another five years, after which it declines slightly. The cessation hazard predicted by the Logit model is shown in Figure 3b and does not have a very similar shape to the observed (nonparametric) hazard. It increases until about 11 years after smoking has started, reaching to over 20 percent probability of quitting. Figure 3c depicts he hazard function predicted by the split-population model. Its shape is not very similar to either of the other hazards. It peaks to a little below 20 percent at around 5 years into the smoking spell and declines relatively sharply afterward. However, in terms of resemblance to the nonparametric hazard, the split-population hazard is perhaps closer than the Logit hazard since its peak occurs at a more similar time.

The conclusion regarding the relative performance of the cessation models is similar to the case of initiation. The split-population model provides a reasonably good fit to the data and has the advantage of taking out the influence of those who will always smoke and for whom anti-smoking policy may not be effective anyway.

6.3. Sensitivity analysis

In the analysis of smoking initiation and cessation discussed so far, I use local-brand cigarette prices. The advantage of local-brand prices is that they are more likely to be considered when the average individual decides to consume cigarettes because local brand cigarettes are typically less expensive. However, there may be a concern about endogeneity of the *Price* variable arising from simultaneous determination of cigarette prices and cigarette demand. The use of micro-level data in this study considerably reduces the danger of such endogeneity because the smoking decision of a single individual could not affect market demand enough to change the price level. Certain characteristics of the local market demand, however, *can* influence the individual smoking decision by affecting the price level. For example, a weak market demand for cigarettes corresponds to higher cigarette prices, which in turn discourages individual smoking. Since market demand can affect both individual smoking decisions and prices, it can present another source of price endogeneity in the form of unobserved market characteristics. This article pays attention to this possibility by including country fixed effects, which can be interpreted as market fixed effects because cigarette prices have country-level variation.

Although using micro-level data and controlling for unobserved market heterogeneity is effective in addressing concerns about econometric endogeneity of *Price*, this research goes a step further in investigating the possibility by substituting the price of Marlboro cigarettes instead of local brand cigarettes. The presumption is that Marlboro prices are more exogenous to cigarette demand because they are imported. This is because the price of imported cigarettes contains a larger exogenous (not determined by market demand) component such as transportation costs, import duties, etc. This larger exogenous component makes Marlboro prices stickier and less vulnerable to changes in market demand than local-brand prices.

Results from the initiation hazard models using Marlboro prices, and the corresponding price elasticities are listed in Tables 3a-b. Cessation model results and price elasticities are in Tables 4a-b. In all specifications, the Marlboro price elasticities of initiation have the same statistical significance as the local-brand price elasticities but differ in magnitude. Marlboro price elasticities are smaller in absolute value, indicating that the decision to start smoking is less responsive to changes in the price of Marlboro cigarettes than to changes in the price of local brand cigarettes. This may be due to the perceived value of the Marlboro brand which makes demand less elastic. It can also be hypothesized that the reduction in the estimated price effect may be due to the stronger exogeneity of Marlboro prices which may have eliminated some of the bias from the estimate thus reducing its magnitude. However, the latter is not a likely explanation because substituting Marlboro prices in the analysis of cessation does not have the same effect on the magnitude of the estimated price coefficient as it does in the analysis of initiation. In fact, the price effect on cessation becomes larger (albeit not statistically larger) when based on Marlboro prices instead of local-brand prices. If substituting Marlboro prices truly helped remove endogeneity bias from the price coefficients, then it would have reduced the absolute value of the price effect for both initiation and cessation, not reduce it for one and increase it for the other. The increase in the price responsiveness of the quitting decision from substituting Marlboro prices is not statistically significant but may reflect a difference in the reasons for smoking Marlboros. It is possible that Marlboro smokers, in particular teenage Marlboro smokers, may smoke cigarettes for reasons other than tobacco consumption alone, for example as a status signal. In such case, addiction may be less of a factor, and smokers may be more inclined to quit when prices increase.

