PATTERNS OF TOBACCO CONSUMPTION IN MEXICO – CURRENT PERSPECTIVE

by

Gustavo A. Ornelas-Almaraz

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

The University of East Anglia

September 2012

Gustavo A. Ornelas-Almaraz

September 2012

"Patterns of Tobacco Consumption in Mexico – Current Perspective"

Abstract

The aim of this study has been to investigate current issues regarding the consumption of tobacco products in Mexico.

The first chapter examines the incidence of the excise of tobacco using data from the National Income and Expenditure Household Survey of 2008. The results showed that, the excise is regressive taking the equivalent of 7.6% of consumption expenditure for smoking households in the lowest, and 3.3% from smoking households in the highest quintile. The unfairness of the excise on tobacco is confirmed by the calculation of the Kakwani index of progressivity which is estimated at -0.196. The results are valid for the 2008 tax schedule.

The second chapter investigates the effect of demographic, socio-economic and psychosocial factors influencing the demand for cigarettes in Mexico. The data used for the analysis comes from two sweeps of the Mexican Family Life Survey of 2002 and 2005-2007, a source individual-level data. A two-part model of cigarette demand is estimated. According to the estimation, a number of significant effects are found to determine the overall level of consumption in both sweeps.

In the third chapter a panel hurdle model is applied to data on cigarette consumption. The model has the feature of applying the Box-Cox transformation to the dependent variable in order to address the skew distribution seen in data. It also includes a non-zero correlation coefficient to account for the temporal linkage of consumption. The data used for estimation comes from the short panel of individuals created from the Mexican Family Life Survey. The results reveal that individuals who are unusually likely to participate in the activity of smoking tend to smoke less intensively. This is confirmed by the estimated correlation parameter which appears to be significantly negative. A number of significant effects are found to determine the overall level of consumption overtime.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my family, specially my parents for their unconditional love and support throughout my studies. Also, I would like to thank the LDS community at the Norwich and Dereham Ward for making me feel like home, especially Mark and Elain Anderton, Pai Masendu and her family. I would like to thank my supervisors along my postgraduate studies: Dr. Bibbas Saha, Dr. Peter Moffatt and Dr. Sara Connolly for their enormous patient and support. I am also in great debt with Dr. Richard Baker. His support has been greatly important in many aspects of this journey. The financial support by the Mexican National Council for Science and Technology (CONACyT) is deeply acknowledged. All the remaining mistakes are entirely my fault.

Introduction

"For thy sake, Tobacco,
I would do anything but die."

- Charles Lamb,
A Farewell to Tobacco, 1830

The World Health Organisation (WHO) has warned about the dangers that are posed to public health by the consumption of tobacco, particularly the smoking of cigarettes. Given decades of scientific research, it is recognised that tobacco use is a leading preventable cause of illness, impoverishment and death. Recent figures from the same organisation indicate that smoking not only kills nearly six million people world-wide, each year, who smoke, but that a further 600,000 people are also harmed through their exposure to tobacco smoke; this is known as second-hand smoking[1]. Individuals who breathe in other people's smoke (passive smokers) are of particular concern as they are also at risk of developing life-threatening illnesses. In all, it is predicted that smoking will kill up to eight million people by 2030[2]. As a consequence, the consumption of tobacco products has become a global concern and it has been classified now as the "tobacco epidemic".

The world consensus of tackling the epidemic has made possible the adoption in 2003 of the Framework Convention on Tobacco Control which is organised by the WHO (WHO FCTC). The creation of the WHO FCTC is, perhaps, one the more important events in recent years of the "tobacco wars" since the publication in 1964 in the US of the 1st Surgeon General's Report on the health hazards of cigarette smoking. To date, 168 nations have signed the WHO FCTC. This framework has laid the foundations for implementing more effective and coordinated tobacco-control policies[3]. It is hoped that, if such policies are enacted, cultural attitudes will change, leading towards the reduction, or even the eradication of the consumption of tobacco. In this context, there is no doubt that the epidemic constitutes a challenge to public health particularly for developing countries where the market is growing. Therefore, the adoption of the WHO FCTC by the Mexican government in 2004 has been perceived as an important step forward in tackling this issue[4]. Mexico is the study case of this thesis.

One efficient way of regulating the consumption of tobacco is through legislation. It expresses predetermined values in society and regulates private actions. However, to legislate on

[1] Tobacco – Fact sheet No. 339, World Health Organisation, May 2012.
http://www.who.int/mediacentre/factsheets/fs339/en/index.html
[2] Ibid.
[3] The convention took effect on 27 February 2005, 90 days after it had been acceded to, ratified, accepted, or approved by 40 States. For more information visit the WHO FCTC's web page:
http://www.who.int/fctc/en/
[4] Mexico signed the convention on 12 August 2003. It was ratified by the Senate on 28 May 2004 and took effect on 27 February 2005.

specific issues creates challenges. Sometimes, there is limited public knowledge about the problem being addressed but also it tests the capacity of institutions such as national and state government departments and health service providers to achieve the goals being pursued. One difficulty has been, for example, the opposition of the tobacco industry which has seen its activities deemed as undesirable. Consequently their business has been negatively affected resulting in a displacement of their activities from developed countries towards developing countries. Therefore, it is important to emphasise that the one and only objective being pursued by regulating the tobacco market is the reduction in the consumption of cigarettes and other tobacco products. Of course, the efficacy of any legislation depends on the institutional capacity of each country, especially, when it comes to following each of the recommendations stipulated by the WHO FCTC.

Institutions have the ability to change the culture of a society. In fact, over the years it has helped to change the view that the consumption of tobacco is “normal” to a view that described it as “socially unacceptable” and “addictive”. Likewise, there is no doubt that the interventions of governments have been crucial in the crusade against tobacco. Without these, it is impossible to reach different groups in society. However, in order to achieve changes of individual behaviour which would oppose smoking, society must change the way tobacco is promoted, sold and consumed.

Cigarette smoking results in a significant health hazard. Therefore, cessation should be analysed in a wider context which would consider smoking a preventable disease. Public actions that promote cessation amongst the youth and the adult population may bring substantial benefits both in the short and the long run and may do so more effectively than any other tobacco-control policy. Individuals who quit smoking reduce the risk of dying prematurely and it also reduces costs of health care in the future. For this reason, the promotion of the reduction of the consumption of cigarettes is the best public health policy.

Mexico, the first Latin American country to ratify the WHO FCTC, has made a decisive commitment to implementing policies for the control of tobacco. Although existing legislation was already in place, the need to strengthen those laws and to move towards more restrictive policies have been the main goal of the government. For instance, tax reform in the early 2000’s that ended the tax differential between filtered and unfiltered cigarettes and has been taken as an important change in the way the government views the tobacco industry. Other measures followed such as the elimination of any cigarette advertisements in the media and the introduction of health warning labels on all tobacco products. Moreover, the creation of smoking cessation clinics and the application of a national survey to monitor the tobacco epidemic (and other addictive substances) have been part of a more integral action supporting the reduction of

tobacco consumption in the country. Notwithstanding, the need for more effective policies has been a concurrent issue in Mexico. Thus, more research is needed to enhance the understanding of tobacco consumption and the effect on society of its control in order to provide further insights for successful anti-smoking policies.

Aim and Objectives of this Thesis

The aim of this research is twofold. Firstly, to provide a normative assessment of the effects of the tobacco excise in Mexico; this tax is the country's leading tobacco-control policy. Secondly, to understand non-price factors influencing the demand of tobacco by using individual-level data.

More specifically this thesis has three main objectives:

1. To provide an answer to whether the tax imposed upon cigarettes is a regressive tax.
2. To investigate the effect of socio-economic and psychosocial factors in determining the demand for cigarettes.
3. To extend the previous objective (given that the vast majority of demand analysis' research has been limited to cross-sectional analysis) this research investigates the demand for cigarettes by estimating the Cragg's model within a panel data context.

The contribution of this research may, on the one hand, improve the understanding of the effects of tobacco taxation on the welfare of Mexican households. Although the fairness of this tax is questioned, any reduction or elimination of this tax is unlikely. However, it is hoped that this research will change the view in regards to its effect and a more effective and efficient tax policy may be developed. This research also aims to achieve an understanding of the ways in which demand is determined. In particular, it aims to understand factors influencing the decision to smoke and, upon smoking, the level of consumption. This may help to formulate specific anti-smoking campaigns.

The Structure of the Thesis

The thesis is organised in six sections:

1. Introduction
2. Literature Review
3. First Chapter: The distributional effects of the tobacco excise in Mexico.
4. Second Chapter: An analysis of adult cigarette demand in Mexico.
5. Third Chapter: A panel double-hurdle approach for modelling cigarette demand in Mexico.
6. Conclusions.

Literature Review

This literature review will fall in three sections:

1. "Smoking Prevalence and Tobacco-Control in Mexico". This section aims to place the context and motivation for conducting this research. By doing so, key figures of the tobacco epidemic are given from Mexico and from other countries to draw international comparisons. An overview of the current state of tobacco-control policies in the country is also provided.
2. "The taxation of tobacco and Mexico". This section reviews the reasons for the government to intervene in the market of tobacco. An emphasis is given on the government's preferred tobacco-control policy, taxation, and the fairness of the excise of tobacco in the tax system. A summary of the objectives being pursued by imposing an excise is given together with some emerging issues regarding its design.
3. "The analysis of the demand for tobacco". This section reviews key findings of a growing trend of economic research that looks to investigate the role non-price factors in determining the demand for cigarettes. The emphasis is given on studies that rely on individual-level data. Advantages and limitation of these studies are discussed. A review of findings from traditional demand analysis is also provided.

Smoking Prevalence and Tobacco-Control in Mexico

Smoking Prevalence

Statistical information on the prevalence of smoking has been collected periodically since 1988 with the implementation of the National Addiction Survey (*Encuesta Nacional de Adicciones*, ENA in Spanish).This is a cross-sectional household survey which collects information on consumption habits of tobacco, alcohol, and other addictive substances (illegal drugs) amongst people between the ages of 12 and 65. So far, information from five sweeps of the survey is available which correspond to the years of 1988, 1993, 1998, 2002 and 2008. This has helped to monitor the epidemiology of addictions in Mexico which has allowed the government to formulate public policies beyond the issue of smoking.

The most recent survey (ENA-2008) showed that almost 14 millions people are smokers in the country representing 18.5% of the population; 17.1% are considered former smokers (quitters) and 64.4% are non-smokers. The same source of information confirmed that the prevalence of smoking is much higher amongst men (27.8%) than amongst women (9.9%).

However, changes on the survey's design across sweeps has made difficult to make comparisons over time, particularly when it comes to the prevailing rate of smoking which is a key

figure of interest in this analysis. One of the most important changes was that, despite the fact that ENA has been implemented since the late 1980's, it was only from the year 2002 that rural population were included in the sample. On the other hand, the definition of what constitutes a "smoker" also has changed across different versions of ENA. For instance, in regards to the definition of a "smoker", it is possible to detect important differences in the last two sweeps of the survey. The ENA of the year 2002 (ENA-2002) defines it as follows:

- Any person who reports having smoked in the last 12 months.

Under this definition, the survey showed that more than 16 million people between the ages of 12 and 65 smoked actively in Mexico representing 23.5% of the population. Furthermore, 17.4% of the population are identified as former smokers and the remaining 59.1% of the population as non-smokers. Stratifying the figure of smokers by gender the survey revealed that more than 11 million men (36.1%) and more than 5 million women (13.1%) smoked. In contrast, the survey of 2008 (ENA-2008) considers a smoker as follows:

- An individual who had smoked some time in his/her life and having smoked in the last 12 months.

Changes in the definition of a smoker and the inclusion of the rural population in the last two surveys have made it difficult to conclude, with certainty, whether the trend of the number of smokers has decreased, increased or has remained stable the last twenty years.

Bearing in mind differences on the survey's designs, additional characteristics of the smoking population are worth mentioning. For instance, in 2002, 26.4% of the urban population were considered active smokers whilst only 14.3% of the rural population were considered as such. Moreover, active smokers are largely concentrated in the age range of 18-29 years old representing 36.2% of the total. In fact, 66.2% of all smokers are younger than 40 years old. This includes individuals who are younger than 17 years old which account for 7.5% of all smokers. In terms of education, the majority of smokers tend to have at least basic education. 58.8% of all smokers have at least either primary or secondary whilst 36.0% have at least high-school education or more.

In 2008, 20.4% of the urban population (almost 12 million people) are active smokers whilst 11.3% of the rural population were considered as such. Official tables of the survey in that year do not break figures of active smokers by age-ranges. It reported, however, that 20.6% of the adult population and 8.8% of all teenagers (12 – 17 years olds) are active smokers. In terms of education, the survey revealed that amongst the smoking population, similar characteristics can

be identified as with data from the year 2002. In fact, only 5.2% of all smokers do not have formal education[1].

Another important key figure to take into account is the rate of second-hand smoking or passive smoking. A passive smoker is a non-smoker who is involuntarily exposed to the fumes of tobacco. The survey of 2002 showed that almost 18 millions people are exposed to such fumes; more in the urban areas (78.8%) than in rural areas (21.2%). Worryingly, almost 30% of passive smokers are teenagers (the largest proportion)[2]. Interestingly though, male teenagers are the most exposed; 43.8% of them are passive smokers whilst this contrast with female teenagers of whom only 23.2% are thus exposed. However, amongst females, those aged 18-29 years old are the most affected; 28.1% are passive smokers.

In turn, figures from ENA-2008 showed that almost 11 million people are exposed to the fumes of tobacco. Similarly as the previous survey, males are more exposed to second-hand smoking (25.5%) than females (22%) and teenagers are more affected by second-hand smoking (27.3%) than adults (22%).

Smoking Households

Another way to monitor the tobacco epidemic is by tracking the level of consumption in the country. However, one of the main challenges is the availability of suitable information. For example, time-series consumption data obtained from the national accounts' records is available. This information constitutes the national production of tobacco products plus net imports. Nevertheless, this information is unreliable and so it is of little use for analysing the trend of consumption over time. For instance, data is not reported periodically, or the information is incomplete or missing for some years [see Saenz de Miera, Jiménez Ruiz et al. (2007) for a further discussion]. Therefore, the obvious source of information to monitor consumption has been the National Household Income and Expenditure Survey (ENIGH in Spanish) undertaken by the National Institute of Statistics and Geography. This survey will be described in more detail in the next section. Nonetheless, in terms of the consumption of tobacco products, the ENIGH collects information on the amount of money that households report spending on cigars, cigarettes and loose leaf tobacco. ENIGH also collects information on quantity being consumed which is available as kilograms rather than number of cigarettes, and price is available as the price paid per kilogram of tobacco. This has made it possible, under restricted assumptions, to calculate estimates of

[1] Perhaps, not surprisingly, in rural areas 73.7% of all smokers have either primary or secondary education. However, 15.2% of all smokers in rural areas do not have formal education.
[2] It seems likely that the number of passive smokers under 18 years old is considerably higher as the survey limits it study only to individuals between the ages of 12 to 65 years old.

"quantity" and "unit prices" for empirical work. Therefore, studies which have used ENIGH as the primary source of information have provided some insights about the determinants of tobacco demand. In particular, some estimates of price sensitivity in Mexico have become available (see below) and these appear to be consistent with other international experiences.

However, the interest here is in studying those households which reported any consumption of tobacco. These are defined as "smoking households" and research has been conducted on the changes of the prevalence over the period of time. Published studies have shown that the prevalence of smoking households have decreased over the years and so the consumption expenditure on tobacco products. For example, Vázquez-Segovia, Sesma-Vázquez et al. (2002) have estimate the "adjusted prevalence rate" of smoking households using seven sweeps of ENIGH between 1984 and 2000[3]. They found a general decrease, from 21% in 1984 to 9% in 2000, in the number of households who reported positive consumption during the period analysed. These estimates were carried out using logistic regressions which included controls for the gender and schooling of the household head and by income decile, and whether the household reported alcohol consumption.

However, findings from the previous study should be taken with caution. Their results did not take into account that versions of ENIGH prior to 1994 are not comparable with those after 1994. Notwithstanding, follow-on studies do seem to confirm a downward trend of the prevalence of smoking households.

Sesma-Vázquez, Campuzano-Rincón et al. (2002) analysed the consumption of tobacco in Mexico over a shorter and earlier period covering the years 1992 to 1998. They also analysed the adjusted prevalence of smoking households using logistic regressions. Their model included controls for gender and schooling of the household head, household size and the sample is stratified by income quintiles and the size of the area (urban or rural). They obtained similar results to Vázquez-Segovia et al. (2002) reporting a fall of the prevalence of "smoking households" from 22.4% in 1992 to 9.9% in 1998.

Miera-Juárez, Jiménez-Ruiz et al. (2007) analysed the prevalence of smoking households, over an eleven year period starting from 1994 to 2005. In addition, this study included information from the ENIGH undertaken during 2005 which was recently made publicly available[4]. Nonetheless, some findings appear to be consistent. The prevalence of smoking households appears to be in decline during the period analysed. This finding appears consistent with the

[3] The years in question are 1984, 1989, 1992, 1994, 1996, 1998 and 2000.

[4] Versions of the ENIGH prior to 1994 are 1984, 1989 and 1992. From 1994, versions available at the time of writing this paper are 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008 and 2010. The 2005 version was supposed to be restricted to government access although eventually it became in the public domain.

results of previous studies described. In 1994, 11.2% of all household reported consumption of tobacco but an overall decline is observed to reach 7.8% of households in 2005.

Alternative Sources of Information and International Comparisons

There are other sources of data that have estimated the rate of smoking prevalence in Mexico. For instance, the Health National Survey of 2000 (*Encuesta Nacional de Salud*, ENSA in Spanish) showed that 21.5% of the population can be considered smokers. In particular, the survey found that 33.7% of men and 10.1% of women smoked. It is worth noting that ENSA follows the internationally accepted definition of smoker given by the WHO. Such institution defines a smoker as:

- Any individual who smokes regularly, and had smoked at least 100 cigarettes at the time of the survey.

A more recent version of survey showed that in 2006, the rate of smoking prevalence amongst adults is 18.9% which stratified by gender, 21.6% of the male population smoked whilst only 6.5% of the female population did so[5]. This is reported in Waters, Saenz-de-Miera et al. (2010). It is worth noting, however, that the rate provided by ENSA is only valid for the population over 20 years old of age.

One additional source of information is the Global Tobacco Surveillance System (GTSS). This "surveillance system" has been developed by the WHO and the US Centers for Disease Control and Prevention (CDC) to track tobacco use across countries using a common methodology (Valdés-Salgado, Lazcano-Ponce et al. 2005). With such information, it has been possible to draw international comparisons on the incidence of smoking and some key figures are presenting here.

The GTSS produces data from four different surveys: a survey that focuses on adults aged 15 years and older (Global Adult Tobacco Survey, GATS), on youth aged 13-15 (Global Youth Tobacco Survey, GYTS), teachers and administrators from the same schools that participate in the GYTS (Global School Personnel Survey, GSPS) and a survey that focuses on 3rd year students pursuing degrees in dentistry, medicine, nursing and pharmacy (Global Health Professions Student Survey, GHPSS).