7. Conclusion

This paper investigates the impact of cigarette prices on smoking initiation and cessation among youth in developing countries. The price impact is identified by using country fixed effects to control for unobserved environmental characteristics such as anti-smoking sentiment. The estimation methods for the hazards of initiation and cessation include a discrete-time Logit model, a Cox model, and a split-population log-logistic model. Unlike the unsplit Logit and Cox models which assume that all subjects have positive hazards of initiation (cessation), the split-population model allows for the possibility that for some individuals the hazard is zero.

A statistically significant impact of cigarette price on the initiation (cessation) hazards is identified in the split-population analysis but not by either of the unsplit Logit and Cox models. The price elasticity of initiation is estimated at -0.165, so that a 10% increase in the price of cigarettes corresponds to about 1.7% decrease in the hazard of starting smoking. I find no evidence that price determines quitting which is possibly due to the very young age of the sample. The results in this paper indicate that once the influence of individuals who never smoke is controlled for, cigarette price policy can be effective in delaying the onset of smoking.

Bibliography

Allison, Paul. 1984. “Event History Analysis.” Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-046. Beverly Hills, CA: Sage.

Cawley, John, Sara Markowitz and John Tauras. 2006. “Obesity, Cigarette Prices, Youth Access Laws and Adolescent Smoking Initiation.” *Eastern Economic Journal* 32(1): 149-170.

Cawley, John, Sara Markowitz and John Tauras. 2004. “Lighting Up and Slimming Down: The Effects of Body Weight and Cigarette Prices on Adolescent Smoking Initiation.” *Journal of Health Economics* 23: 293-311.

DeCicca, Philip, Don Kenkel, and Alan Mathios. 2008. "Cigarette Taxes and the Transition from Youth to Adult Smoking: Smoking Initiation, Cessation, and Participation." *Journal of Health Economics* 27(4): 904-917.

DeCicca, Philip, Donald Kenkel, and Alan Mathios. 2002. "Putting Out the Fires: Will Higher Taxes Reduce the Onset of Youth Smoking?." *Journal of Political Economy* 110(1): 144-169.

Douglas, Stratford. 1998. “The Duration of the Smoking Habit.” *Economic Inquiry* 36: 49-64.

Douglas, Stratford and Govind Hariharan. 1994. “The Hazard of Starting Smoking: Estimates form a Split Population Duration Model.” *Journal of Health Economics* 13: 213-230.

Forster, Martin and Andrew M. Jones. “The Role of Tobacco Taxes in Starting and Quitting Smoking: Duration Analysis of British Data” *Journal of Royal Statistical Society*, Series A 164: 517-547.

Madden, David. 2007. “Tobacco Taxes and Starting and Quitting Smoking: Does the Effect Differ by Education?” *Applied Economics* 39: 613-627.

Nicolas, Angel Lopez. 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 Economics* 11(6): 521-535.

Ross, Hana, Frank J. Chaloupka, and Melanie Wakefield. 2006. "Youth Smoking Uptake Progress: Price and Public Policy Effects." *Eastern Economic Journal* 32(2): 355-367.

Schafer, Joseph L. and Maren K. Olsen. 1998. “Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst’s Perspective.” *Multivariate Behavioral Research* 33(4): 545-571.

Schmidt, Peter and Ann Dryden Witte. 1989. "Predicting Criminal Recidivism Using “Split Population” Survival Time Models." *Journal of Econometrics* 40: 141-159.

Tauras, John A., and Frank J. Chaloupka. 1999. "Determinants of Smoking Cessation: An Analysis of Young Adult Men and Women." NBER Working Papers: 7262.

Wooldridge, Jeffrey M. 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press: Cambridge, MA.

Figure 1a. Initiation hazard rates, by region

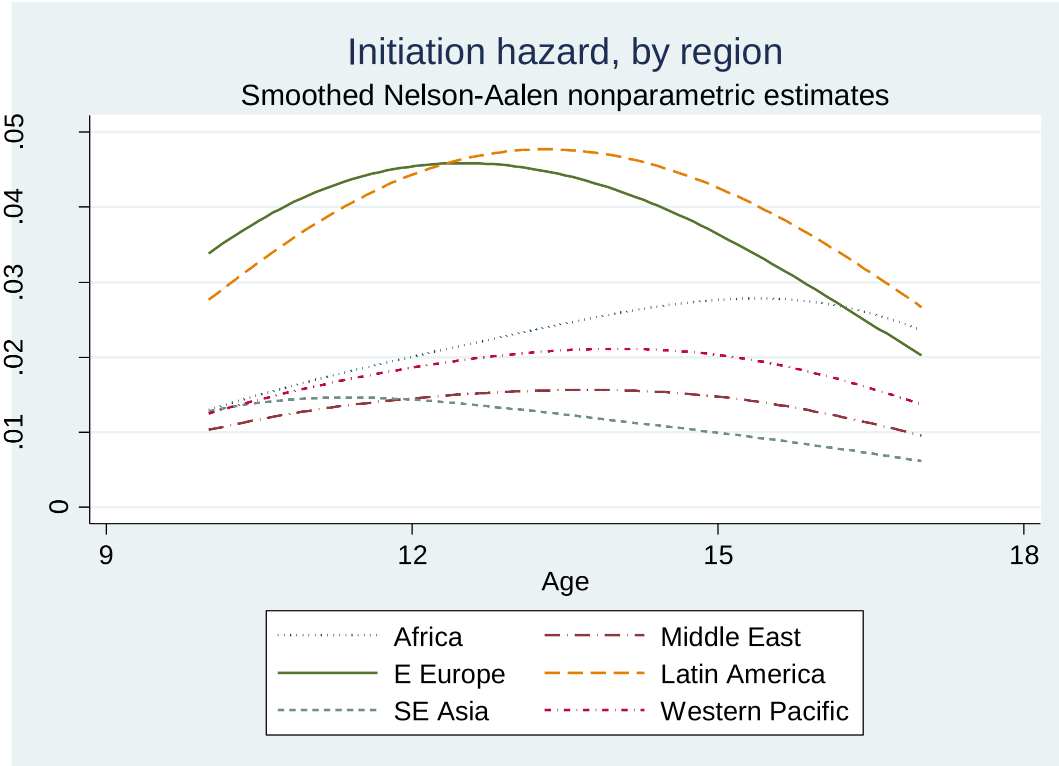

Figure 1b. Cessation hazard rates, by region

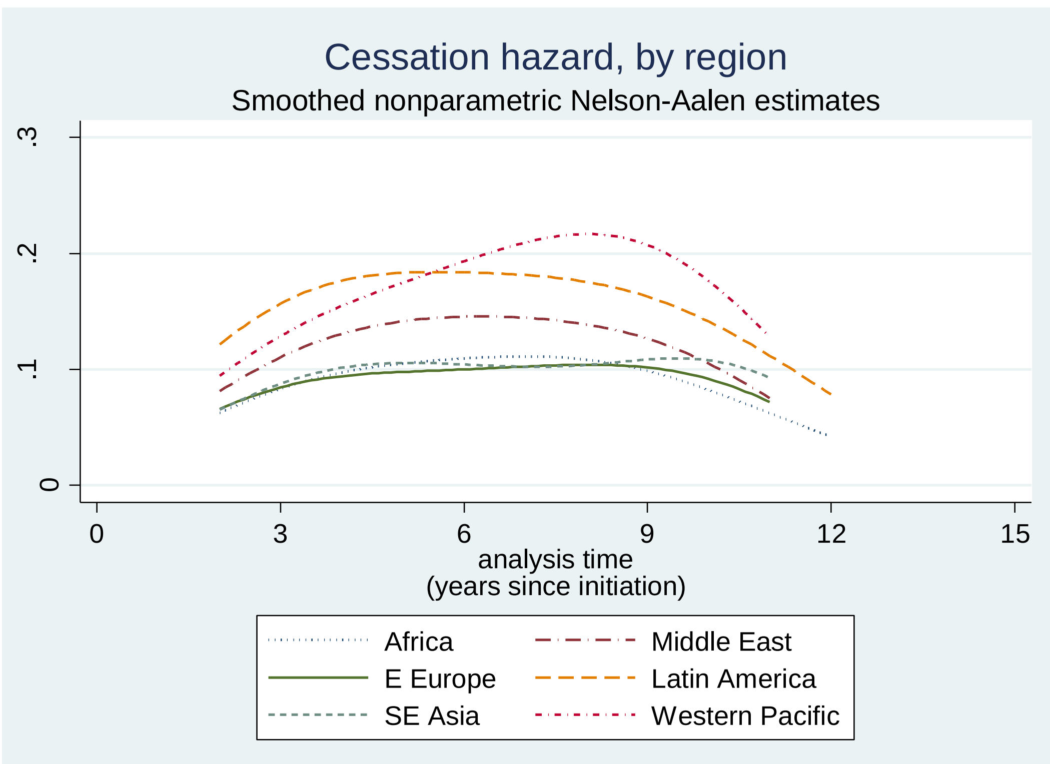

Figure 2a. Initiation hazard rate, non-parametric

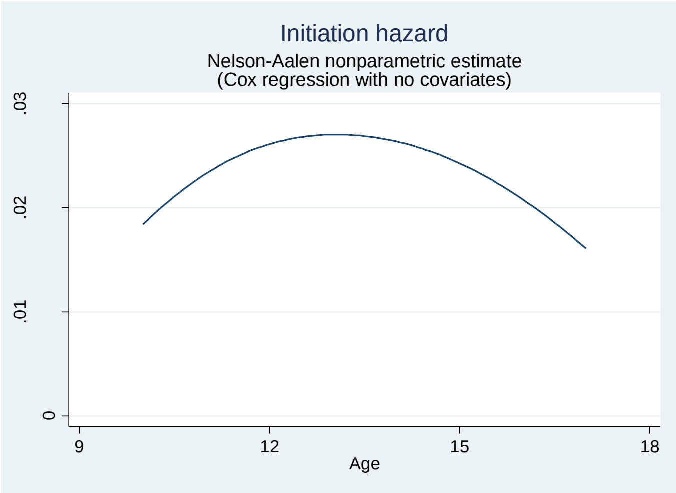

Figure 2b. Initiation hazard rate, discrete-time logit hazard model

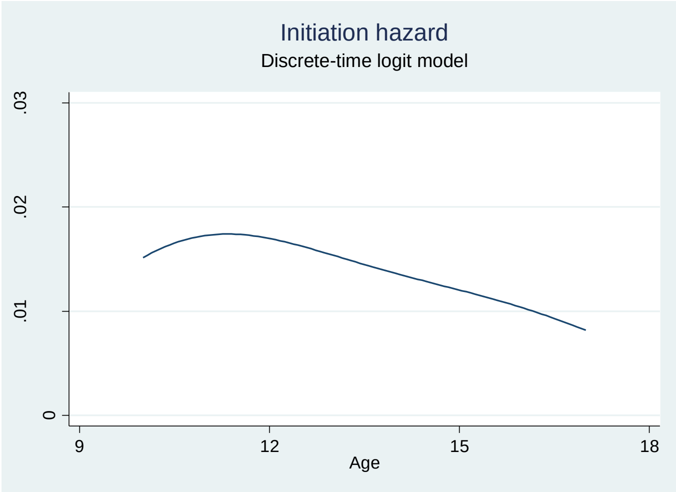