In Mexico, one survey of the GATS is available corresponding to the year 2009 and one survey of the GHPSS corresponding to the year 2006 and these are nationally representative surveys. Versions of the GYTS are available only for selected cities within Mexico. In the year of 2000 for example, data has been collected only from one major city (Monterrey). By 2003,

[5] In 2006, the name of the survey changed to National Health and Nutrition Survey, ENSANut in Spanish.

samples for 10 cities were collected whereas for the period 2005-2006 samples were collected from 22 cities.

Focusing only on the adult population, the GATS of the year 2009 showed that 10.9 millions people smoked representing 15.9% of the adult population in Mexico of whom 8.1 million are men (24.8% of all men) and 2.8 million are women (7.8% of all women)[6]. The rate of prevalence is comparatively similar to countries of the same income levels such as Brazil where 17.2% of the population smokes. In India the rate of smoking prevalence is 14.0% whilst in Turkey is much higher at 31.2%. In contrast, the proportion of smokers relative to the adult population in Mexico is much lower than in some high-income countries such as the UK where smoking prevalence is around 24% of the population or in Germany where it is 29%.

Tobacco Control Policies in Mexico

The government's concern on the consequences of tobacco consumption is, however, not new in Mexico. Almost three decades ago the Mexican government created the National Council Against Addictions (CONADIC in Spanish) which main task is to define and conduct the national policy on tackling the addiction of illegal drugs and legal ones such as alcohol and tobacco[7]. However, given the recent emphasis on addressing the tobacco epidemic, the federal government in conjunction with state governments have worked together to implement more effective tobacco controls under the five policy-reduction provisions of the WHO FCTC. As summarised by WHO (2003) these include:

a) Banning smoking in public places
b) Prohibiting cigarette sales to minors
c) Banning cigarette advertisements
d) Enlarging the health warning label on cigarette packages
e) Increase tobacco taxes.

From the year of 2001, measures to contain the epidemic were agreed which, under a national framework, it set up three main courses of action: prevention, treatment and direct tobacco-control policies. The following policies which have been enacted so far include:

[6] ENA-2008 also provides an estimation of the rate of smoking prevalence following the international definition of a smoker. It calculates that 17.2% of the populations can be considered as a smoker which appears somewhat consistent with the estimation a year later if a downtrend is accepted.
[7] For more information visit the web page: http://www.conadic.salud.gob.mx/

I. Prevention

Laws and norms with the intention to protect the rights of non-smokers have been enacted. In this context, the main aim has been to regulate the consumption of tobacco in closed spaces. In other words, to promote areas free of smoking. Direct results of this policy include:

- Primary and secondary schools are places free of smoking.
- Government buildings (either from the federal or state government) are also places of smoke-free.

However some areas have been exclusively reserved for non-smoking even before new regulations were discussed. For example, hospitals and health clinics have been free smoking areas. Lastly, as an additional preventive measure is the recognition of the "World No Tobacco Day" which WHO made it official since 1988 every 31st May.

II. Treatment

The creation of help centres to provide medical and physiological treatment to those affected by addiction. Actions focusing on treatment also included the re-training of health workers which will enable them to provide the required counselling. It is worth mentioning that since 1998, health workers in all hospitals or similar institutions are obliged to interrogate any patient on smoking habits so that treatment may be recommended.

III. Direct Tobacco-Control Policies

New laws have been reformed which aim to strengthen the pre-existent health regulations which had controlled the trade of cigarettes. This is because evaluations of these laws showed serious enforcement problems. For instance, the access of tobacco to minors has been restricted since health related laws were enacted in 1984. Notwithstanding, the sale of cigarettes to minors, mainly from corner shops, is a wide spread phenomenon which has been socially tolerated and related with economic interests Kuri-Morales, Cravioto et al. (2000), Kuri-Morales, Cortés-Ramírez et al. (2005) and Kuri-Morales, Cortés-Ramírez et al. (2006). Another example is the regulations on packaging tobacco products. Pre-existing laws had stipulated the inclusion of health warning labels on each package of cigarettes. However, these labels were too small and printed in a colour which made them poorly visible to consumers. Therefore regulations to control the consumption of tobacco in Mexico have been reformed or augmented consistently since the period of 2001-2002.

Other specific and direct actions taken following the implementation of the WHO FCTC include the prohibition of any advertisement of tobacco products on radio, TV, magazines, newspapers or any printed material. However, it will be legal to advertise tobacco at the cinema as long as the movies is of classification “C” or “D” therefore for adults. Advertisement is also restricted in the Internet though it has the provision that if technology is in place that would restrict the access to minors, the promotion of tobacco is possible.

Regulation on advertising also includes specific guidelines regarding the promotion of the product itself. For example, it is forbidden to create ads targeting minors or to create ads which may be appealing to them. This includes the prohibition of using cartoons or animated pictures and the ban of presenting celebrities in their campaigns. Furthermore, any advertising must not suggest, in any way, a sense of success such as athletic, sexual or professional success by using tobacco. Therefore, ads promoting tobacco products must target adult smokers and they explicitly must exclude non-smoker adults and minors.

In terms of health warnings labels on tobacco products, these must cover an area of at least 25% of the packaging. Labels must include information on where the consumer can have access to information. Specific health warning labels which are common to be present on packages of cigarettes includes “currently there is not a cigarette that reduces risks to the health” or “there are good reasons to quit smoking - the consumption of tobacco causes different types of cancer […] and it is recommended to stop smoking. Request professional help to the following number: 01 800 …”.

One of the new regulation imposed the prohibition of the selling and distribution of cigarette packages containing less than 14 cigarettes, or if loose tobacco, less than 10 grams. Moreover, the distribution of free samples has been forbidden for minors and non-smokers. Furthermore, it has been prohibited any exhibition of advertisements less than 200 metres from schools (kindergartens, primary, secondary and high schools), parks and hospitals. Additionally, if any advertisement boards are to be made, their size should not exceed an area of 35 square metres. In addition, tobacco companies are allow to sponsor only events in which all participants are adults or when all individuals who are sponsored are adults.

Lastly, the use of taxation is perhaps one of the most significant tobacco-control measures taken by the Mexican government. It consisted in increasing the level of excise on tobacco to international standards. The first step focused on closing the gap on the level of excise that prevailed between filtered and unfiltered cigarettes. When this was achieved in 2005, subsequent increases in the excise rate followed. Particularities of this policy will be outlined in some detail later in this review as this is particular of interest for this analysis. However, it is worth mentioning

that policies have continued to be strengthened and augmented at the time of writing this thesis. The activism of the medical community, in particular, has made possible to continue pressing for tougher regulations including higher taxes.

There have been signs that regulations have improved the panorama of smoking in Mexico. However, it will take time before substantial changes on the rate of smoking prevalence and level of tobacco consumption become evident. For instance, an environment which reinforces positive attitudes towards smoking seems to be the most relevant barrier in reducing the prevalence of smoking in Mexico, especially amongst the youth. This is revealed by the study of Valdés-Salgado, Thrasher et al. (2006) which analyse different characteristics of students who declared having smoked. Using the GYTS of the years 2003 and 2006, they found that teenagers who live with parents who smoke have greater possibilities to become a smoker especially if both parents do so, relative to teenagers whose parents do not smoke. Surprisingly, if the mother is the only parent who smokes, the probability of becoming a smoker is greater (2.2 times) than if only the father does (1.6 times). In addition, teenagers who have some friends who smoke are two times more likely to smoke themselves, but 3.2 times more likely if all their friends smoke. All these reveals the importance of acceptability of smoking in which teenagers live that influence the decision to smoke.

Two additional findings are worth mentioning by the study of Valdés-Salgado, Thrasher et al. (2006). Teens who admit having consumed other tobacco products are 2.7 times more likely to smoke cigarettes which reveal the addictive power of tobacco. However, the access to pocket money is positively related to the consumption of cigarettes. In fact, the researchers found that those who have between 50 to 100 MXN pesos a month for pocket money are 5.6 times more likely to smoke in comparison with those who cannot afford it.

Within two years of the implementations of the WHO FCTC in Mexico, no major improvement in critical areas of tobacco control had been achieved. This is confirmed by the study of Valdés-Salgado, Reynales-Shigematsu et al. (2007) who analysed two surveys of the GYTS also from the years 2003 and 2006. In particular, they found that it has not been possible to reduce the prevalence of smoking amongst teenagers. According with the survey, 25% of secondary students admitted having smoked and such rate has not changed in either version of the survey. Furthermore, the susceptibility to smoking amongst non-smokers has not changed either. The study by Valdés-Salgado, Reynales-Shigematsu et al. (2007), however, provides some explanations for this. For instance, the exposition to second-hand smoking, particularly at home and in public places did not decrease. In fact, most of the students interviewed (80%) continue to perceive directly or indirectly advertisements of tobacco products. In addition, and perhaps more importantly, the access to cigarettes to minors continue to be a reality. Students interviewed

declared that being underage is not an impediment to acquire cigarettes from convenient stores. The only tangible progress of the convention, however, is the implementation of anti-smoking campaigns in the schools by integrating information on the dangers of smoking in the school curricula.

Tobacco Taxation & Mexico

Market Failures in the Tobacco Market and Government Intervention

Free market economists invoke the principle of consumers' sovereignty as a necessary condition for the well functioning of markets. This principle suggests that consumers make the best choice in their decisions by maximising their benefit under conditions of scarcity or of a limited budget. If this principle is true, interference in the market by institutions such as governments is not necessary. However, the principle of consumers' sovereignty relies on two important assumptions. Firstly, all consumers are completely informed about the costs and benefits of their choices, and secondly, all consumers assume all costs associated with their choices and do not impose costs onto others. The consumption of tobacco violates these two assumptions.

Market Failure in the Market of Tobacco

Consumers are not completely informed about the costs and benefit of their choices. This is rarely seen, in particular, for the case of teenagers. The initiation of smoking usually starts during the teenage years. According with ENA-2008, the age of smoking initiation is 13.7 years old amongst teenagers and 16.8 years old for adults. In the past, however, smoking was reserved only for adults. According to Franco-Marina and Lazcano-Ponce (2010), the age of smoking initiation in Mexico for individuals born in the 1930's was 20.6 years old whilst for those born during the period of 1975 to 1978, the initiation age decreases to 16.5. This is of particular significance because according with the information collected by GYTS, most young smokers do not have precise information regarding the health risk associated with smoking thereby they lack the capacity to recognise the addictive power of cigarette smoking.

Consumers do not assume the risks and costs associated with the consumption of tobacco. It is widely recognised that smoking imposes physical and financial costs to the individual and the society as a whole. These costs are associated with the exposure to tobacco smoke, the loss of labour productivity due to the risk of premature death and the medical costs attributable to the health care treatment of illnesses related with smoking.

Regarding health care costs, a considerable amount of research has been conducted. Although all related studies face important methodological issues, given that it is almost impossible to quantify all cost associated with the treatment of a disease, tentative estimations provide a valuable reference about the magnitude of the tobacco epidemic. Lightwood, Collins et al. (2000) found that in high-income countries, the gross healthcare cost of tobacco use may range between 0.10% and 1.1% of the gross domestic product (GDP). The availability of studies from low and middle-income countries is scarce but they reckon that the costs could be comparable to those found for high-income countries.

Some studies on the healthcare cost associated with tobacco use have recently emerged for Mexico. Reynales-Shigematsu, Rodríguez-Bolaños et al. (2006) estimated the cost of medical care for four diseases which are usually attributable to tobacco consumption. These are acute myocardial infarction (heart attack), cerebrovascular disease (hardening of the arteries in the brain), chronic obstructive pulmonary disease and lung cancer. In their study, the focus of analysis is on patients of at least 35 years old of age diagnosed with such diseases between the years of 2001 and 2004 at *Instituto Mexicano del Seguro Social* (IMSS). IMSS is the main public health care provider in Mexico with covers its affiliates, their direct relatives and pensioners; their services are provided for 44% of the Mexican population.

The methodology employed was the "direct cost approach" which involves calculating the value of each treatment needed when a disease is diagnosed. These included basic medical care and rehabilitation. Therefore, the total cost incurred for treating a medical condition is the sum of all unit costs of all treatments needed. Given this methodology Reynales-Shigematsu, Rodríguez-Bolaños et al. (2006) found that the overall cost for treating all the diseases studied is almost 12 billions MXN per year. However, only 7 billions MXN of this cost is attributable to tobacco consumption which represents 4.3% of IMSS's budget during the year of 2004. The most expensive disease to treat is "heart attacks" which cost the institute 178,266 MXN on average per year per patient. Incidentally, this disease the biggest numbers of cases (39,906 cases per year on average) albeit only 25,323 are directly attributable to smoking (media value). Cerebrovascular disease cost 162,561 MXN to treat per year. This is the second most expensive and the second most common illness of all diseases considered, with 21,061 cases per year but only 10,263 cases are smoking related. Lastly, lung cancer and chronic obstructive pulmonary disease cost 148,837 MXN and 99,669 MXN respectively although the former is less common (449 smoking related cases) than the latter (10,152 cases).

The same team of researchers estimated the cost of treating heart attacks at IMSS for the Mexican State of Morelos. Using the same methodology Reynales-Shigematsu, Campuzano-

Rincón et al. (2006) found that the cost of medical care is 2.9 million USD per year in 2001[8]. As 342 cases where documented, their study showed that the annual average cost per case was 8,657 USD. However, just as before, only 56% of all cases (192) are directly attributable to tobacco consumption.

It is worth emphasising that although smokers impose a financial burden to the health system of a country, the externality argument on this regard should be taken with caution. In particular, it is important to recognize the "net cost of smoking" which, as defined by Warner (2000), is 'the costs of treating smoking related illness minus the additional expenditures on non-smokers because they live longer' (p. 81). The definition has two elements. On the one hand, it is related to the ability to account for the actual financial cost associated with smoking which, in view of the evidence provided above, is significant. On the other hand, it is important to recognize that smokers, on average, die younger that non-smokers. This raises the issue that during a life-time, smokers receive less social security payments than non-smokers which suggests that smokers subsidise non-smokers (Warner, Chaloupka et al. 1995). Hence, if the additional expenditures incurred by the health system on non-smokers are taken into account, the net impact of smoking may be negligible. All this implies that smoking may also bring some benefits, not only costs. Bluntly speaking, 'smoking ensures that there are fewer people around to utilise health care services' (Warner 2000, p.81). Thus, the ability to account for the totality of all the externalities caused by it should be taken into account for determining the best correcting policy.

Intervening in the Tobacco Market through Taxation

Failures in the market of tobacco justify the intervention of non-market institutions. In fact, it is now widely accepted that governments must intervene in this market. Many have done so in recent years by undertaking various tobacco-control policies to reduce, in particular, the incidence of smoking on the population. Although what exactly should be the role of governments has been subject to debate, the consensus is that a government intervention is necessary. A comprehensive discussion on this matter is provided by Jha, Chaloupka et al. (2003). In this review, however, the discussion will focus its attention to the use of excise taxation which has become one of the most efficient tools employed by governments to intervene in the market of tobacco (Chaloupka, Hu et al. 2000).

The tax imposed to tobacco products comes in the form of an excise. An excise is a levy or a specific tax which, unlike VAT or other consumption taxes, has the characteristics of being selective in coverage (only specific goods or services are targeted e.g goods from the tobacco

[8] 1 USD = 9.18680 Mexican pesos MXN.

market) and discriminatory in intent (only those who consume those goods or services are affected by the tax e.g. smokers) (Cnossen 2005). Excises may be levied at specific rates (a fixed amount per quantity or by weight) or levied at *ad valorem* rates (a fixed amount of the price) or some combination of both. Usually, revenue collected from excises is reported to the tax authorities by the producer or the importer of the good that it is intended to be traded in a determined country. The implementation of excises pursues three main objectives: To raise tax revenue, to discourage the use of certain goods and to correct for externalities.

The first objective pursued by an excise is to increase fiscal revenue. One of the main mechanisms employed by governments to pay for the goods and services they provide and/or finance their debt is through taxation. Therefore an excise is implemented to this end. Although the importance of the fiscal revenue that derives from the taxing tobacco is relatively less important than other taxes such as the income tax, for some countries it is still an important and profitable source of revenue (Yurekli 2004). One explanation for this is that excisable goods are usually much easier to administer than any other tax given that revenue collection is linked to physical control (Cnossen 2005). Thus, governments can monitor their production (or trade) and distribution. In addition, excisable goods are normally provided by few producers which make the collection of tax revenue simpler. In the case of the tobacco market, it is largely dominated by two global companies, Philip Morris and British American Tobacco.

Moreover, excisable goods have few or no substitutes. As a consequence, their demand is inelastic. Goods that lack close substitutes have desirable economic properties which make them suitable for a tax. In fact, fiscal revenue is expected to rise if taxes are increased (other things being equal). This gives rise to the presence of differential tax rates; usually higher rates on tobacco products in comparison to any other goods in an economy.

The economic rationale for the presence of differential taxes lies on the well-known Ramsey rule due to Ramsey (1927) which prescribes the basic principle of optimal taxation. This states that tax rates on goods should be determined inversely to their elasticity of demand[9]. Empirical evidence over the years has shown that tobacco products, particularly cigarettes, have an inelastic demand (see below) thereby it comes as no surprise that these are an ideal candidate for an excise.

The second objective being pursued by taxing tobacco is to discourage its consumption. The causal effect of smoking and the developing of death-threating illnesses is well established and confirmed by decades of scientific medical research therefore discourage the consumption of

[9] Goods with inelastic demand also provide an efficiency argument for the implementation of excises. This relates to the notion that economic distortions are caused by an excise should be small under the Ramsey rule under the assumption of horizontal supply curve.

tobacco seems a consistent prescription. Thus, a tobacco tax is imposed for “health reasons” or to “improve public health” (Yurekli 2004). Although the ideal scenario would be for smokers to quit, the initial aim is for consumption to be reduced.

There are health benefits from reducing the consumption of tobacco although they appear to be still rather small. In a review of studies where health effects of reduced smoking (without quitting) has been analysed, Pisinger and Godtfredsen (2007) found some but small short-run benefits from reduced smoking. In particular they found from 25 studies reviewed, improvements in respiratory symptoms and cardiovascular risks as well as 25% decrease in biomarkers and incidence of lung cancer for a 50% reduction in tobacco consumption. This may be translated into possible long-term benefits such as a reduction in mortality and morbidity.

However, as Chaloupka, Hu et al. (2000) pointed out, discouraging consumption does not only mean reduction in consumption. It also means reducing smoking initiation rates particularly amongst youth and young adults. According to Cnossen (2005) the price elasticity of demand for cigarettes amongst the young is, on average, twice the price-elasticity amongst adults therefore the tobacco tax also works as a deterrent from smoking. The youth are a particular targeted population group when it comes to tobacco-control policies and so calls for even higher tobacco taxes aim to focus on this group (IFS 2009). Of course, tobacco taxes are also meant to discourage former smokers from relapsing and to deter non-smokers from initiation (Yurekli 2004).

The third reason to justify the tobacco tax is to correct for externalities. Earlier in this review it was documented that the existence of health care costs that arise from smoking and these are imposed, at large, to non-smokers. Although the estimations provided are far from accounting for the true cost of smoking, it poses the issue of who should be the bearer of these. In correcting for this negative externality, the institution of a tobacco tax appears to be an ideal government’s solution (Jha and Chaloupka 2000). Taxation, however, is only a “second best” solution. Jha, Chaloupka et al. (2003) discusses some “first best” policies which include educating children about the dangers of smoking or restrict advertisements of cigarettes. From the available evidence though, they admit that such solutions are relatively less effective.