Figure 2c. Initiation hazard rate, split-population log-logistic hazard model

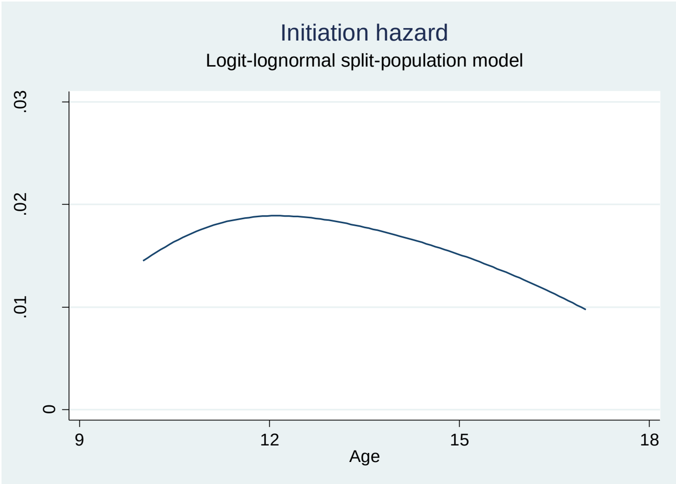

Figure 3a. Cessation hazard rate, non-parametric

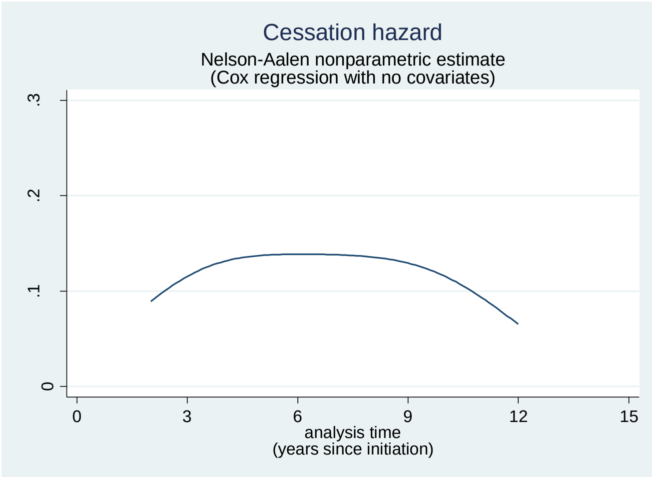

Figure 3b. Cessation hazard rate, discrete-time logit hazard model

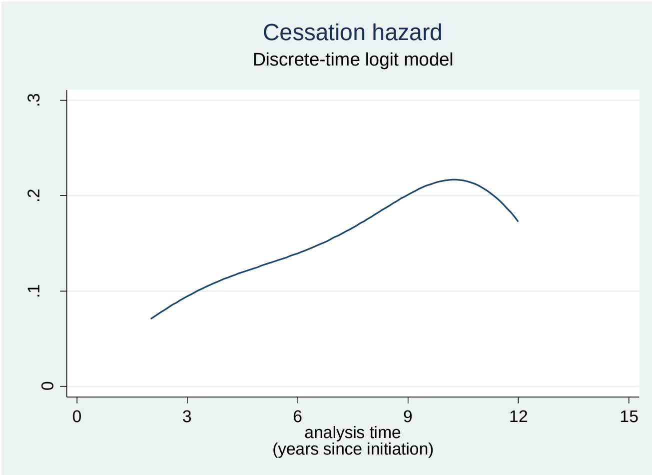