However, is the fact that the price of tobacco products do not reflect the true costs of their consumption, then, the Pigouvian prescription becomes clearer. As explained by Cnossen (2005):

> Efficient consumption of tobacco can be achieved through the tax system by imposing a tax on cigarettes equal to the cost of the damage cause to other people. That is, the Pigouvian tax should be equivalent to the marginal external cost per cigarette consumed. Cnossen (2005)

Although the Pigouvian prescription is appealing, the difficulty lies in the identification of the marginal cost, which is hard to measure. Altogether the institution of the tobacco tax is mainly justified on fiscal grounds, to work as a deterrent from smoking and to correct for the negative externality in the tobacco market. However, unintended consequences may arise. Some of those consequences are discussed in the next sub-section.

The Regressitivity of the Tobacco Tax

The implementation of an excise is usually justified to reduce activities considered "socially undesirable", which in this case, is the act of smoking. An excise, then, helps to make producers and consumers accountable for the externality caused by producing or consuming tobacco (Frey 2005). The same applies for other socially undesirable activities include drinking of alcohol, gambling and environmental pollution. Notwithstanding, its implementation has led to a variety of concerns that affect their effectiveness in real scenarios. One of those concerns is the fairness of this tax. This is usually ignored in favour of one of the supposed benefits which are to reduce the consumption of tobacco. Whilst the benefits of imposing a tobacco tax may outweigh the cost, there is no reason why unintended consequences of doing should be overlooked. In this review, issues that arise from imposing a tax on tobacco are discussed with an emphasis on the distributional impact of this excise in any tax system although the main concern is on the Mexican context. The immediate limitation of this review is that issues concerning the design and administration of the tobacco tax are not discussed but the interested readers may consult Yurekli (2004). The main objective of this review is to provide a normative evaluation of its existence in the tax system.

There has been always a concern that tobacco taxes may impose a greater burden to low-income smokers. They spend a bigger share of their income or consumption expenditure on tobacco products than high-income smokers. It is not surprising then that one of the arguments against high or higher tobacco taxes is the distributive impact of the tax burden created which, in general, falls disproportionally towards the poor. Therefore, the main concern is whether the principle of vertical equity is fulfilled. This is one of the main normative principles of tax policy which states that an individual with the greatest ability to pay should pay more on taxes as proportions of their income (Ibid). However, it would appear that, by nature, tobacco taxes are regressive.

According to Remler (2004) there are two main reasons for the tax on tobacco to be regressive. The first relates to the notion that richer individuals invest a larger share of their income than the poor, and so consumption represents a smaller share than their counterparts. The second reason relates to the observation that the rate of smoking prevalence among the poor

is higher than amongst the rich thereby there is a disproportionate share on consumption by both groups. However, in the influential paper by Warner (2000), he states that even if the tax on tobacco is fundamentally unfair, a tax increase may not be regressive. His argument relies on the poor to be more price responsive than the rich. Therefore an increase in the tax would make any tax increase progressive. This is the idea of marginal progressivity discussed in some detail by Peck (2004).

Nevertheless, is it important to look into the progressivity of a single tax? Warner (2000) points out that what it matters is to consider the overall progressivity of the entire tax system. In deed, the tobacco tax represents only one element what would include the distributional effect of public spending. Therefore, the fact that one tax is in particular regressive, their effect can be offset by lowering other tax or by any other distributive measure (Peck 2004). However, according to Remler (2004) 'standard principles for assessing the equity of taxes should not be forgotten'.

The Analysis of the Demand For tobacco

The discipline of economics has contributed to understanding factors that influence the demand for cigarettes. This has made possible the formulation of better policies which aim to reduce consumption. One of the most important findings is that the demand for cigarettes responds to changes in price. This suggests that, like many other goods and services traded in the market, consumption of cigarettes follows the law of demand (Chaloupka 1999). This has supported the use of taxation as discussed in the previous sub-section. Understanding the relationship between the price-sensitivity of consumers and their demand for cigarettes has not been, however, the only focus of attention. The increasing availability of national representative surveys in which the unit of analysis is the individual, has provided further insights of the ways in which the demand is determined. Given the increasing public attention of social issues and health problems related to smoking, a more detailed analysis of demand for cigarettes appears essential.

In this sub-section, a review of studies of demand analysis is provided. Firstly, a summary of the main findings from "traditional" tobacco demand literature is given. Traditional means, in this sense, studies where the focus of research is on investigating the role of "macro-social influences" such as price and tobacco-control policies. Secondly, a review of studies in which micro data at individual-level has been used is provided.

Traditional Tobacco Demand Analysis

In understanding how the demand for tobacco responds to changes in price, empirical studies have been mostly concerned in estimating the "price elasticity of demand". Doing so different estimation techniques and variety of data sources have been used to measure how quantity demanded responses to changes in prices. The most comprehensive review on empirical estimates of price elasticity can be found in (Chaloupka 1999) and (Chaloupka and Warner 2000). Their review concludes that the price elasticity of tobacco demand falls in the *narrow* range centred on -0.4. This estimate is interpreted as follows: for a 10% increase in prices, the consumption of tobacco will be reduced by approximately 4%. It should be noted that this estimate places tobacco in the category of goods that are "price-inelastic". This means that any change in price will cause a proportionately smaller change in the quantity demanded. This additional interpretation is sometimes overlooked. However, when discussing the benefits of tobacco taxation, the estimate of price elasticity also suggests that tax revenue will increase given increases in the price.

It is worth mentioning that even though most empirical estimates agree that tobacco products have a low level of price sensitivity, different estimates tend not to be in a narrow range as suggested by Chaloupka (1999) and Chaloupka and Warner (2000). Gallet and List (2003) conducted a meta-analysis of 86 studies that produced estimates of price elasticity (over the last fifteen years of their article's publication). They found that the range of estimates reported in the literature is wide; from -3.12 to 1.41. According to their study, this is explained by several "heterogeneous influences" including the econometric model employed; it also includes the model's functional form, estimation technique and data structure. All these elements seem to confound the magnitude of the estimates. Nonetheless, they did find that tobacco products are, in general, price inelastic. Moreover, they found that the mean estimate of the price elasticity found across studies is indeed -0.4.

Meta analysis has been used once before to summarise econometric results of price elasticity of tobacco demand (and other determinants). Andrews and Franke (1991), for example, found an average estimate of -0.363 across 48 studies reviewed produced in the period from 1933 to 1990. Interestingly, the obtained mean estimate is similar in magnitude with the estimate found more than a decade later by Gallet and List (2003). However, their study restricts the analysis only to "time-series research". In other words, their analysis only includes studies where aggregate time-series data has been used. This is, perhaps, not surprising given that other sources of data, such as cross-sectional data from households or individual surveys, were not as widely available as they are now.

One important characteristic of all studies previously cited is worth mentioning. All these studies targeted estimates of price elasticity produced from developed nations. In recent years, a small body of research is targeting developing countries given the availability of micro data. An early review by Chaloupka, Hu et al. (2000) found that for low and middle-income countries, the estimate of price elasticity is usually double to the one found for high-income countries at around -0.5 to -1.00.

Traditional analysis not only has been concerned with the role of price when investigating factors influencing the demand for tobacco. Some other factors have been also considered for analysis. For example, empirical research has found that as income increases, the demand for tobacco increases as well. This finding is summarised by the calculation of the income elasticity which, given its positive value, tobacco products are considered as a "normal good"[10]. However, the magnitude of this elasticity, at around 0.42 according to Gallet and List (2003), also suggests that cigarettes are "necessities". This means that, as income rises, the demand for cigarettes will increase at a smaller proportion than income[11].

The effect of smoking advertisements also seems to appear routinely in studies of tobacco demand. The general hypothesis has been that advertisements would have a positive influence on cigarette consumption. This hypothesis has been verified by the positive value of the calculated elasticity too, at around 0.10 (Gallet and List 2003). Although estimates of the effect of advertisements appears to differ amongst studies, its effect has tended to decrease over time (Andrews and Franke 1991)[12]. In fact, due to the implementation of tobacco-control policies around the world, the effect of cigarette advertisements has been diminished. On the other hand, the role of counter-advertisements has been more relevant in recent studies.

Tobacco Demand Analysis in Mexico

Three published studies of the determinants of tobacco demand have become available in recent years in Mexico. Two studies have relied on cross-sectional data from household surveys whilst only one, the most recent one, relies on aggregate time-series consumption data.

[10] In contrast, goods are said to be "inferior goods" when the demand falls as income increases. Thus, the income elasticity is of negative value.
[11] This is the case when the income elasticity is less than 1. When it is greater than 1, goods are said to be luxuries. This means that as income rises, the demand of cigarettes will increase at a greater proportion than income.
[12] Given that referred studies are a "meta analysis" of measures of elasticity, when it comes to "advertisement" it is unclear what is the actual metric used for such calculation(s). However, in the study by Chaloupka (1999) it is revealed that the econometric evidence that has measured the responsiveness of advertising on cigarette demand is based on "aggregate expenditure on total cigarette sales".

The first study conducted by Jiménez-Ruiz, de Miera et al. (2008) used a sample selection model (Heckman model with a two-stage estimator) to estimate the price and income elasticity of demand using seven rounds of the ENIGH (1994-2005). Their estimated price and income elasticity conditional on smoking participation is -0.06 and 0.25 respectively. The conditional price and income elasticity of cigarette consumption is -0.45 and 0.25 respectively, as well. The cited estimates produced, according to the authors' account, an overall price elasticity of -0.52 and an overall income elasticity of 0.49. These two estimates, however, may be biased as the method used for calculating the overall elasticity in both cases is incorrect[13]. In addition, their study used the estimated elasticities to simulate the effect of increases in the level of tax on tobacco on consumption and tax revenue. They found, in particular, that an increase of 10% in taxes would reduce the consumption of cigarettes by 6.4% and increase the government's tax revenue by 15.7%.

Expanding on the study cited above, Waters, Saenz-de-Miera et al. (2010) provides an estimate of the price elasticity of demand. Their estimated values are -0.70 for the year of 2006, and -0.55 for the year of 2008. The study also used data from the ENIGH. However, there is no explicit discussion about the methodology used but one can infer that a similar approach was taken as in Jiménez-Ruiz, de Miera et al. (2008). Similarly, they provided simulated results on tax revenue caused by increases on the level of tobacco tax. The simulation also includes an estimation of the number of preventable deaths that would generate the tax rise. Waters, Saenz-de-Miera et al. (2010) found that by increasing the share of tax on the final price of cigarettes to 75%, 0.98 million of premature deaths can be prevented and it would generate 35 billion MXN of tax revenue by 2013.

Lastly, Olivera-Chávez, Cermeño-Bazán et al. (2010) estimated the price and income elasticity of demand using quarterly aggregate consumption data from 1994 to 2005. Aggregate tobacco consumption is restricted to filtered cigarettes which are converted into consumption per capita taking into account the population over 15 years of age only. The estimated value of long-run price elasticity is -0.15 and the estimated value of income elasticity is 0.54. The estimation methods employed were Ordinary Least Squares (OLS) and Dynamic OLS regressions using a double-log model.

[13] In the presence of censored data, the unconditional and conditional elasticity can be summed-up only when a two-part model is employed [see Wilkins et al. (2004) for a discussion].

Demand Analysis with Individual-Level Data

Using individual-level data for analysing the demand for tobacco is not entirely new in the literature. However, it has become more common in recent years given that it allows investigating in more detail smoking behaviour by focusing on the individual rather than the aggregate. For instance, these studies have showed the role of socio-demographic characteristics such gender, level of education and ethnicity in explaining smoking patterns seen amongst individuals. Therefore, the econometric analysis reviewed here provides some useful insights in to what extent individual's characteristics explained consumption.

Using individual-level data, however, imposed some methodological issues than need to be addressed in empirical analysis. The most important issue, perhaps, has to do with the presence of a mass of zeros seen in the data. This reflects the fact that, for reasons to be determined, individuals do not consume tobacco products. As the large number of zeros cannot be ignored econometrically, a large body of research has focused its attention in addressing this issue. Although this is discussed in some extent later in this thesis, it would appear that the literature has settled for the "Double – Hurdle Model" originally due to Cragg (1971). This model has brought in two important features when modelling the consumption of tobacco. It allows investigating factors influencing the decision to smoke, known as "propensity", and investigating factors influencing the level of consumption, known as "intensity". These two realisations have invariably increased our understanding of how the demand is determined; factors affecting it and their direction of influence. For this reason, using individual-level data has become more relevant in empirical work.

In the traditional tobacco demand literature using individual-level data has several advantages too. For example, Chaloupka and Warner (2000) explains that relatively to "aggregate data studies" potential simultaneity biases are less likely to be encountered given that socio-demographic determinants are less correlated with price. It also allows investigating the effect of price, say, for various population sub-groups. Nonetheless, potential biases are also present. For instance, the use of individual-level data relies on the assumption that any under-reporting of self-reported consumption of tobacco is proportional to true consumption across population groups (Wilkins, Yurekli et al. 2004). In term of econometric modelling, the presence of ecological bias is latent given that 'omitted variables affecting tobacco use may be correlated with the included determinants of demand' [Wasserman et al. (1991) cited Chaloupka and Warner 2000].

Recent studies have been continually concerned with understanding the effect of price. On the other hand, a substantial number of studies have focused in addressing some methodological

issues within the discipline of econometrics. Studies which have been fully concerned with non-price factors include Raptou, Mattas et al. (2005), Yen (2005a), Yen (2005b) and Bauer, Göhlmann et al. (2007). The most comprehensive study of all these is given by Raptou, Mattas et al. (2005) which investigates the effect of smoking restrictions in workplaces and schools and psychosocial factors controlling for various demographic and socio-economic variables. Their results showed some differences regarding those factors affecting the risk of being a smoker and those affecting the level of consumption. In particular, they found that the most important determinants of smoking behaviour are the psychosocial factors. For example, the smoking behaviour of relatives and friends are positively associated with consumption whilst the perceived health effect of smoking is negatively associated with participation.

The study by Yen (2005a), Yen (2005b), and Bauer, Göhlmann et al. (2007) focused on investigating gender difference in factors determining the consumption of cigarettes. The studies by Yen (2005a) and Yen (2005b) formally tested the hypothesis of equal consumption parameter between men and women. Although both studies employed a different econometric model (in the former study the Double-Hurdle model was used, whilst in the latter a Sample-Selection model is used), Steven T. Yen rejects the hypothesis of equal consumption between genders. He concludes that for a more appropriate modelling of cigarette demand data samples should be segmented by gender.

Concerning traditional studies of demand analysis, Sheu, Hu et al. (2004) investigates the price sensitivity of smokers (price elasticity) in California following a major tax rise in 1999. Bishop, Liu et al. (2007) and Mao, Sung et al. (2008), two recent studies from China, also aim to provide calculations of price (and income) elasticity. The former based their calculation on a sample of urban males with data for 1995 whilst the latter provide estimates for the general population and by income groups with data from 2002.

Concerning methodological issues, Yen (1999) investigates the results from two competing views representing the structure of the Double-Hurdle model. Under this model, the demand for cigarettes is estimated in two-stages. Commonly, the first stage would involve a Probit model for smoking participation and in the second stage, an Ordinary Least Square regression with a logarithm transformation of the dependent variable with the sub-sample of smokers reporting consumption. This is the “Gaussian distribution” option. Alternatively, a Logit model for smoking participations in the first stage may be employed followed by a zero-inflated negative binomial model (ZINB). This is the “Poisson distribution” option. Thus given that the interest is in modelling the consumption of cigarettes, a question arises of whether consumption should be taken as continuous distribution (first option) or as a discrete distribution (second option). The study by Yen (1999) provides comparisons of estimates by using different both distribution densities. He

did not find systematic differences between these two though the Gaussian option is the most commonly framework used in empirical work.

Other methodological issues have been discussed in recent studies of tobacco demand. Tauras (2005) investigates the estimation bias in elasticity calculations produced by applying logarithm transformation of the dependent variable in the Cragg's model. This bias is generated when non-normal and/or heteroscedastic errors are encountered in the second stage of the estimation. Thus, Tauras (2005) suggests a Generalised Linear Model (GLM) as an alternative to OLS. By the same token, other studies focused their attention on discussing the possibility of employing alternative estimators. Su and Yen (2000) suggest a system of censored equations closely related to the Heckman model. Similarly, Madden (2008) discusses the viability of the heckit estimator given that it is considered a viable alternative to the Cragg's model.

Summary of main findings

The studies reviewed focus their attention on the impact of addiction on the level of tobacco consumption. In addition, this review targets those studies who have investigated non-price determinants. However, only a handful of studies meet this last criterion. Nonetheless, irrespective of whether the focus of the study was on price or other issues, their results provided are a valuable source of information in enhancing our understanding on factors determining the demand for tobacco.

The findings of all studies reviewed have been summarised in Table 1 and Table(s) 2. All studies have been divided by whether the data has been pooled or whether it is still divided by gender. Twelve main categories have been considered which represent individuals' demographic, socio-economic and psychosocial factors. Given the reliance on the Double-Hurdle framework, the effect of each factor on the propensity (1^{st}) and intensity (2^{nd}) is reported albeit only when a statistically significant effect has been encountered. In some cases, the "overall" effect on the mean has also been reported though, in general, the aim of the majority of studies has been in analysing the direction of influence of each factor in each realisation. It is worth noting the not all studies have considered investigating the same factors in order to explain the propensity or the intensity, so the results presented should be taken with caution.

Admittedly, it has been difficult to draw general conclusions as to which factors are more relevant in determining the demand given the mixture and variability of results across studies. Notwithstanding, some regularities seem apparent. For instance, males are in general more at risk of becoming a smoker than women and upon being a smoker, they would smoke in greater intensity than their counterpart. Education also appears to have a negative effect on the

likelihood of an individual in becoming a smoker. Individuals with more years of education appear to smoke less upon smoking. Almost consistently, individuals who are divorced or separated are both more likely to become a smoker and consume tobacco in greater intensity. Furthermore, some factors reflecting social status such as being a homeowner makes and individual less likely to become a smoker and to smoke in lesser intensity. In contrast, factors reflecting employment have an effect, mostly, on the propensity rather than in intensity. Lastly, other factors with no defined direction of influence include the area of residence (urban or rural), ethnicity and regional differences within a specific country. All these factors appear routinely in studies of this kind.

Review of studies from Mexico

Studies of tobacco demand using individual data are not yet available for Mexico. However, from the analysis of the available surveys which have monitored the tobacco epidemic, it has been possible to derive some demographic and socio-economic characteristics that explain smoking. In effect, the only available literature has focused on explaining the propensity. One example of this is the study conducted recently by Buttenheim, Wong et al. (2010). Their study aims to explain health outcomes such as smoking status and obesity base on "social gradients". These social gradients are determined by education and household assets (wealth). Focusing only on smoking status, the study tests the hypothesis that higher levels of socio-economic status are associated with healthier behaviour. Buttenheim, Wong et al. (2010) found that higher educational attainment and greater wealth is associated with a higher prevalence of smoking amongst women. This is not the case for the case of men in which higher level of smoking prevalence are association with higher educational attainment but low levels of assets.

ANNEX TO LITERATURE REVIEW

Table 1. Studies of Tobacco Demand Using Individual-Level Data.	
Reference	**Data source**
Jones (1989)	▪ United Kingdom ▪ General Household Survey of 1980 ▪ Sample of individuals over 18. ▪ $Y =$ Cigarette smoked. ▪ $N = 1{,}573$, 38.6% of individuals reported consumption of cigarettes. ▪ Different specifications of the Double-Hurdle Model.
Yen (1999)	▪ U.S ▪ 1989-91 Continuing Survey of Food Intakes by Individuals. ▪ Sample for women over 18. ▪ $Y =$ Cigarette smoked per day. ▪ $N = 5{,}814$, 24.7% of individuals reported consumption of cigarettes. ▪ $\bar{Y}\|Y > 0 = 17.0$ ▪ Double Hurdle (Gaussian & count-data models)
Su & Yen (2000)	▪ U.S ▪ 1989-91 Continuing Survey of Food Intakes by Individuals. ▪ $Y =$ Cigarette smoked per day[14]. ▪ $N = 7{,}823$, 27% of individuals reported consumption of cigarettes. ▪ $\bar{Y}\|Y > 0 = 19.12$ ▪ General censored system of equations.
Farrelly et al. (2001	▪ U.S. ▪ 1976-1993 National Health Interview Survey (pooled data) ▪ $Y =$ Cigarette smoked every day. ▪ $N = 367{,}106$ ▪ $\bar{Y}\|Y > 0 = 20.2$ ▪ 2PM (Probit + OLS)
Sheu et al. (2004)	▪ U.S (State of California) ▪ 1996-1999 Behavioural Risk Factor Survey. ▪ $Y =$ Cigarette smoked every day. ▪ $N = 11{,}180$, 17.9% of individuals reported consumption of cigarettes (20.3% men & 16.1% women). ▪ $\bar{Y}\|Y > 0 = 17.0;\ 15.3$ (Men & women respectively). ▪ 2PM (Logit +count data model)
Yen (2005a)	▪ U.S ▪ 1994-96 Continuing Survey of Food Intakes by Individuals. ▪ Pool regression and by gender. ▪ $Y =$ Cigarette smoked per day. ▪ $N = 9{,}613$, 23.8% of individuals reported consumption of cigarettes (1,257 men & 1,039 women). ▪ $\bar{Y}\|Y > 0 = 20.45;\ 16.94$ (Men & women respectively). ▪ Double-Hurdle Model

Continue in the next page…

[14] It also estimates a demand equation for beer and for wine.

Reference	Data source
Yen (2005b)	▪ U.S ▪ 1994-96 Continuing Survey of Food Intakes by Individuals. ▪ Pool male-female regression and by gender. ▪ $Y =$ Cigarette smoked per day[15]. ▪ $N = 8{,}479$, 25.5% of individuals reported consumption of cigarettes (1,186 men & 979 women). ▪ $\bar{Y}\|Y > 0 = 20.8;\ 17.3$ (Men & women respectively). ▪ Multivariate Sample Selections Model.
Eleni Raptou et al. (2005)	▪ Greece (Urban area of Thessaloniki) ▪ Primary data collected during May-September 2002 ▪ $Y =$ Cigarette smoked per day. ▪ $N = 603$, 59.4% of individuals reported consumption of cigarettes (33.7% of them are women). ▪ $\bar{Y}\|Y > 0 = 16.6.$ ▪ 2PM (Probit + OLS)
Tauras (2005)	▪ U.S ▪ 1993-1994 National Health Interview Survey. ▪ $Y =$ Cigarette smoked per month. ▪ $N = 40{,}943$, 24.82% of individuals reported consumption of cigarettes ▪ 2PM (Probit + GLM)
J. A. Bishop et al. (2007)	▪ China ▪ 1995 Chinese Household Income Project. ▪ Sample of urban males. ▪ $Y =$ Cigarette smoked per day. ▪ $N = 7{,}707$, 52.5% of individuals reported consumption of cigarettes. ▪ $\bar{Y}\|Y > 0 = 13.99$
Bauer (2007)	▪ Germany ▪ German Socio-Economic Panel (1998-20040 ▪ $Y =$ Cigarette smoked per day. ▪ $N = 43{,}025$,(22,264 Women & 20,761 Men) 25.4% of Women and 34.1% of Men reported consumption of cigarettes ▪ 2PM (Logit + Truncated Negbin)
Zhengzhong Mao et al. (2008)	▪ China ▪ 2002 National Smoking Prevalence Survey ▪ $Y =$ Cigarette smoked per day. ▪ $N = 16{,}056$, 29.2% of individuals reported consumption of cigarettes. $\bar{Y}\|Y > 0 = 14.2$ ▪ 2PM (Probit + OLS)
Madden (2008)	▪ Ireland ▪ 1998 Saffron Survey. Sample of women. ▪ $Y =$ Cigarette smoked per day[16]. ▪ $N = 1{,}257$, 28.6% of individuals reported consumption of cigarettes. $\bar{Y}\|Y > 0 = 15.19$ ▪ 2PM (Probit + OLS)

[15] It also estimates a demand equation for beer and for wine.
[16] It also estimates an equation for alcoholic drinks.

Table 2. The Effect of Non-Price Factors in Determining the Demand, Selected Studies.

	Source	Income				Gender				Age				Education				Marital State			
		Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall
POOL	Jones (1989); [1]	e	(+)	(+)	x					Age		(+)	x								
	UK									Age2		(+)	x								
	Su & Yen (2000)	e	x	x	-0.12	Male	x	x	(+)	e	x	x	-0.868	e	x	x	-1.032				
	US																				
	Farrelly et al. (2001)	e	(-)		x	Male	(+)	(+)	x	Age	(+)	(+)	x	High School dropout	(+)	(+)	x	Separated	(+)	(+)	x
	US									Age2	(-)	(-)	x	Some college	(-)	(-)	x	Widowed	(+)	(+)	x
														College graduate	(-)	(-)	x	Never married	(-)	(-)	x
														Postgraduate	(-)	(-)	x	Divorced	(+)	(+)	x
	Sheu et al. (2004)	10k - 19,999 USD				Male	(+)	(+)	x	25-34			x	High school graduate		(+)	x	Single	(+)		x
	US	20K - 34,999 USD								35-44		(+)	x	Some college	(-)		x	Divorced	(+)	(+)	
		35K +		(-)	x					45-54		(+)	x	College graduate	(-)		x	Widowed	(+)		
										55-64	(-)	(+)	x								
										65+	(-)	(+)	x								
	Eleni Raptou et al. (2005)	e	(+)		x	Female	(+)	(-)	x	15-18		(-)	x	High school	(-)		x				
	Greece									19-29				University	(-)		x				
										30-40											
										41-51											
										51+											
	Tauras (2005); [2]	Family Income [e]	(-)		x	Male	(+)	(+)	x	18-24	(+)	(-)	x	High school graduate	(-)		x	Married		(+)	x
	US									25-44	(+)	(+)	x	Some college	(-)	(-)	x	Sep/Div/Wid	(+)	(+)	x
										45-64	(+)	(+)	x								
	Zhengzhong Mao et al. (2008)	Low Income		(+)	x	Male	(+)	(+)	x	e	(+)	(+)	x	Moderately educated	(-)		x				
	China	Middle Income		(+)	x									Highly educated	(-)		x				
		High Income		(+)	x																
	Cheng & Kenkel (2010)									Age	(+)	(+)	x	Less than High School		(+)	x				
	US									Age2	(-)	(-)	x	High School dropout	(+)		x				
														Two year college	(-)		x				
														College graduate or above	(-)	(-)	x				

	Source	Area of residence				Ethnicity/Race				Regional				Social status				Other Smoker				Other effects			
		Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall
POOL	Jones (1989); [1]													Homeowner	(-)		x	Other Smoker	(+)	(+)	x	Smoke risk	(-)		x
	UK																								
	Su & Yen (2000)	Urban	x	x	(-)	Black	x	x	(-)	Northeast	x	x	(+)												
	US					Hispanic	x	x	(-)	Midwest	x	x	(+)												
										South	x	x	(+)												
	Farrelly et al. (2001)	MSA Central City	(+)		x	White, non-Hispanic	(+)	(+)	x													Family size	(-)	(+)	
	US	Central City resid.	(+)		x	African-American, Non Hispanic	(+)	(-)	x																
						Hispanic	(-)	(-)	x																
	Sheu et al. (2004)					Black		(-)	x					Unemployed	(+)		x								
	US					Hispanic	(-)	(-)	x																
	Eleni Raptou et al. (2005)																	Family	(+)	(+)	x	Religion	(-)		x
	Greece																	Friend	(+)	(+)	x				
	Tauras (2005); [2]	MSA Central City	(-)		x	Black	(-)	(-)	x					Not employed	(+)		x								
	US	Not in MSA Farm	(-)		x	Black Hispanic	(-)	(-)	x																
						White Hispanic	(-)	(-)	x																
						Other Race	(-)	(-)	x																
	Zhengzhong Mao et al. (2008)																					Smoking initiation age [e]	x	(-)	x
	China																								
	Cheng & Kenkel (2010)					White		(+)	x													Regulation Index	(-)	(-)	x
	US																								

	Source	Income				Gender				Age				Education				Marital State			
		Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall
FEMALE	Yen (1999), [3]	e	(-)							e	(-)		-0.154	e	(-)	(-)	-0.549				
	US																				
	Yen (2005a); [4]									e	(-)	(-)	-0.42	e	(-)	(-)	-1.06				
	US																				
	Yen (2005b)									e	(-)			e	(-)		-0.44				
	US																				
	Bauer (2007); [5]	1k-1999 €	(+)		x					Age	(-)	(+)	x	In educational training	(+)	(-)		Married	(+)		
	Germany									Age2	(+)	(-)	x					Sep/Div/Wid	(-)		
	Madden (2008); [6]									Age		(+)	x	Inter Cert.				Married	(-)		x
	Ireland									Age2		(-)	x	Leaving Cert.	(-)			Sep/Div	(-)		x
														Third Level	(-)	(-)	x	Widowed			x
MALE	Yen (2005a); [4]									e	(-)	(+)	-0.44	e	(-)	(-)	-1.24				
	US																				
	Yen (2005b)									e	(-)	(+)		e	(-)		-0.71				
	US																				
	Bauer (2007); [5]	1k-1999 €								Age	(-)	(+)	x					Sep/Div/Wid	(-)	(+)	x
	Germany	2k +	(+)	(+)	x					Age2	(+)	(-)	x								
	J. A. Bishop et al. (2007)	e	(-)	(+)	0.0027					Age	(+)	(+)	x	High School or above	(-)	(-)	x	Married	(+)	(-)	x
	China									Age2	(-)	(-)	x	Middle school	(-)	(-)	x				

	Source	Area of residence				Ethnicity/Race				Regional				Social status				Other Smoker				Other effects			
		Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall	Effect	1st	2nd	Overall
FEMALE	Yen (1999), [3]					Black	(-)	(-)	x	Northeast	(+)	(+)	x	Homeowner	(-)		x					Household size		(+)	
	US					Hispanic	(-)	(-)	x	Midwest	(+)		x	White Collar	(-)		x								
										South	(+)		x												
	Yen (2005a); [4]	Suburban	(-)	(-)	(-)	Black	(-)	(-)	(-)	Northeast	(-)	(-)	(-)	White Collar	(-)	(-)	(-)								
	US					Hispanic	(-)	(-)	(-)	Midwest		(+)	(+)												
										South															
	Yen (2005b)	City		(-)		White	(+)	(+)	(+)	Northeast	(-)	(-)	(-)	Homeowner	(-)		(-)								
	US	Suburban	(-)		(-)	Black	(+)		(+)	Midwest		(+)		White Collar	(-)	(-)	(-)								
										South		(+)													
	Bauer (2007); [5]	Urban	(-)	(+)	x	Non-German	(+)	(-)	x	East Germany	(+)	(-)	x	White Collar - FT	(+)		x								
	Germany													White Collar - PT	(+)		x								
														Other job - PT	(+)		x								
														Not in the labour force	(+)		x								
	Madden (2008); [6]													Working	(-)		x								
	Ireland																								
MALE	Yen (2005a); [4]	City	(-)	(-)	(-)	Black	(-)	(-)	(-)					White Collar	(-)	(-)	(-)								
	US	Suburban	(-)		(-)	Hispanic	(-)	(-)	(-)																
	Yen (2005b)	City	(-)	(-)	(-)	White		(+)	(+)					Homeowner	(-)	(-)	(-)								
	US	Suburban	(-)		(-)	Black								White Collar	(-)		(-)								
						Hispanic	(-)	(-)	(-)					Employed	(+)		(+)								
	Bauer (2007); [5]	Urban	(-)	(+)	x	Non-German				East Germany		(-)	x	White Collar - FT	(+)	(-)	x								
	Germany													White Collar - PT	(+)	(-)	x								
														Other job - FT	(+)	(+)	x								
														Unemployed	(-)		x								
														Not in the labour force	(+)	(-)	x								
	J. A. Bishop et al. (2007)																								
	China																								

Notes:

[1] Results from estimates of the Cragg's model specification

[2] Results from the preferred model specification

[3] Results from the Negative Binomial model

[4] Results from the Double-Hurdle Model

[5] Results from using date from SOEP

[6] Results from the preferred model specification

Chapter I

The distributional effects of the tobacco excise in Mexico

Introduction

Having signed and ratified the Framework Convention on Tobacco Control (FCTC WHO), an international agreement in response to the globalization of the tobacco epidemic, the Mexican government implemented a series of policies to that end. One of the first policies introduced was a reform on the structure of tobacco taxation. This reform first aimed to end the different tax treatment between filtered and unfiltered cigarettes. Before 2002, filtered and unfiltered cigarettes were differently taxed because the latter were typically smoked by poor people. But the tax level for unfiltered cigarettes rose to match that of filtered ones in 2006. Annual tax increases to the level of excise on all cigarettes followed thereafter. The implementation of higher taxes on tobacco (particularly cigarettes) has been welcomed by the medical community in Mexico and calls for even higher taxes have been even suggested [see Jiménez-Ruiz, de Miera et al. (2008) and Waters, Saenz-de-Miera et al. (2010)]. There is no doubt of the effectiveness of using fiscal policy as part of the government strategy to curb the consumption of tobacco products, but there has been also a concern on the distributive impact of this excise. This concern is the focus of this chapter.

The aim of this chapter is to provide an assessment of the impact of the tobacco excise duties on the welfare of Mexican households. Particular attention will be given to equity considerations since a comprehensive analysis on this regard is not yet available. Thus, this work is an analysis of the incidence of the tobacco excise with the objective of providing an answer to how the burden of this excise is distributed among households of different standards of living. From the distributional point of view, it is well known that the excise on tobacco is a highly regressive tax (Cnossen and Smart 2005), and so the increases in rates since 2002 have been perceived as unfair. This perceived unfairness of the tax is based on the fact that low income households spend, as a proportion of their income or consumption expenditure, more on tobacco than high income households (Sesma-Vázquez, Campuzano-Rincón et al. 2002) and (Miera-Juárez, Jiménez-Ruiz et al. 2007). This suggests that taxing tobacco heavily would intensify the already regressive nature.

This analysis will provide a local measure of progressivity from a welfare perspective. It involves calculating the burden of tax through a monetary measure of welfare loss, namely equivalent and compensating variation to assess equity aspects. Empirically, this work will rely on the methodology proposed by Creedy (1998) in which equivalent and compensating variation are calculated by a parametric approach derived from using the Linear Expenditure System (LES) in modelling the demand. Moreover, results will be compared with measures of inequality and their respective index of progressivity.

Given that the household is the unit of analysis, budgetary data from the National Income and Expenditure Household Survey of 2008 is used. The sample survey is for 29,468 households which is representative at national levels. This is the same survey used by all fiscal incidence reports available on tax and spending in Mexico. With this data, tax incidence for the main indirect taxes, VAT and excises have been determined corresponding to the 2008 tax schedule.

The main finding of this chapter is that, in terms of vertical equity, the excise on tobacco appears to be regressive among households of different standards of living. Therefore, tobacco taxation may hit families harder at the lower end of income distribution. The structure of this Chapter is as follows. Section 2 gives a description of the structure of excises in Mexico; particular attention is given to the structure of the tobacco tax. Section 3 provides a summary of the methods used to measure tax progressivity together with a description of the data to be used. Also, the methodology followed in calculating the tax liability will be outlined. Section 4 presents the results of how the burden of taxes is distributed across household of different income levels and results for alternative measures of tax progressivity are also provided. Section 5 concludes.

Background

In 2001, legislators in Mexico approved a law that makes it mandatory for the *Secretaría de Hacienda y Crédito Público* (Ministry of Finance, SHCP in Spanish) to provide an assessment of the distributional effect of taxes and public spending. This law, included in the statue that regulates the Federal budget, states:

> *Con el propósito de coadyuvar a conocer los efectos de la política fiscal en el ingreso de los distintos grupos de la población, la Secretaría de Hacienda y Crédito Público deberá realizar un estudio de ingreso-gasto con base en la información estadística disponible que muestre por decil de ingreso de las familias su contribución en los distintos impuestos y derechos que aporte, así como los bienes y servicios públicos que reciben con recursos federales, estatales y municipales.*
>
> *Articulo 25 de la Ley de Ingresos de la Federacion para el Ejercicio Fiscal de 2001*
>
> "In order to assist in understanding the effects of fiscal policy on income of different groups of the population, the Ministry of Finance shall conduct a study of income-expenditure based on the statistical information available to show by decile of family income its contribution in the various taxes and duties, and the public goods and services they receive from federal, state and municipal resources".

The importance of this law is that it signals the explicit government interest in the fairness of the tax system. As a consequence, the SHCP has released annually "distributive tables" which

summarise the contribution of households to each tax of the federal fiscal code[1]. This complies with the law's indication of how to assess the incidence of taxes.

Distributive tables are a typical instrument employed by government agencies to analyse the impact of taxation. It involves calculating average tax rates by income groups. Although these studies may also target "representative taxpayers" by income class, family type or age etc., for the specific case of Mexico this is not well suited given the level of income inequality and poverty. Thus, households by income groups are the best option as the unit of analysis, especially given the lack of alternative statistical information.

Table 1 shows average tax rates (ATRs) by income deciles for the specific case of the tobacco tax. These ATRs have been obtained for the official reports which reflect the share of the contribution by households according to their respective income group[2]. Such calculations derived from a micro-simulation using household data from the ENIGH. This micro-simulation is simply an analysis of the average monthly impact of the current or proposed tax schedule. Although the tax code is adjusted every year, data used for such analysis comes from the latest household survey available which is the ENIGH of the year 2008. In summary, Table 1 shows that on average during the period of 2001 and 2008 households in the upper end of the income distribution (deciles 8, 9 and 10) have contributed to 54.11% of the revenue obtained through this tax whereas households in the lower end (deciles 1, 2 and 3) have contributed only to 11.13% (SHCP 2000-2010)[3]. According to these results, the reports have suggested that the excise on tobacco is a progressive tax. However, an important issue to address in this analysis is whether these figures provide a true assessment of the fairness of this tax.