Figure 3c. Cessation hazard rate, split-population log-logistic hazard model

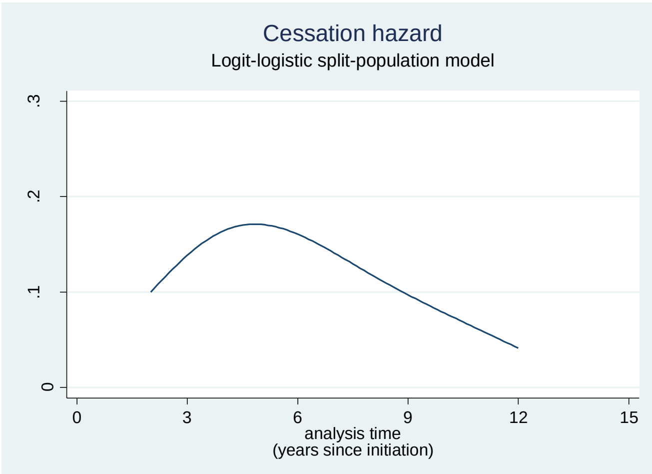

Table 1a. Sample means, individual variables

	Initiation sample (full sample)	Cessation sample (smokers only)
Age at start of risk	9.0	11.2
Age at failure or cessation	13.6	14.3
Male	0.50	0.61
At least one parent smokes	0.46	0.60
Has pocket money	0.63	0.82