Using distributive tables to present information about tax incidence have drawn criticisms in the past. Kesselman and Cheung (2004), for example, concluded that this method can […] 'be seen to fall short of the best practice in the economic literature on tax incidence and tax distribution' (p. 719). Some of the usual criticism may stem from the fact that these types of studies share some of the weaknesses that are inherent in fiscal incidence analysis (FINC) which usually assumes that:

a) Taxes are fully burned by the taxpayer.
b) Ignores any lifetime impact.

[1] The report also shows distributive tables of the government's spending by categories.
[2] The methodology employed by the Ministry of Finance ranks households according with their income per capita and divides the sample in deciles of income. Thus, 10 income groups are obtained which range from the poorest group (1st decile) to the richest (10th decile).
[3] This information may also suggest that richer households smoked more than poorer households. However, given that the data reported is only on money spend, it could be the case that richer households smoker higher quality tobacco products which are more expensive (i.e. filtered cigarettes).

The most contentious issue, however, is that the results do not seem to reflect the real impact of taxation on the welfare of (Mexican) households. In particular, the calculations do not reflect how the burden of the tobacco tax is born by households of different income levels. This is usually how progressive or regressive taxes are defined (Miller 2005). This suggests that perhaps policy makers may have been misinformed. Table 2 for example shows the amount of tax paid by households which reported consumption expenditure on tobacco. The information corresponds only to the year of 2008. However, it is possible to see that, in general, tax paid as a proportion of income falls as income increases. Although the amount of tax paid constitute only 1.6% of total income on average, the data do suggests and regressive tobacco tax. Therefore, the objective of this analysis is to bring light into the true distributional impact of the tobacco excise with the aim to apply a more comprehensive methodology that would provide more accurate estimations, and possibly, to reconcile academic research with the results provided in official reports.

While upcoming sections will deal with describing the methodology used to assess the distributional effect of the tobacco tax, the following subsections will provide a brief description of its structure and relevance in the public finances.

Excises in Mexico

Excises in Mexico are regulated by the *Impuesto Especial sobre la Produccion y Servicios* code (Special Tax on Production and Services code, IEPS hereafter) which was introduced 1981 and is part of the tax system valid at federal level. For the 2008 tax schedule, three categories of products were subject to an excise which covered those items labelled as “motor fuel”, “alcohol” and “tobacco”. In essence, excises are an additional tax due to the fact that the products included in the IEPS code are also part of the VAT base. Overall the structure of these taxes is comparable with others in existence in the majority of countries around the world. Perhaps the only exception is the levy on fuel which is variable and depends on a national price policy. Since the emphasis of this analysis is on the excise of tobacco, a brief account of its structure and their relevance in public finances is provided here. A brief account of the excises on alcohol and motor fuel can be found in the Appendix.

The structure of the tobacco tax

For the 2008 tax schedule, the excise on tobacco is levied on three different categories of products: “cigarettes”, “cigars and others” and “cigars and others made by hand”. The last category could be regarded as luxury since it is considerable more expensive that the usual manufactured cigarette or cigar.

Cigarettes and *cigars* are subject to a 150% tax rate *ad valorem*[4] whilst *cigars and other tobacco products made by hand* are subject to a 30.4% tax rate which is considerably lower than its manufactured counterpart.

The level of excise and the categorisation of tobacco products, however, has varied over the years since its formal inception in 1981. Prior to 2006, cigarettes (in particular) had to be classified as either "filtered" or "unfiltered", and this was translated into levying a different tax rates on them. This differential tax treatment 'was based on the consideration that levying products consumed disproportionately by the poor, which is specifically the case of unfiltered cigarettes, would be regressive' [(Waters, Saenz-de-Miera et al. 2010) p. 26]. Until 2002, the tax rates on unfiltered cigarettes where significantly lower, between 20-25% compared with tax rates usually exceeding 100% on filtered cigarettes.

However, due to the FCTC WHO, the tax differential between filtered and unfiltered cigarettes has been applied since was gradually eliminated 2002-6. The harmonization of rates was accomplished in 2006 and subsequent increases followed as a continuation of tobacco control policies. As a consequence, tax revenue has increased on average by 15.4% annually during the period of 2002-2008. This can be seen in Figure 1 as well which shows the evolution of tax revenue (shaded area) and the levels of excise rates (dotted and solid lines) over time. It is clear the sharp increase in revenue since the tobacco tax reform. In 2008 for example, the federal government collected 24,761.9 MXN Millions (nominal) from taxing tobacco.

International Comparison

Recent efforts by the Mexican government to set a high price for tobacco products by increasing the tax rate, has made the excise to be in accordance with international standards. This can be seen by calculating the tax as a share of the retail selling price. For the 2007 tax schedule, the tax on cigarettes represented 50.72% of the consumer price[5]. In Table 3, data on tobacco taxes from selected countries belonging to the OECD is presented. In terms of the excise on value, Mexico imposes comparatively high taxes on tobacco. However, it should be noted that given the lack of specific taxes or unit taxes, the overall impact on the consumer's price may not be substantial. In fact, tobacco companies may have partially absorbed the impact of tax increases which may explain the sharp increase in revenue collected since the reform. Although given the

[4] This is the statutory tax rate which can also be thought as the tax as % of the retail price.
[5] To make this conversion, the following formula is used: $IEPS/[(1+VAT)\times(1+IEPS)]$. If we would like to add the effect of VAT, we simply add its effect as in: $VAT/(1+VAT)$. These formulae also reveal that the calculation of the excise is carried out before considering the VAT. The year of 2007 has been chosen as data for 2008 is not available for other countries at the time of writing this dissertation.

objective being pursued by this tax, calls for establishing specific taxes have been in the public debate in recent years (Hernández-Ávila, Rodríguez-Ajenjo et al. 2007).

Taxes on imports

Imported cigarettes are subject to a duty tax of 67% and cigars of 45% tax rate *ad valorem*. However, the impact of duties on imported tobacco products were expected to remain low. 51.7% of all imported cigarettes and 49.6% of all imported cigars come from countries from which Mexico has trade treaties. For example, the free-trade agreement with the US and Canada (North American Free Trade Agreement, NAFTA) stipulated that tobacco duties would be set to zero by 2003. In turn, the commercial agreements with Bolivia and Nicaragua would see tobacco products free of any duty by 2004 and 2007 respectively. Thus, the relevance of the tax revenue coming from import duty taxes is minimum (Waters, Saenz-de-Miera et al. 2010). Nevertheless, it is worth noting that even if these treaties did not exist, imports of cigarettes represent 1.62% of the national production (Ibid). This make the IEPS and VAT the most important levies on tobacco as they represents a substantial proportion of the consumer's price (see above). Lastly, it is important to point out that Mexico does not tax exports of tobacco products.

How progressive is the tobacco tax in Mexico?

Data

Data for this analysis comes from the *Encuesta Nacional de Ingresos y Gastos de los Hogares* of 2008 (National Income and Expenditure Household Survey, ENIGH in Spanish) taken by the National Institute of Statistics and Geography in Mexico. This is a cross-sectional household survey taken biannually since 1992 which is public and freely available on the internet[6]. It collects detailed information for a variety of topics, mainly on income and consumption expenditure patterns and socio-economic characteristics for almost 30,000 households. The actual sample used for this study is for 29,468 households and is representative at national levels. This survey is the main data source available for studying, among other things, tax progressivity.

An important piece of information for this study is the computation of the tax liability, the theoretical amount of tax paid by each household. This will be related across the distribution of living standards among households. The computation of tax liability is described in the next sub-section. Some considerations regarding the preferred welfare indicator are summarised here.

[6] For more information consult:
http://www.inegi.org.mx/est/contenidos/Proyectos/Encuestas/Hogares/regulares/Enigh/default.aspx

In terms of measurements of standards of living, different statistical information is available for each household within the ENIGH database such as household's total income and consumption expenditure. Regarding consumption expenditure, this indicator is the result of aggregating the money spent across all of the 732 different items recorded in the survey by each household (monetary expenditure) plus the imputed value of all items reported as auto-consumption, transfers in kind (payments or gifts) and household rent (these last three are known as non-monetary expenditure); minus the imputed value of all items received in kind as payments or gifts.

Given that households' budget shares are an important piece of information to be analysed, the welfare indicator used to calculate them is the "total net household consumption expenditure". This guarantees to meet the budget share rule[7]. Notwithstanding, as the order of households depends on the level of income per capita rather than consumption expenditure per capita, income has been also calculated. In fact, this indicator has been calculated following the methodology established by the National Council for the Evaluation of Social Development Policy (CONEVAL in Spanish), an independent public organization with the responsibility to measure poverty in the country[8]. The final database created by following CONEVAL's "open code" (CONEVAL 2008) generates several measurements income available at gross and net values, as well as at household and per capita levels[9]. Per capita levels have been obtained by dividing each measurement by household size and also by the square root of the family size. This provides a measure of *equivalised income* an enable us to test whether the results are sensitive to family composition.

In terms of budget shares, 17 commodity aggregates have been created which closely followed those suggested by CONEVAL. The list of items included in each commodity aggregate can be found in the Appendix. Descriptive statistics of all measures of standard of living created by CONEVAL are also given in the Appendix. Information has been disaggregated by net consumption expenditure quintiles and, for comparison, households have been ordered both by net income per capita and by net consumption expenditure per capita using a weighting selection process. The weights used for this selection and for calculating any statistical figures in this analysis is the "sampling weight", a factor that provides the probability of the household being

[7] Irrespective of the number of commodity aggregates considered, the following rule should be satisfied: $\sum w_i = 1$ where w_i is the commodity's share to a welfare indicator.

[8] CONEVAL makes all its documents public and freely available on the internet, using its peer-review methodology results is convenient given that calculations are easy to handle and reproduce.

[9] This analysis has found that although the use of net income as opposed to net consumption expenditure assign some households in different places of the income distribution, the overall conclusions do not change direction by the use of either welfare indicator.

selected. These weights are available in the database. All final measurements are reported as monthly values in Mexican pesos (MXN) and in real terms.

The calculation of tax liability

In this sub-section, a description of the methodology followed to calculate the tax liability is provided. The tax liability has been calculated for each item recorded in the ENIGH therefore it corresponds to the level of all indirect taxes paid by households in Mexico. The computation closely followed the methodology applied by the *Centro de Estudios de las Finanzas Publicas* (CEFP in Spanish), a research agency attached to the Chamber of Deputies (Low Chamber), in their "fiscal incidence simulator" of VAT and excises (CEFP 2009) and (CEFP 2009) and the micro-simulation of the Ministry of Finance for its fiscal incidence report (SHCP 2000-2010).

Firstly, the following considerations have been assumed for the computation of the tax liability:

a. The bearer of the tax burden is the consumer.
b. Households and firms do not change their behaviour by the presence of taxes.
c. The preferential VAT treatment in border areas in Mexico has been ignored as well as the excise (tax/subsidy) to motor fuel.

The first assumption is commonly applied in studies of this kind and, as far as indirect taxes are concerned, it is assumed that taxes are passed onto consumers because of a horizontal supply curve (Haughton and Khandker 2009). Moreover, as this exercise falls within a partial equilibrium framework, only the direct effect of taxes, the effect on the household cost of living are of concern (Coady 2005). Thus, the second assumption ignores any 'welfare impacts that results from demand (and supply) side responses […] with associated implications for efficiency and revenue' (Ibid). The third assumption has been considered for the sake of simplification. Mexico applies a preferential VAT rate of 10% on border areas (as opposed to 15% as in the rest of the country) but given that only a small number of households in the sample live within this jurisdiction, to ease the computation of tax liability, it has been assumed that they face the national VAT rate. Lastly, as the excise applied to motor fuel was variable throughout 2008, it is difficult to determine the level of excise paid therefore it has not been considered in the overall computation.

Computation of the Tax Liability

For the 2008 tax schedule, the Mexican fiscal code taxed items differently. Some may be part of the VAT base, or they may be exempted from a tax or be subject to an excise. Items subject to VAT face either a preferential tax rate of 0% or the standard tax rate of 15% depending

of the item in question. Furthermore, items that fall into the category of alcohol and tobacco are subject to an additional *ad-valorem* tax which is the excise ("additional" because they are also subject to a VAT of 15%). Thus, for the calculations of the tax liability, five taxable groups where considered: "Exempt", "0%", "15%", "Excise" and "Not Classified". Table 1 in the Appendix shows the list of commodity aggregate created together with the number of taxable goods included in each category. Some items have received the status of "Exempt" because they are considered necessities for low income households. In turn, there are items for which is not possible to include them in the calculations given their unknown tax liability (such as fuel or items not included in the tax code) and so they are labelled as "Not Classified". In practice, "Exemptions" and "Not classified" act like 0% rate therefore tax paid from those items classified as "Exempt", "VAT 0%" and "Not classified" is clearly zero.

Households who reported positive expenditure on items classified with "VAT 15%" paid the following amount in tax:

$$T_i^H = [1 - (100/(100 + \tau_i))] \times m_i^H \qquad (1)$$

where T_i^H is the amount of tax paid by a household H on item i. τ_i is the statutory tax rate (15% in this case or VAT) and m_i^H is the consumption expenditure, m, reported by household H on the respective item i. If the item in question has been classified as an "Excise", then households paid the following amount:

$$T_i^H = \left(\frac{\tau_i}{100}\right) \times \left[m_i^H \times \left(\frac{100}{100 + \tau_i^{VAT}}\right) \times \left(\frac{100}{100 + \tau_i}\right)\right] \qquad (2)$$

Modelling strategy

Tax incidence analysis, both theoretical and empirical, is an area of constant and active research. Having knowledge about who bears the tax burden is important to assess the fairness of the tax system, so that it contributes to improving the design of taxes. Although tax policy debates deals with many issues surrounding taxation, it is mostly concerned on efficiency and equity aspects[10]. In this respect, it is desirable that a tax should be fair, which is related to the ability to pay. In other words, an individual with higher income should pay a higher proportion of income as tax than an individual with lower income. This is called progressive taxation. Regressive taxation is exactly the opposite, and under proportional taxation, individuals would pay the same proportions of their income as tax. Therefore, equity considerations require that taxation should

[10] See, for instance, (Ebdon 2005) on the evaluation of tax systems.

be progressive. Furthermore, when the tax burden is fairly distributed across consumers with different abilities to pay, then a tax or tax system has vertical equity (Rosen 2001).

Also, a tax should be efficient. Given that taxes inevitably distort consumers' behaviour and their economic well-being, such distortions must be compensated by redistributive gains of taxation. When the distributive gains are not directly measureable, the proposed tax must minimise the distortionary effects; this is what is called efficiency of a tax system.

A consensus among policy makers that a tax should be fair and should not create distortions in the economy appears to be the basic policy recommendations. The extent in which governments decide to ignore or take into account these considerations in designing their tax schedules would depend largely upon their particular objectives.

This chapter narrows the scope of analysis by focusing on the excise of tobacco (or simply tobacco tax) which has been used as a tool for tobacco control. The perceived unfairness of this tax however, has called for an examination of its equity characteristics. In the previous sub-section, an approach was outlined for determining the tax liability that households (the taxpayer) are subject to given the prevailing indirect tax system (including the tobacco tax) using budgetary data. In this section, an approach for determining the effect of indirect taxes across households of different standards of living is provided. The effect of taxes on the individual's welfare is analysed using microeconomic theory.

Suppose that an individual faces an initial situation in which his or her utility is set U^0 and prices are set at p^0. Suppose further that prices change from p^0 to p^1 due to a tax, so an individual attains a new level of utility denoted by U^1. The amount of money that an individual would be prepared to pay, in the new situation, to avoid the price change is known as the equivalent variation (EV) and is represented as:

$$EV = E(p^1, U^1) - E(p^0, U^1) \qquad (3)$$

where $E(p, U)$ is the expenditure function that gives the minimum cost of achieving a fixed level of utility given a certain level of prices; the superscripts denote the pre-tax and post-tax situation. Notice that $E(p^1, U^1)$ in (3) is simply income after the price change. Denoting this as y^1, EV can be rewritten as:

$$EV = y^1 - E(p^0, U^1) \qquad (4)$$

A similar concept can be applied, namely compensating variation (CV) which provides the amount of income that an individual would require in order to stay in the same level of utility given the tax. This is represented by the following expression:

$$CV = E(p^1, U^0) - E(p^0, U^0) \quad (5)$$

Notice as well that $E(p^0, U^0)$ in (5) is simply income before the price change as well. Therefore CV can be rewritten as:

$$CV = E(p^1, U^0) - y^0 \quad (6)$$

Estimating (4) and (6) would make it possible to analyse how much utility has been lost given the tax. This approach has been widely used in tax incidence research since the seminal work by King (1983).

Prices and welfare changes

Following the work by Creedy (1998), a parametric approach derived from estimating the individual's expenditure function through the Linear Expenditure System (LES) will be employed to estimate empirically both the EV and CV. This will allow deriving a local measure of vertical progressivity. Using the LES is empirically convenient given the minimum amount of information required although other demand systems may be used [see the Appendix in King (1983)]. Creedy (1998) is the main reference for a full exposition of the welfare effect of price changes. Alternatively, the reader is advised to consult the documentation contained in TaxLab (see below). Here, only key results are presented.

Defining A and B respectively as $\sum_i p_i \gamma_i$ and $\prod_i (p_i/\beta_i)^{\beta_i}$, the indirect utility function, $V(p, y)$, is defined as:

$$V = (y - A)/B \quad (7)$$

The expenditure function, $E(p, U)$, is found by inverting (7) so that:

$$E(p, U) = A + BU \quad (8)$$

Suppose that due to a tax, prices change from p^0 to p^1 which in such a case, (8) can be rewritten as:

$$EV = y^1 - (A^0 + B^0 U^1) \quad (9)$$

The resulting expression for EV using the LES is found by substituting for U^1, using equation (7), into (9) which after some rearranging gives:

$$EV = y^1 - A^0 \left\{ 1 + \frac{B^0}{B^1} \left(\frac{y^1}{A^0} - \frac{A^1}{A^0} \right) \right\} \quad (10)$$

where A^1/A^0 is a Laspeyres type of price index which uses, γ_i, the committed consumption of the ith commodity group, as the weight. The term B^0/B^1 which simplifies to $\prod(p_i^0/p_i^1)^{\beta_i}$ is a geometric mean of relative prices. Notice as well that $p_i\gamma_i$ and β_i are parameters that vary with income (y). In a similar fashion, the expression for the compensating variation is given by:

$$CV = A^0\left\{\frac{A^1}{A^0} + \frac{B^1}{B^0}\left(\frac{y^0}{A^0} - 1\right)\right\} - y^0 \qquad (11)$$

which can be compared with the expression in (10).

Consider Δp_i as the proportionate price change in the ith commodity group, so that $p_i^1 = p_i^0(1 + \Delta p_i)$, then from (10), A^1/A^0 can be written as:

$$\frac{A^1}{A^0} = 1 + \sum_i \left(p_i^0\gamma_i / \sum_i p_i^0\gamma_i\right)\Delta p_i \qquad (12)$$

Using the same rationale, B^0/B^1 can be written as:

$$\frac{B^0}{B^1} = \frac{1}{\prod_i (1 + \Delta p_i)^{\beta_i}} \qquad (13)$$

With all these results, EV can parametrically be represented having only in its arguments Δp_i, $p_i\gamma_i$ and β_i.The proportionate price change, Δp_i, can be simply the tax rate, whilst the last two parameters are the most problematic since they are not directly observable. However, following the approach proposed by Frisch (1959), β_i can be obtained as:

$$\beta_i = e_i w_i \qquad (14)$$

where e_i is the income elasticity for commodity i and w_i is its respective budget share (or weight). In turn, $p_i\gamma_i$ is given by:

$$p_i\gamma_i = \frac{y w_i(1 + e_{ii})}{(1 - \beta_i)} \qquad (15)$$

where e_{ii} is the own-price elasticity and the remaining arguments have been stated above. A proof of equation (15) is given in the Appendix. Therefore this methodology relies on having estimates of income elasticity, own-price elasticity, budget shares and tax rates. Households in the ENIGH are divided into income quintiles, and within each quintile, k, average budget shares for each commodity aggregate, $w_{k,i}$, (see below). For $k = 5$ income groups, this would create matrices with k rows and n columns for $\beta_{k,i}$ and $p_{k,i}\gamma_{k,i}$. In order to ease the estimation of all required parameters, the fiscal simulator TaxLab (Creedy and Foster 2002) has been employed which only needs information on budget shares, income levels, tax rates and the Frisch

parameter. Information on budget shares and income has been obtained from ENIGH previously described. Budget shares correspond to 17 commodity aggregates and calculated for each quintile of net expenditure. An estimate for the Frisch parameter and tax rates had to be obtained separately. The following sub-sections describe the approach taken to obtain these.

Frisch Parameter

The elasticity of the marginal utility of income or better known as the "Frisch parameter" is an important element to be estimated for computing the own-price elasticity as presented in expression (15). For directly additive utility functions such as the LES, the own-price elasticity is given by:

$$e_{kii} = e_{k,i}\left\{\frac{1}{\xi_k} - w_{k,i}\left(1 + \frac{e_{k,i}}{\xi_k}\right)\right\} \qquad (16)$$

where ξ stands for the Frisch parameter. This parameter should vary as income varies as established by (Frisch 1959) after whom the parameter is named. In the literature however, there is no consensus as to how this parameter must be calculated empirically. TaxLab assigns the value of the Frisch parameter according to each income level using the equation in (17) below.

$$\log(-\xi) = a - \alpha \log(y + \theta) \qquad (17)$$

where y can be regarded as income (or a suitable proxy) and a, α and θ are parameters to be specified. Equation (17) above is based on the empirical work by Lluch et al. [1977 cited Cornwell and Creedy (1997)] and is a flexible functional form that allows the absolute value of ξ_k to decline as y increases, though empirically $|\xi| \geq 1$.

To estimate the required set of parameters, thirty random values of ξ found in Cornwell and Creedy (1997) were matched with thirty average values of net total expenditure ordered from low to high from the database constructed. Then, (17) was estimated using ordinary least squares under the constraint that $\theta = 0$. The regression analysis gives estimates of $\hat{a}$ =9.67 and $\hat{\alpha}$ =-0.9 which provides plausible values of the elasticity of the marginal utility of income for different levels of net expenditure.

Average Tax Rates

Another important component that needs to be specified in TaxLab is the proportion in which prices change. Price changes for a variety of reasons but it will be assumed that it is due to taxation. The approach followed is to calculate the overall effect of the entire indirect tax system

from which it is possible to separate the effect of the tobacco tax and assess its impact on household welfare. Thus, it will be consider that "price changes" are the current level of indirect taxation.

Given the structure of the tax system in Mexico and that this study is at aggregate level, it would be impractical to use current statutory tax rates of VAT and/or excise rates as a proxy for "price changes". Instead, TaxLab has been given 17 values of "average tax rates" for each of the commodity aggregates constructed. These rates then, represent the level of current indirect taxation prevailing in the nation. Having calculated the tax liability for each item and household, average tax rates by commodity aggregates are computed following the methodology suggested by Decoster (2005) as:

$$\tau_J = \frac{T_J}{m_J - T_J} \tag{18}$$

where τ_J is the average tax rate for the commodity aggregate J so that $i \in J$, T_J is the sum of indirect tax paid by all households from a subset of different commodities and m_J is the sum of consumption expenditure by all households also from a subset of different commodities. Hence, calculated ratios in (18) are thought to be "average rates" since they 'reflect the interplay of differential tax rates and excises' (Ibid).

Table 4 presents the calculated average tax rates for each commodity aggregate and its corresponding budget share valid for the entire dataset. For seven commodity aggregates, the resulting average tax rate is 15%. This is simply the rate of VAT given that all of the items included are subject only to such rate. In the remaining groups, it is possible to detect the effect of having items with different tax rates. "Food" for example, which is the most important group in the list by budget share, captured a small average tax rate given that a small amount of items are subject to a rate of VAT at 15%. "Public transport" and "Household rent" both have an average tax rate of 0%. In the first case, it is because all items included are exempt of taxes whilst in the second case the average tax rate is set at 0% by assumption given that household rent is an imputed value. "Alcohol" and "Tobacco" captured the highest levels of tax. The average tax on alcohol is driven by different excise rates within the category whilst in the case of tobacco, the excise was uniform. Note however that both alcohol and tobacco represent a small share in the household budget. This is important given that household's budget shares will dictate the impact of the calculated tax rates on the distribution of the tax burden across household of different standards of living.

Alternative Methods

One of the most common methods used to measure progressivity and vertical equity is the use of concentration curves, a type of Lorenz curve. Concentration curves are a descriptive and normative tools based on the theory of welfare dominance which application has been useful in tax incidence analysis (Duclos and Araar 2006). It involves constructing a Quasi-Lorenz curve of the distribution of a tax or taxes paid T_k. It is called "Quasi" because households (or individuals) are ordered by a welfare indicator in increasing values (poor to rich). The Quasi-Lorenz curve is defined as, (Sahn and Younger 2003):

$$C_{(i)} = \left(\sum_{k=1}^{i} T_k \Big/ \sum_{k=1}^{n} T_k \right) \tag{19}$$

where $C_{(i)}$ gives the share of taxes by the poorest i/n households. The idea then is, if the curve for tax paid is below a curve for a welfare distribution, say income, the burden of a tax is more concentrated among the rich. This suggests that the tax is progressive. A tax is regressive if the curve for tax paid is above a curve for a welfare distribution. This method is simple to use given that it requires minimal data. In fact, all the necessary data has been already obtained from the ENIGH. Thus, concentrations curves are employed to reinforce the assessment of the distributional effect of the tobacco tax.

Results

Before presenting the results, some considerations are worth mentioning. EV and CV are approximate measures of the tax burden and so they have been computed for the entire indirect tax system (all average tax rates) and for the tobacco tax only. In each case, a comparison between "all households" and "smoking households" has been provided (see below). The computations have been carried out using expression (10) and (11) above. The calculations of both measures of utility have been broken down by net expenditure quintiles (from lowest to highest). The parameters involved for computing the CV and EV (income and own-price elasticities, beta values and committed expenditures) are reported in Tables 5-8. These correspond to calculations where households have been sorted by net income per capita.

In terms of the actual estimates, the majority of parameters obtained fall within the expectations of previous empirical research. For instance, income elasticities (Table 5) are largely positive reflecting that the majority of commodity aggregates considered are normal goods. In some cases the degree of sensitivity falls (or increases) with the level of income. In addition, all

own-price elasticities (price elasticity of demand) are negative as required which are shown in Table 6. Focusing on the aggregate for tobacco (good 12 in Tables 5-8), it has been found that tobacco is a normal good. When the analysis focused on *all household* the magnitude of the income elasticity suggests that it is a luxury good except for the richest quintile. For the case of *smoking households,* tobacco appears to be a necessity except for the 4th quintile. In both cases, tobacco appears to have an inelastic demand just as the wider empirical evidence has found.

Additional parameters estimated from the LES includes the betas values (β_i) or the marginal budget shares as they are also known; all are positive and less than one as required. Depending on the commodity aggregate in question, it is possible to discern some patterns. For example, in the case of food (good 3 in Tables 5-8) the marginal budget share decreases as income increases reflecting the importance of that good in the household's budget. Similarly, the parameters for committed expenditure or "subsistence" expenditure ($p_i\gamma_i$) are positive and in general all estimates increase with the level of income. Not surprisingly, for basic commodity aggregates such as food, the estimate is large.

Progressivity

The main focus will be on the overall calculation of CV and EV. These measures are presented as a proportion of net expenditure. The ratios CV/y and EV/y are local measurements of tax progressivity which is the main interest in this analysis. If these ratios increase with the level of net expenditure, a tax or tax system would be considered progressive. If they remain constant with the level of net expenditure a tax or tax system would be considered proportional and finally if CV/y and EV/y decrease with the level of net expenditure, a tax or tax system would be considered regressive. Results are presented graphically to ease the understanding of the main findings.

Figures 2 show the impact of the entire indirect tax system on all households. It is found that, on average, the tax burden represent around 6.66% of total net expenditure if CV is considered or 6.21% if EV is the preferred measure. Therefore, all indirect taxes are found to be progressive given that both CV/y and EV/y increase monotonically across the net expenditure quintiles, though the upward tendency is clearer when tax progressivity is measured with EV/y.

The effect of the tobacco tax has been isolated by the computation of CV and EV; this is shown in Figures 4. The magnitude of the tax burden is found to be small relative to net expenditure, at around 0.29% on average across quintiles. This proportion is in accordance with the observed budget share for the entire dataset. However, it is difficult to make a judgment of the progressivity of this tax as neither CV/y nor EV/y increases/decreases monotonically across

households of different standards of living. One explanation for this is that the results are being affected by the reduced number of households in the survey reporting positive consumption expenditure on tobacco; these will be referred to as “smoking households”. Although the proportion of households with positive consumption expenditure does increase as income increases, in absolute terms the distribution of observations across quintiles are not uniform. In fact, only 7.39% of all households (2,178 households) reported consumption expenditure on tobacco. This may explain why the budget share on this commodity varies notably among net expenditure quintiles thereby a pattern of progressivity cannot be established.

In order to better assess the distributional impact of the excise on tobacco, similar computations have been carried out only for smoking households. This could be regarded as an analysis of conditional progressivity. When it comes to excises, this distinction is important because of the participatory nature of the tobacco tax. Only smokers are affected by it therefore to investigate its distributional impact among them is of most relevance. The results from smoking households are considerably different to those presented above. Figures 6 show that for smoking households, indirect taxes seem to be somewhat regressive. Calculations show that the overall tax burden represents 10.56% to 11.48% of household’s net expenditure (EV/y & CV/y respectively) though households in the lowest and lower-mid quartiles appear to be more affected by the prevailing level of indirect taxation.

However, when the effect of the tobacco tax is separated from all taxes, the magnitude of its effect is found to be large. Figures 8 show the burden of the tobacco tax. This burden represents around 4.8% on average relative to total net expenditure. The important finding though, is that as expected, the tax appears to be highly regressive across households of different standards of living. In fact, CV/y for the highest quintile is less than a half of that in the lowest quintile and a similar result is observed for the case of EV/y. Both CV/y and EV/y decrease monotonically across quintiles and on average, the proportion on the top quintiles is smaller than their counterpart in the bottom quintiles.

All the results previously discussed have been tested by calculating EV/y and CV/y using *equivalised income* as the leading welfare indicator that would help to rank households from poor to rich. The results are show in Figures 3, 5, 7 and 9. Although the magnitude of the estimated tax burden differs by a small margin, the overall conclusions are the same when using net income per capita as the ranking factor.

Table 9 provides more detailed information on smoking households. These households appear to be a sub-group with high tax liability. In fact, the average net expenditure (both total and per capita) is higher than the rest of the population. Table 9 reveals that the largest

concentration of smoking households appears to be in the richest quintile, with a majority of households living in urban areas. It has been known that in Mexico, smokers tend to be richer that non-smokers which is usually not seen for the wider international experience [see WHO (2011)]. Therefore it is not surprising that the average consumption expenditure on tobacco is considerably higher for households in the upper-end of the income distribution. But as a proportion of net expenditure, it is noticeable lower than amongst their counterparts at the lower end of the income distribution. This confirms the progressivity of the tobacco tax given that the amount of tax paid as a proportion of net expenditure is greater for poor households than in rich households.

Further evidence of the regressivity of the tobacco tax is illustrated in Figure 10. In this figure, a quasi-Lorenz curve (concentration curve) of the cumulative proportion of tobacco tax paid is plotted against the cumulative proportion of households. In addition, a quasi-Lorenz curve for the cumulative proportion of households' net expenditure has been superimposed to compare which of the two distributions is more equally distributed. The graph shows that the burden of the tobacco tax is more equally distributed than net expenditure therefore it is a regressive tax. To verify whether the tax moves towards regressivity, the computation of the Kakwani index is provided in Table 10. This index is given by computing the difference of the Gini coefficient for net expenditure (represented by the quasi-Lorenz curve in Figure 10) and the quasi-Gini coefficient from the concentration of tobacco tax such that[11]:

$$K = -\left[G_y - C_T\right] \qquad (20)$$

From which a positive value would imply a progressive tax, zero for proportional tax and negative value for a regressive tax. The calculated index estimate gave a value of $G_y = 0.306$ and $C_T = 0.110$, so $K = -0.196$ which means that the tobacco tax moves towards regressivity. Standard errors have been computed which shows that the estimate is statistically different from zero[12].

Concluding Remarks

In this chapter, attention has been paid to examining the distributional effect of the excise on tobacco in the Mexican indirect tax system. The assessment concludes that in terms of vertical equity, the tobacco tax appears to be regressive. This result is based on comparing smoking

[11] This index is the area between the two curves shown in Figure 10.
[12] For comparisons purposes the index has been calculated using a different welfare indicator for ranking households. In both cases, the same conclusion is reached.

households which belong to different living standards, so that the tax burden for the top 20% of households is less than a half of those in the bottom 20%. This conclusion is confirmed by the calculation of the Kakwani index of progressivity which estimate is negative.

The results contrast with those presented in official reports, and are more consistent with academic studies and other international experiences. The results being presented in this study, however, should be taken with caution. Firstly, the results are valid for the tax schedule effective in 2008 only, and so lifetime effects of the tax are ignored. This means that the measure of progressivity corresponds to average progressivity and so it measures only the short-run effect. Secondly, even though behavioural components have been included in the analysis, it follows a partial equilibrium approach. Thirdly, this study is specifically concerned with the excise on tobacco and its equity characteristics which constitute only one aspect of a complete tax policy analysis. This may also include the efficiency of this tax and the distributive gains of public spending. Nevertheless, evidence that the burden of a particular tax is unevenly distributed is still relevant even though it is just normative evaluation. However, given that calls for higher taxes on tobacco prevail in the public discourse, this study may contribute to understand the limits of taxation and some unintended consequences of high taxes.

ANNEX TO CHAPTER I

PAGE INTENTIONALLY LEFT BLANK

Table 1. Distribution of the Tobacco Tax Contribution by Income-Group, Mexico 2001-2008.									
	2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**All**
Decile	*Contribution*	*Contribution*	*Contribution*	*Contribution*	*Contribution*	*Contribution*	*Contribution*	*Contribution*	**Average**
1	2.17	0.39	0.69	0.04	0.04	0.73	0.73	1.70	0.81
2	8.30	2.01	4.09	3.70	3.70	5.65	5.65	2.00	4.39
3	7.90	6.84	4.87	5.10	5.10	6.46	6.46	4.70	5.93
4	7.25	5.37	7.65	6.60	6.60	5.54	5.54	7.60	6.52
5	9.18	11.89	7.75	5.30	5.30	6.28	6.28	7.00	7.37
6	12.63	13.37	10.38	6.30	6.30	9.18	9.18	10.90	9.78
7	10.74	11.19	13.94	9.80	9.80	11.42	11.42	10.00	11.04
8	9.44	17.33	13.96	14.50	14.50	14.42	14.42	14.40	14.12
9	12.82	15.77	18.74	16.70	16.70	15.50	15.50	17.80	16.19
10	19.56	15.84	17.93	31.70	31.70	24.82	24.82	24.00	23.80
Concentrations (Deciles)									
1, 2, and 3	18.37	9.24	9.65	8.84	8.84	12.84	12.84	8.40	11.13
8, 9, and 10	41.82	48.94	50.63	62.90	62.90	54.74	54.74	56.20	54.11

Source: Taken from SHCP (2000-2010)
Data Source:
2001 = ENIGH 1998
2002 = ENIGH 2000
2003 = ENIGH 2000
2004 = ENIGH 2002
2005 = ENIGH 2002
2006 = ENIGH 2004
2007 = ENIGH 2004
2008 = ENIGH 2006

Table 2. A Portrait of Smoking Households, Mexico 2008.

Decile	% of Households with Tobacco Consumption	Total Net Expenditure	Total Net Expenditure per Capita	Household Size	% of Households Living in Rural Areas	Average Expenditure on Tobacco	Exp. On Tobacco relative to Net Exp. (%)	Average Tobacco Tax paid	Tobacco Tax Paid relative to Net Exp. (%)
I	3.27%	3,789.17	754.92	5.1	66.9%	154.23	4.7%	90.09	2.4%
II	3.99%	5,168.94	1,091.45	5.5	57.9%	201.64	6.2%	117.78	2.3%
III	6.11%	5,759.93	1,266.18	4.7	37.5%	180.29	4.2%	105.31	1.8%
IV	5.87%	7,069.84	1,539.53	4.7	24.3%	211.82	3.2%	123.73	1.8%
V	6.13%	8,244.63	1,720.37	5.0	15.6%	187.11	3.0%	109.30	1.3%
VI	7.60%	8,805.20	2,264.47	4.1	18.9%	218.96	3.1%	127.90	1.5%
VII	9.08%	9,650.51	2,684.80	3.7	13.5%	259.62	3.4%	151.65	1.6%
VIII	9.45%	11,396.48	3,375.22	3.5	11.6%	241.22	3.0%	140.90	1.2%
IX	10.13%	13,638.64	4,390.32	3.1	15.9%	269.31	2.7%	157.31	1.2%
X	12.23%	23,281.10	10,431.73	2.5	5.1%	374.52	2.5%	218.76	0.9%
All	7.39	11,651.99	3,888.48	3.8	20.0%	251.53	3.3%	146.93	1.6%

Table 3. Tobacco Taxes Around the World, Selected OECD Countries 2007**.**

	Cigarettes				Cigars				Tax on rolling tobacco for cigarettes			
	Specific excise per 1 000		Excise on value	VAT	Specific excise per 1 000		Excise on value	VAT	Specific excise per 1 000 grams		Excise on value	VAT
	National currency	USD	(% of RSP)[1]	%	National currency	USD	(% of RSP)	%	National currency	USD	(% of RSP)	%
Belgium	15.93	18.04	52.41	21.00	0.00	0.00	10.00	21.00	7.96	9.01	31.50	21.00
France	7.50	8.36	58.00	19.60	0.00	0.00	27.57	19.60	0.00	0.00	58.57	19.60
Germany	82.70	88.07	24.66	16.00	14.00	14.91	1.47	16.00	34.06	36.27	18.57	19.00
Greece	5.14	7.37	53.83	19.00	0.00	0.00	26.00	19.00	0.00	0.00	59.00	19.00
Ireland	151.37	149.87	17.78	21.00	217.39	215.24	0.00	21.00	183.44	181.62	0.00	21.00
Italy	6.20	7.39	58.50	20.00	0.00	0.00	23.00	20.00	0.00	0.00	56.00	20.00
Mexico*	0.00	0.00	50.72	15.00	0.00	0.00	18.27/50.72	15.00	0.00	0.00	18.27/50.72	15.00
Norway	1 870.00	197.26	0.00	25.00	1 870.00	197.26	0.00	25.00	1 870.00	197.26	0.00	25.00
Poland	75.12	18.14	31.30	22.00	134.00	46.05	-	22.00	52.00	17.87	21.67	22.00
Spain	4.20	5.47	54.95	16.00	0.00	0.00	12.71	16.00	0.00	0.00	38.46	16.00
Sweden	200.00	21.46	39.20	25.00	560.00	60.09	0.00	25.00	630.00	67.60	0.00	25.00
U.K.	99.80	161.23	22.00	17.50	145.35	234.81	0.00	17.50	104.47	168.77	0.00	17.50

Source: Taken from OECD (2008). Source of data are from national delegates; position as at 1 January 2007.