Table 1b. Sample means, country variables

GDP per capita, 2000 US dollars	4,324
Price, local brand cigarettes, PPP-adjusted 2000 US dollars	2.36
Price, Marlboro cigarettes, PPP-adjusted 2000 US dollars	3.40

Table 2a. Initiation hazard models using price of local brand cigarettes

	Logit discrete-time hazard model (marginal effects)		Cox hazard model (coefficients)		Split-population log-logistic model (coefficients)	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (local brand cigarettes)	-0.007***	-0.002	-0.319***	-0.084	-0.617***	-0.140***
	(0.000)	(0.210)	(0.000)	(0.154)	(0.000)	(0.000)
Male	0.011***	0.011***	0.502***	0.530***	0.945***	1.047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Parents smoke	0.013***	0.010***	0.563***	0.481***	0.992***	0.868***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	0.018***	0.014***	0.881***	0.805***	1.424***	1.294***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.004***	0.003*	0.202***	0.210***	0.440***	0.410***
	(0.001)	(0.055)	(0.000)	(0.001)	(0.000)	(0.000)
			Chi2 test statistic of proportional hazards assumption			
			315,623	29,562		
			(0.000)	(0.000)		
Number of observations	2,333,179	2,333,179				
Number of subjects	418,753	418,753				
Number of initiations	67,330	67,330				

Table 2b. Price elasticity of initiation hazard (local brand cigarettes)

	Logit discrete-time hazard model		Cox hazard model		Split-population log-logistic model	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price elasticity	-0.716***	-0.204	-0.715***	-0.203	-0.752***	-0.165***

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend
Price elasticities are calculated by predicting the % change in the individual hazards from a 1% price increase, then averaging across observations

Table 3a. Cessation hazard models using price of local brand cigarettes

	Logit discrete-time hazard model (marginal effects)		Cox hazard model (coefficients)		Split-population log-logistic model (coefficients)	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (local brand cigarettes)	0.006	0.005	0.084	0.092	0.139	0.199
	(0.105)	(0.542)	(0.147)	(0.561)	(0.000)	(0.000)
Age	0.011***	0.009***	0.163***	0.136***	0.353***	0.294***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.014***	-0.010***	-0.236***	-0.199***	-0.527***	-0.474***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Parents smoke	-0.011***	-0.008***	-0.190***	-0.156***	-0.434***	-0.363***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	-0.023***	-0.022***	-0.345***	-0.366***	-0.766***	-0.835***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.001	-0.014**	0.009	-0.248**	0.046***	-0.651***
	(0.820)	(0.027)	(0.859)	(0.022)	(0.000)	(0.000)
			Chi2 test statistic of proportional hazards assumption			
			16,269	15,760		
			(0.000)	(0.000)		
Number of observations	273,690	273,690				
Number of subjects	65,932	65,932				
Number of cessations	30,050	30,050				

Table 3b. Price elasticity of cessation hazard (local brand cigarettes)

	Logit discrete-time hazard model		Cox hazard model		Split-population log-logistic model	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price elasticity	0.235	0.219	0.196	0.215	0.155	0.273

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend
Price elasticities are calculated by predicting the % change in the individual hazards from a 1% price increase, then averaging across observations

Table 4a. Initiation hazard models using price of Marlboro cigarettes

	Logit discrete-time hazard model (marginal effects)		Cox hazard model (coefficients)		Split-population log-logistic model (coefficients)	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (Marlboro cigarettes)	-0.004***	-0.001	-0.194***	-0.053	-0.367***	-0.046***
	(0.000)	(0.361)	(0.001)	(0.110)	(0.000)	(0.000)
Male	0.012***	0.011***	0.498***	0.530***	0.955***	1.046***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Parents smoke	0.014***	0.010***	0.570***	0.482***	1.021***	0.870***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	0.019***	0.015***	0.888***	0.805***	1.471***	1.296***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.003**	0.003**	0.150***	0.211***	0.362***	0.451***
	(0.015)	(0.027)	(0.008)	(0.001)	(0.000)	(0.000)
			Chi2 test statistic of proportional hazards assumption			
			701,092	32,629		
			(0.000)	(0.000)		
Number of observations	2,316,505	2,316,505				
Number of subjects	416,098	416,098				
Number of initiations	67,325	67,325				