***NOTE:**

1. Retail selling price

Mexico: A rate of 26.6% (18.27% of the RSP) for cigars or rolling tobacco applies as long as these products are handmade; otherwise a 140% rate applies (50.72% of the RSP).

Table 4. Average Tax Rates and Budget Shares, ENIGH 2008.		
Commodity Aggregate	Average Tax Rate (%)	Budget Share (%)
Food	4.1	34.41
Alcohol	38.2	0.22
Tobacco	140.4	0.23
Rent	0.0	15.88
Household and Services.	8.7	9.18
Household Furniture	15.0	0.69
Household Glass	15.0	0.38
Cleaning Items	15.0	3.67
Clothing and Shoes	15.0	4.08
Personal Care	15.0	5.05
Health	4.3	2.57
Education and Leisure	4.8	7.66
Accessories	15.0	0.11
Communication	15.0	3.36
Transport	14.2	4.76
Public Transport	0.0	5.29
Others	13.5	2.46
All	20.8	100.00

Table 5 Income Elasticities.

Expenditure class		Commodity group																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Figures 2	1	1.000	1.721	0.695	2.591	0.852	0.887	1.312	1.919	0.524	1.309	1.527	1.408	1.892	1.453	0.823	1.273	1.274
	2	1.517	1.063	0.592	2.406	1.297	0.781	1.249	1.701	0.458	1.990	1.290	1.074	2.100	1.132	0.897	1.090	1.420
	3	2.236	1.210	0.610	2.027	1.425	0.675	1.185	1.891	0.733	2.330	1.163	1.699	2.199	0.579	1.040	0.868	1.353
	4	1.921	1.743	0.633	1.446	1.152	0.697	1.291	1.701	1.017	2.128	1.352	1.761	1.842	0.394	0.997	0.942	1.137
	5	1.681	1.287	0.375	1.189	1.449	0.515	1.437	1.338	1.434	2.040	1.416	0.926	1.690	-0.666	0.989	0.879	1.150
Figures 4	1	1.000	1.721	0.695	2.591	0.852	0.887	1.312	1.919	0.524	1.309	1.527	1.408	1.892	1.453	0.823	1.273	1.274
	2	1.517	1.063	0.592	2.406	1.297	0.781	1.249	1.701	0.458	1.990	1.290	1.074	2.100	1.132	0.897	1.090	1.420
	3	2.236	1.210	0.610	2.027	1.425	0.675	1.185	1.891	0.733	2.330	1.163	1.699	2.199	0.579	1.040	0.868	1.353
	4	1.921	1.743	0.633	1.446	1.152	0.697	1.291	1.701	1.017	2.128	1.352	1.761	1.842	0.394	0.997	0.942	1.137
	5	1.681	1.287	0.375	1.189	1.449	0.515	1.437	1.338	1.434	2.040	1.416	0.926	1.690	-0.666	0.989	0.879	1.150
Figures 6	1	1.000	3.521	0.672	3.462	0.339	1.725	1.514	0.261	0.724	1.363	1.787	0.190	1.538	1.926	0.698	1.244	1.090
	2	1.895	1.560	0.637	2.357	0.564	1.124	1.713	1.171	0.639	1.998	1.235	0.115	1.470	1.407	0.929	0.972	1.346
	3	1.341	1.435	0.596	1.841	1.872	0.424	1.427	2.168	0.648	2.654	1.896	0.690	2.793	0.312	1.061	0.785	1.409
	4	0.869	1.708	0.479	1.459	1.649	0.669	1.173	1.412	0.944	2.361	1.994	1.018	2.375	-0.243	0.807	1.061	1.210
	5	1.982	0.485	0.298	1.283	1.144	0.271	1.502	0.951	1.386	2.038	1.183	0.450	1.627	-0.308	0.983	0.930	1.318
Figures 8	1	1.000	3.521	0.672	3.462	0.339	1.725	1.514	0.261	0.724	1.363	1.787	0.190	1.538	1.926	0.698	1.244	1.090
	2	1.895	1.560	0.637	2.357	0.564	1.124	1.713	1.171	0.639	1.998	1.235	0.115	1.470	1.407	0.929	0.972	1.346
	3	1.341	1.435	0.596	1.841	1.872	0.424	1.427	2.168	0.648	2.654	1.896	0.690	2.793	0.312	1.061	0.785	1.409
	4	0.869	1.708	0.479	1.459	1.649	0.669	1.173	1.412	0.944	2.361	1.994	1.018	2.375	-0.243	0.807	1.061	1.210
	5	1.982	0.485	0.298	1.283	1.144	0.271	1.502	0.951	1.386	2.038	1.183	0.450	1.627	-0.308	0.983	0.930	1.318

Table 6 Own-Price Elasticities.

Expenditure class		Commodity group																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Figures 2	1	-0.1126	-0.1951	-0.3572	-0.3185	-0.0979	-0.1463	-0.2172	-0.2209	-0.0806	-0.1599	-0.1963	-0.1599	-0.2459	-0.2305	-0.1242	-0.2386	-0.2844
	2	-0.2320	-0.1639	-0.2960	-0.4053	-0.2009	-0.1576	-0.2618	-0.2661	-0.0849	-0.3218	-0.2218	-0.1656	-0.3645	-0.2343	-0.1676	-0.2550	-0.3770
	3	-0.4256	-0.2313	-0.3031	-0.4292	-0.2745	-0.1585	-0.2918	-0.3663	-0.1601	-0.4673	-0.2426	-0.3246	-0.4712	-0.1412	-0.2312	-0.2324	-0.4216
	4	-0.4654	-0.4237	-0.3207	-0.3903	-0.2814	-0.1957	-0.3814	-0.4195	-0.2696	-0.5417	-0.3510	-0.4283	-0.5037	-0.1139	-0.2715	-0.2945	-0.4163
	5	-0.7124	-0.5468	-0.2351	-0.5317	-0.6161	-0.2338	-0.6637	-0.5724	-0.6293	-0.8780	-0.6181	-0.3935	-0.7563	0.3071	-0.4420	-0.4203	-0.5955
Figures 4	1	-0.1126	-0.1951	-0.3572	-0.3185	-0.0979	-0.1463	-0.2172	-0.2209	-0.0806	-0.1599	-0.1963	-0.1599	-0.2459	-0.2305	-0.1242	-0.2386	-0.2844
	2	-0.2320	-0.1639	-0.2960	-0.4053	-0.2009	-0.1576	-0.2618	-0.2661	-0.0849	-0.3218	-0.2218	-0.1656	-0.3645	-0.2343	-0.1676	-0.2550	-0.3770
	3	-0.4256	-0.2313	-0.3031	-0.4292	-0.2745	-0.1585	-0.2918	-0.3663	-0.1601	-0.4673	-0.2426	-0.3246	-0.4712	-0.1412	-0.2312	-0.2324	-0.4216
	4	-0.4654	-0.4237	-0.3207	-0.3903	-0.2814	-0.1957	-0.3814	-0.4195	-0.2696	-0.5417	-0.3510	-0.4283	-0.5037	-0.1139	-0.2715	-0.2945	-0.4163
	5	-0.7124	-0.5468	-0.2351	-0.5317	-0.6161	-0.2338	-0.6637	-0.5724	-0.6293	-0.8780	-0.6181	-0.3935	-0.7563	0.3071	-0.4420	-0.4203	-0.5955
Figures 6	1	-0.1249	-0.4451	-0.3417	-0.4550	-0.0434	-0.2817	-0.2479	-0.0342	-0.1162	-0.1837	-0.2457	-0.0338	-0.2233	-0.3110	-0.1153	-0.2572	-0.2381
	2	-0.3208	-0.2728	-0.3140	-0.4341	-0.0967	-0.2483	-0.3608	-0.2021	-0.1280	-0.3566	-0.2305	-0.0237	-0.2830	-0.3099	-0.1875	-0.2511	-0.3447
	3	-0.2947	-0.3205	-0.3065	-0.4404	-0.4131	-0.1132	-0.3862	-0.4831	-0.1597	-0.6056	-0.4367	-0.1687	-0.6491	-0.0869	-0.2678	-0.2329	-0.4396
	4	-0.2299	-0.4594	-0.2510	-0.4203	-0.4389	-0.2037	-0.3702	-0.3806	-0.2701	-0.6516	-0.5546	-0.2923	-0.6804	0.0762	-0.2387	-0.3538	-0.4345
	5	-0.8849	-0.2195	-0.1919	-0.5974	-0.5131	-0.1298	-0.7180	-0.4293	-0.6375	-0.9203	-0.5463	-0.2103	-0.7636	0.1480	-0.4612	-0.4663	-0.6768
Figures 8	1	-0.1249	-0.4451	-0.3417	-0.4550	-0.0434	-0.2817	-0.2479	-0.0342	-0.1162	-0.1837	-0.2457	-0.0338	-0.2233	-0.3110	-0.1153	-0.2572	-0.2381
	2	-0.3208	-0.2728	-0.3140	-0.4341	-0.0967	-0.2483	-0.3608	-0.2021	-0.1280	-0.3566	-0.2305	-0.0237	-0.2830	-0.3099	-0.1875	-0.2511	-0.3447
	3	-0.2947	-0.3205	-0.3065	-0.4404	-0.4131	-0.1132	-0.3862	-0.4831	-0.1597	-0.6056	-0.4367	-0.1687	-0.6491	-0.0869	-0.2678	-0.2329	-0.4396
	4	-0.2299	-0.4594	-0.2510	-0.4203	-0.4389	-0.2037	-0.3702	-0.3806	-0.2701	-0.6516	-0.5546	-0.2923	-0.6804	0.0762	-0.2387	-0.3538	-0.4345
	5	-0.8849	-0.2195	-0.1919	-0.5974	-0.5131	-0.1298	-0.7180	-0.4293	-0.6375	-0.9203	-0.5463	-0.2103	-0.7636	0.1480	-0.4612	-0.4663	-0.6768

Table 7 Beta values.

Expenditure class		Commodity group																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Figures 2	1	0.0007	0.0029	0.3029	0.0399	0.0028	0.0521	0.0824	0.0077	0.0232	0.0156	0.0305	0.0025	0.0431	0.0809	0.0353	0.1121	0.1653
	2	0.0011	0.0023	0.2261	0.0611	0.0040	0.0437	0.0884	0.0094	0.0163	0.0267	0.0314	0.0023	0.0653	0.0747	0.0357	0.1067	0.2050
	3	0.0020	0.0022	0.2119	0.0722	0.0054	0.0349	0.0864	0.0117	0.0243	0.0450	0.0280	0.0032	0.0926	0.0353	0.0421	0.0811	0.2219
	4	0.0025	0.0045	0.1981	0.0630	0.0044	0.0330	0.1010	0.0146	0.0318	0.0568	0.0362	0.0053	0.1059	0.0206	0.0404	0.0868	0.1952
	5	0.0032	0.0039	0.0909	0.0569	0.0070	0.0200	0.1410	0.0137	0.0562	0.1053	0.0470	0.0027	0.1430	-0.01965	0.0399	0.0767	0.2120
Figures 4	1	0.0007	0.0029	0.3029	0.0399	0.0028	0.0521	0.0824	0.0077	0.0232	0.0156	0.0305	0.0025	0.0431	0.0809	0.0353	0.1121	0.1653
	2	0.0011	0.0023	0.2261	0.0611	0.0040	0.0437	0.0884	0.0094	0.0163	0.0267	0.0314	0.0023	0.0653	0.0747	0.0357	0.1067	0.2050
	3	0.0020	0.0022	0.2119	0.0722	0.0054	0.0349	0.0864	0.0117	0.0243	0.0450	0.0280	0.0032	0.0926	0.0353	0.0421	0.0811	0.2219
	4	0.0025	0.0045	0.1981	0.0630	0.0044	0.0330	0.1010	0.0146	0.0318	0.0568	0.0362	0.0053	0.1059	0.0206	0.0404	0.0868	0.1952
	5	0.0032	0.0039	0.0909	0.0569	0.0070	0.0200	0.1410	0.0137	0.0562	0.1053	0.0470	0.0027	0.1430	-0.01965	0.0399	0.0767	0.2120
Figures 6	1	0.0010	0.0148	0.2819	0.0447	0.0014	0.0861	0.0740	0.0018	0.0290	0.0176	0.0307	0.0105	0.0401	0.0948	0.0314	0.1216	0.1189
	2	0.0019	0.0133	0.2315	0.0608	0.0017	0.0726	0.1012	0.0057	0.0227	0.0296	0.0280	0.00425	0.0467	0.0953	0.0366	0.1044	0.1522
	3	0.0021	0.0096	0.2025	0.0632	0.0064	0.0224	0.1077	0.0171	0.0209	0.0597	0.0377	0.0210	0.0983	0.0202	0.0466	0.0739	0.1899
	4	0.0011	0.0166	0.1428	0.0581	0.0073	0.0332	0.0883	0.0133	0.0284	0.0770	0.0612	0.0328	0.1453	-0.0114	0.0331	0.1029	0.1697
	5	0.0046	0.0038	0.0679	0.0584	0.0054	0.0102	0.1454	0.0087	0.0503	0.1229	0.0395	0.0117	0.1378	-0.00916	0.0400	0.0876	0.2153
Figures 8	1	0.0010	0.0148	0.2819	0.0447	0.0014	0.0861	0.0740	0.0018	0.0290	0.0176	0.0307	0.0105	0.0401	0.0948	0.0314	0.1216	0.1189
	2	0.0019	0.0133	0.2315	0.0608	0.0017	0.0726	0.1012	0.0057	0.0227	0.0296	0.0280	0.00425	0.0467	0.0953	0.0366	0.1044	0.1522
	3	0.0021	0.0096	0.2025	0.0632	0.0064	0.0224	0.1077	0.0171	0.0209	0.0597	0.0377	0.0210	0.0983	0.0202	0.0466	0.0739	0.1899
	4	0.0011	0.0166	0.1428	0.0581	0.0073	0.0332	0.0883	0.0133	0.0284	0.0770	0.0612	0.0328	0.1453	-0.0114	0.0331	0.1029	0.1697
	5	0.0046	0.0038	0.0679	0.0584	0.0054	0.0102	0.1454	0.0087	0.0503	0.1229	0.0395	0.0117	0.1378	-0.00916	0.0400	0.0876	0.2153

Table 8 Committed Expenditures.

Expenditure class		Commodity group																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Figures 2	1	2.53	5.59	1,637.00	44.51	12.15	215.20	218.10	12.79	169.80	41.35	67.51	6.17	73.16	189.90	158.60	307.60	453.10
	2	3.09	10.57	1,991.00	92.23	14.26	282.80	328.70	23.36	190.40	53.53	111.90	10.07	121.20	313.10	197.00	468.10	648.20
	3	3.79	10.15	2,248.00	160.30	20.28	330.50	413.50	29.09	209.10	78.77	137.40	9.42	179.50	397.30	237.90	570.80	891.90
	4	6.66	14.40	2,533.00	271.40	26.23	377.10	514.70	48.46	225.90	124.10	172.60	16.49	305.30	452.60	294.10	681.30	1,190.00
	5	9.78	24.35	3,639.00	423.40	33.11	539.80	685.40	78.90	274.70	125.50	237.40	31.47	429.20	674.80	417.90	978.00	1,689.00
Figures 4	1	2.53	5.59	1,637.00	44.51	12.15	215.20	218.10	12.79	169.80	41.35	67.51	6.17	73.16	189.90	158.60	307.60	453.10
	2	3.09	10.57	1,991.00	92.23	14.26	282.80	328.70	23.36	190.40	53.53	111.90	10.07	121.20	313.10	197.00	468.10	648.20
	3	3.79	10.15	2,248.00	160.30	20.28	330.50	413.50	29.09	209.10	78.77	137.40	9.42	179.50	397.30	237.90	570.80	891.90
	4	6.66	14.40	2,533.00	271.40	26.23	377.10	514.70	48.46	225.90	124.10	172.60	16.49	305.30	452.60	294.10	681.30	1,190.00
	5	9.78	24.35	3,639.00	423.40	33.11	539.80	685.40	78.90	274.70	125.50	237.40	31.47	429.20	674.80	417.90	978.00	1,689.00
Figures 6	1	4.00	10.79	1,754.00	33.57	17.92	178.90	181.20	30.90	166.50	48.91	61.07	246.40	96.36	170.80	187.10	376.90	430.40
	2	4.37	40.19	2,082.00	99.74	17.42	335.90	269.70	25.23	203.80	62.96	115.30	242.60	153.50	331.30	213.20	576.10	560.90
	3	9.69	39.38	2,532.00	175.50	17.20	411.10	444.90	35.59	236.70	80.85	99.78	221.80	117.30	516.40	288.80	667.60	798.10
	4	10.57	56.21	2,748.00	258.20	26.22	430.70	548.40	62.20	238.40	129.70	153.50	248.40	241.20	526.10	340.30	736.60	1,007.00
	5	5.03	117.10	3,732.00	368.00	43.52	623.50	604.20	100.20	262.10	103.60	298.40	391.40	439.30	639.00	432.10	1,041.00	1,272.00
Figures 8	1	4.00	10.79	1,754.00	33.57	17.92	178.90	181.20	30.90	166.50	48.91	61.07	246.40	96.36	170.80	187.10	376.90	430.40
	2	4.37	40.19	2,082.00	99.74	17.42	335.90	269.70	25.23	203.80	62.96	115.30	242.60	153.50	331.30	213.20	576.10	560.90
	3	9.69	39.38	2,532.00	175.50	17.20	411.10	444.90	35.59	236.70	80.85	99.78	221.80	117.30	516.40	288.80	667.60	798.10
	4	10.57	56.21	2,748.00	258.20	26.22	430.70	548.40	62.20	238.40	129.70	153.50	248.40	241.20	526.10	340.30	736.60	1,007.00
	5	5.03	117.10	3,732.00	368.00	43.52	623.50	604.20	100.20	262.10	103.60	298.40	391.40	439.30	639.00	432.10	1,041.00	1,272.00

Table 9. A Portrait of Smoking Households, Mexico 2008.						
	1	2	3	4	5	All
% of Households with Tobacco Consumption	3.64	5.99	6.87	9.26	11.2	7.39
Total Net Expenditure	4,562.00	6,415.00	8,566.00	10,542.00	18,913.00	11,652.00
Total Net Expenditure per Capita	943.50	1,403.00	2,032.00	3,037.00	7,695.00	3,888.00
Household Size	5.3	4.7	4.5	3.6	2.8	3.8
% of Households Living in Rural Areas	61.9	30.9	17.5	12.5	10.0	20.0
Average Expenditure on Tobacco	179.70	195.00	204.00	250.90	322.00	249.90
Exp. On Tobacco relative to Net Exp. (%)	5.53	3.71	3.05	3.22	2.59	3.28
Average Tobacco Tax paid	105.60	114.50	119.90	146.10	191.40	146.90
Tobacco Tax Paid relative to Net Exp. (%)	3.25	2.18	1.79	1.88	1.52	1.92

***Note*:** Households ordered by per Capita Net Income.

Table 10. Concentration Coefficients of Distribution, Tobacco Tax and Expenditure, Mexico 2008.				
Quasi-Gini Coefficients	Estimate	Std. Err.	Estimate	Std. Err.
	(1)		(2)	
Tax	0.110	0.012538	0.122	0.011529
Net Consumption Expenditure	0.306	0.003097	0.346	0.002896
Difference (Kakwani Index)	-0.196	0.015378	-0.224	0.015034
[95% Confidence Interval]	[-0.214 -0.154]		[-0.253 -0.194]	

(1) Ranking of Household by Net Income per Capita. (2) Ranking of Household by Consumption Expenditure per Capita.