Table 4b. Price elasticity of initiation hazard (Marlboro cigarettes)

	Logit discrete-time hazard model		Cox hazard model		Split-population log-logistic model	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price elasticity	-0.608***	-0.107	-0.604***	-0.175	-0.604***	-0.074***

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend
Price elasticities are calculated by predicting the % change in the individual hazards from a 1% price increase, then averaging across observations

Table 5a. Cessation hazard models using price of Marlboro cigarettes

	Logit discrete-time hazard model (marginal effects)		Cox hazard model (coefficients)		Split-population log-logistic model (coefficients)	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (Marlboro cigarettes)	0.002	0.008*	0.018	0.154	-0.004	0.227***
	(0.452)	(0.088)	(0.717)	(0.131)	(0.737)	(0.000)
Age	0.011***	0.008***	0.166***	0.134***	0.359***	0.296***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.013***	-0.009***	-0.228***	-0.200***	-0.507***	-0.476***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Parents smoke	-0.010***	-0.007***	-0.187***	-0.158***	-0.421***	-0.362***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	-0.021***	-0.020***	-0.346***	-0.363***	-0.762***	-0.834***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.001	-0.010*	0.010	-0.191*	0.051***	-0.588***
	(0.704)	(0.091)	(0.832)	(0.086)	(0.000)	(0.000)
			Chi2 test statistic of proportional hazards assumption			
			13,572	16,665		
			(0.000)	(0.000)		
Number of observations	271,500	269,945				
Number of subjects	64,955	64,955				
Number of cessations	29,503	29,503				

Table 5b. Price elasticity of cessation hazard (Marlboro cigarettes)

	Logit discrete-time hazard model		Cox hazard model		Split-population log-logistic model	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price elasticity	0.121	0.522*	0.055	0.509	-0.006	0.317***

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend
Price elasticities are calculated by predicting the % change in the individual hazards from a 1% price increase, then averaging across observations

Table 6. Initiation hazard models by gender.
Logit discrete-time hazard models (marginal effects)

	Male		Female	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (Marlboro cigarettes)	-0.009***	-0.002	-0.006***	-0.002*
	(0.000)	(0.406)	(0.000)	(0.056)
Parents smoke	0.016***	0.013***	0.011***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	0.024***	0.019***	0.013***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.004***	0.001	0.005***	0.005***
	(0.005)	(0.692)	(0.001)	(0.000)
Number of observations	1,145,696		1,187,483	
Number of subjects	209,292		209,461	
Number of initiations	26,399		40,931	
Price elasticity of initiation	-0.664***	-0.143	-0.854***	-0.317*

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend

Table 7. Cessation hazard models by gender.
Logit discrete-time hazard models (marginal effects)

	Male		Female	
	Without fixed effects	With fixed effects	Without fixed effects	With fixed effects
Price (Marlboro cigarettes)	0.008	0.005	0.005	0.004
	(0.123)	(0.670)	(0.133)	(0.527)
Age	0.016***	0.013***	0.009***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)
Parents smoke	-0.018***	-0.012***	-0.008***	-0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
Pocket money	-0.026***	-0.023***	-0.021***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.000)
Log GDP per capita	0.000	-0.029***	0.001	-0.009*
	(0.985)	(0.007)	(0.735)	(0.061)
Number of observations	173,187		100,503	
Number of subjects	40,059		25,873	
Number of cessations	17,316		12,734	
Price elasticity of cessation	0.239	0.222	0.223	0.172

Notes:
* p<.1, ** p<.05, *** p<.01
All specifications include a quartic calendar time trend