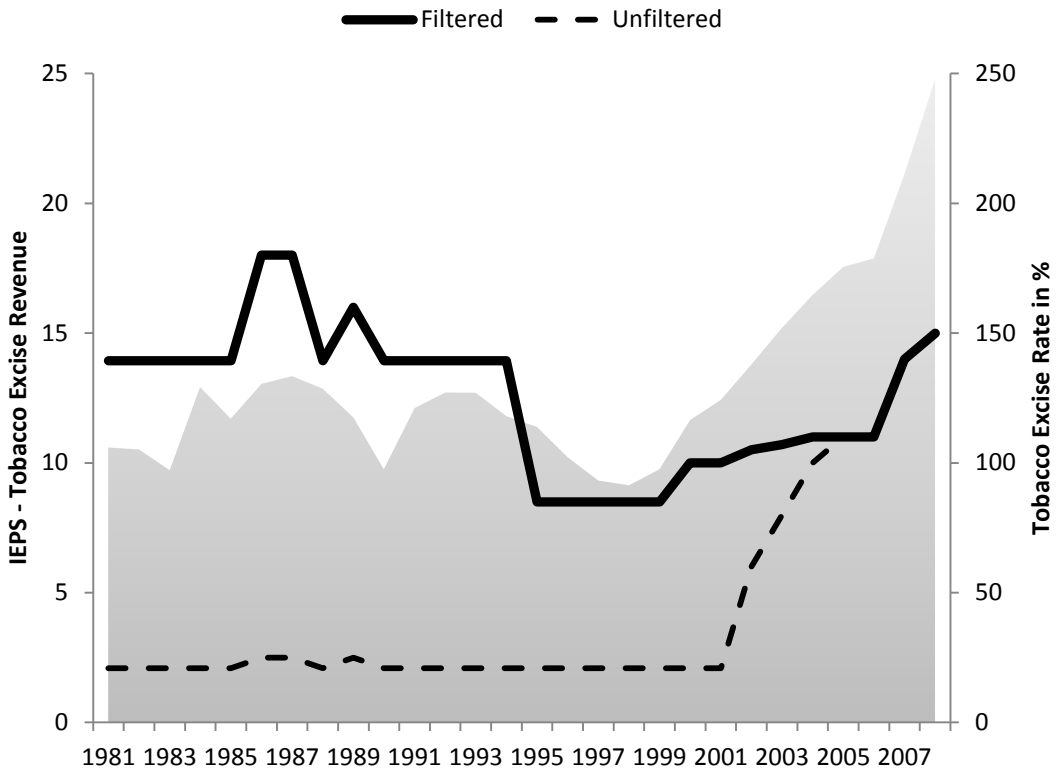
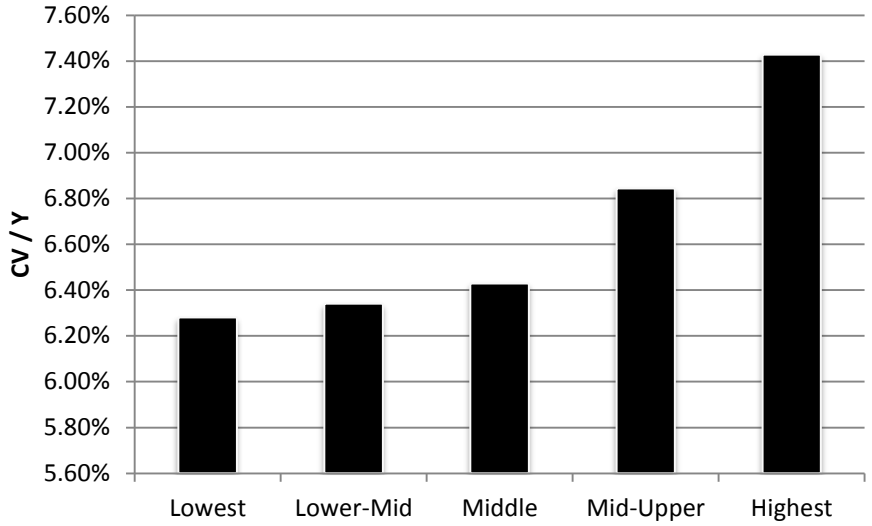
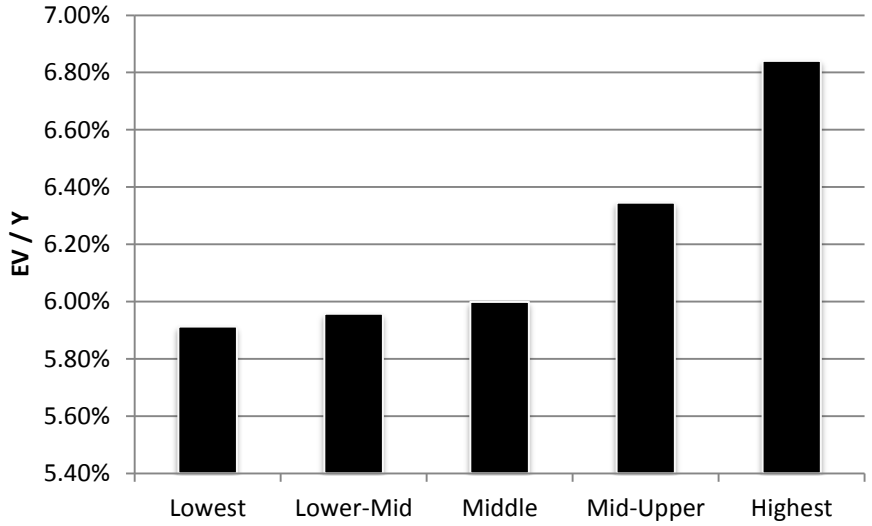


Figure 1 Tobacco Excise Rates and Fiscal Revenue, Mexico 1981-2008.
Note: Revenue in thousands of millions of MXN in real terms.
Source: Waters, Saenz-de-Miera et al. (2010).

Figures 2 Effect of Existing Indirect Taxes at Mean Net Expenditure Levels: All Households. Households Sorted by Net Income per Capita.

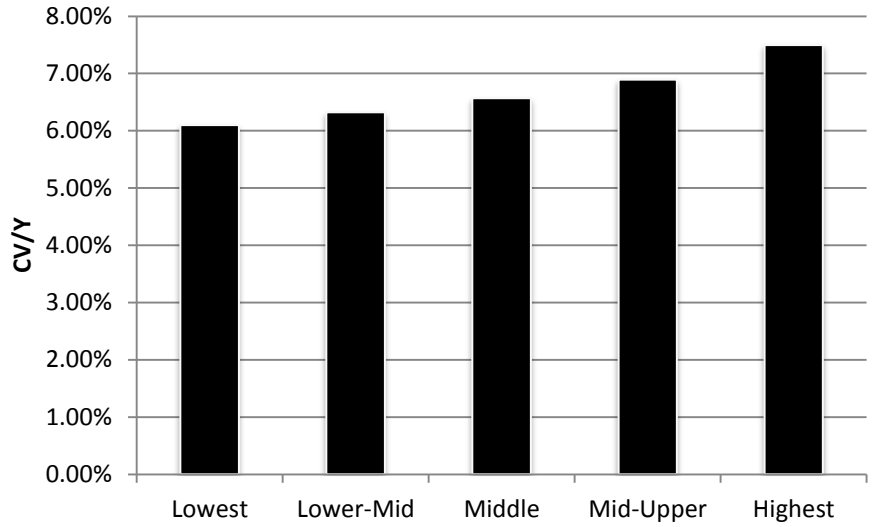
Class	y	cv	cv/y	ev	ev/y
Lowest	4,071.00	255.70	6.28%	240.70	5.91%
Lower-Mid	5,733.00	363.50	6.34%	341.50	5.96%
Middle	7,317.00	470.40	6.43%	439.00	6.00%
Mid-Upper	9,565.00	654.60	6.84%	607.00	6.35%
Highest	17,840.00	1,325.00	7.43%	1,221.00	6.84%





Figures 3 Effect of Existing Indirect Taxes at Mean Net Expenditure Levels: All Households. Households Sorted by ***Equivalised Income***.

Class	y	cv	cv/y	ev	ev/y
Lowest	2,675.00	163.00	6.10%	153.50	5.74%
Lower-Mid	4,812.00	304.00	6.32%	284.90	5.92%
Middle	6,749.00	443.30	6.57%	413.60	6.13%
Mid-Upper	9,609.00	662.10	6.89%	615.70	6.41%
Highest	20,680.00	1,551.00	7.50%	1,428.00	6.90%



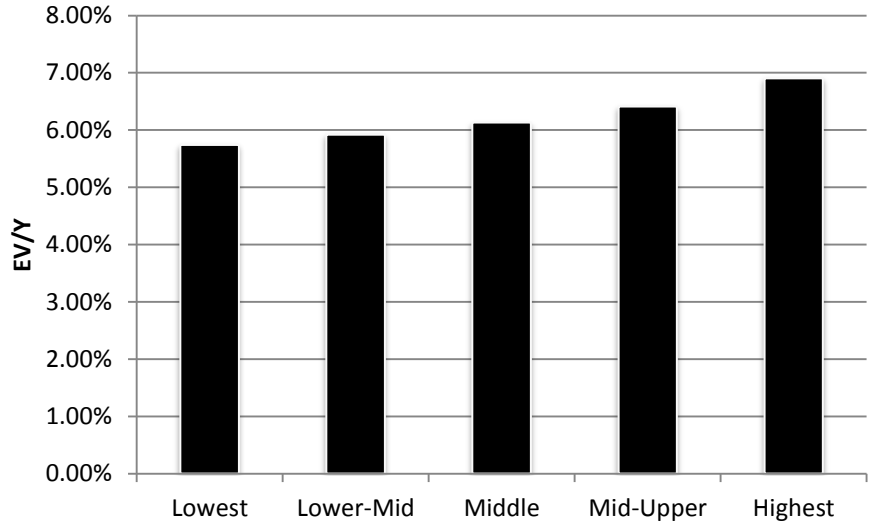


Figure 4 Effect of Existing Tobacco Taxes at Mean Net Expenditure Levels: All Households. Households Sorted by Net Income per Capita.

Class	y	cv	cv/y	ev	ev/y
Lowest	4,071.00	9.70	0.24%	9.68	0.24%
Lower-Mid	5,733.00	15.90	0.28%	15.87	0.28%
Middle	7,317.00	17.24	0.24%	17.19	0.23%
Mid-Upper	9,565.00	34.08	0.36%	33.92	0.35%
Highest	17,840.00	62.32	0.35%	62.17	0.35%

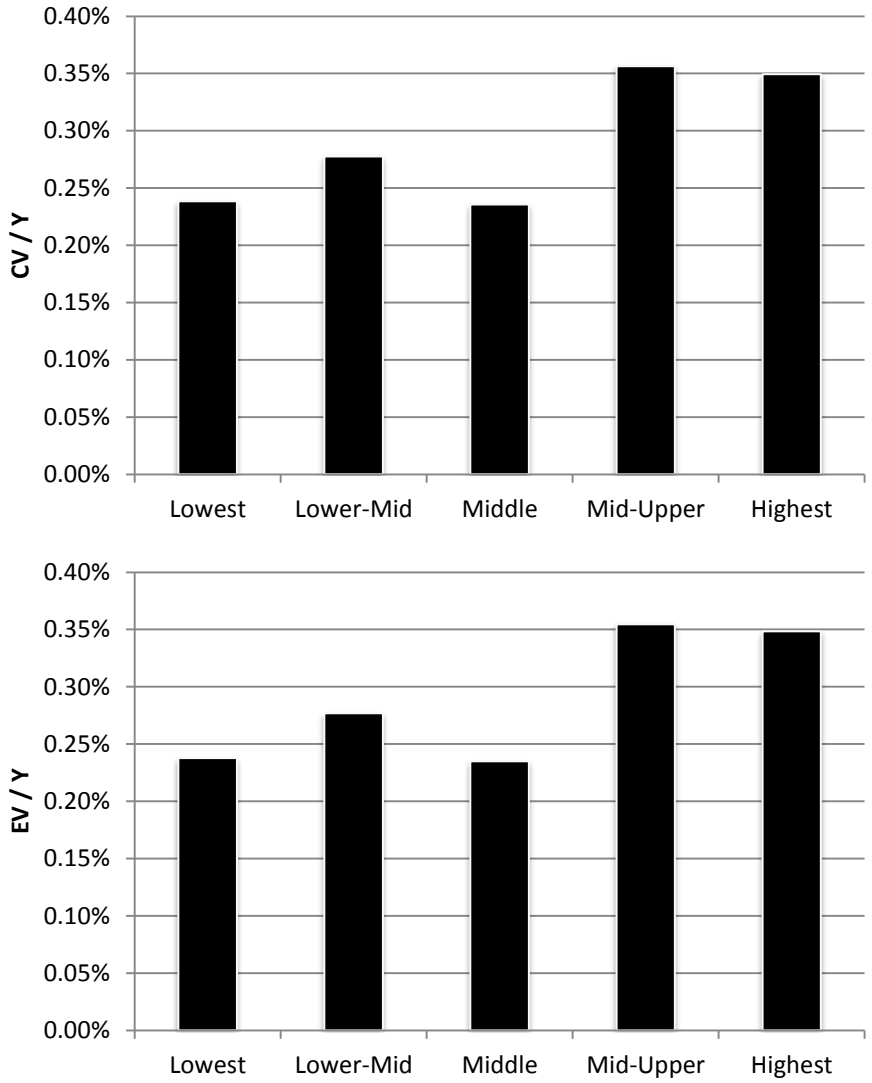


Figure 5 Effect of Existing Tobacco Taxes at Mean Net Expenditure Levels: All Households. Households Sorted by ***Equivalised Income***.

Class	y	cv	cv/y	ev	ev/y
Lowest	2,675.00	9.23	0.35%	9.22	0.34%
Lower-Mid	4,812.00	12.17	0.25%	12.15	0.25%
Middle	6,749.00	21.81	0.32%	21.74	0.32%
Mid-Upper	9,609.00	30.00	0.31%	29.94	0.31%
Highest	20,680.00	55.89	0.27%	55.79	0.27%

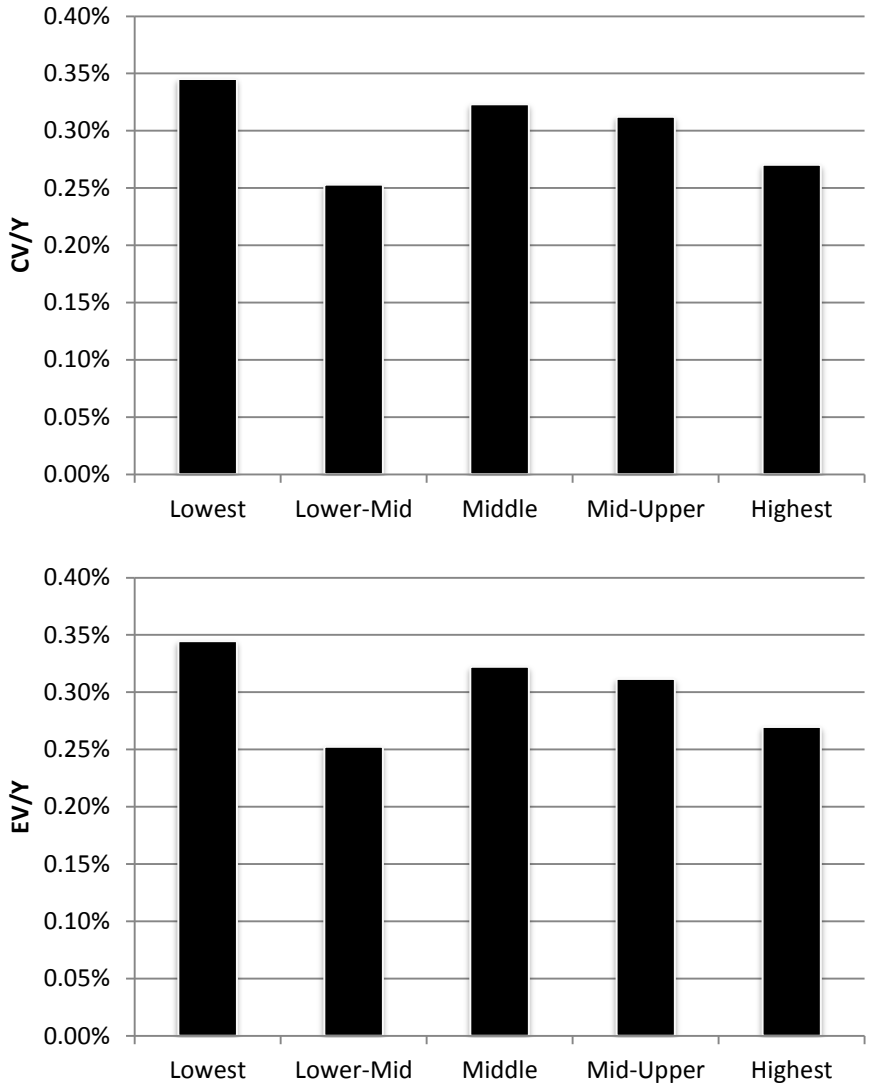


Figure 6 Effect of Existing Indirect Taxes at Mean Net Expenditure Levels: Smoking Households. Households Sorted by Net Income per Capita.

Class	y	cv	cv/y	ev	ev/y
Lowest	4,562.00	624.70	13.69%	579.40	12.70%
Lower-Mid	6,415.00	746.10	11.63%	700.70	10.92%
Middle	8,566.00	892.80	10.42%	817.40	9.54%
Mid-Upper	10,540.00	1,159.00	11.00%	1,040.00	9.86%
Highest	18,910.00	2,019.00	10.68%	1,848.00	9.77%

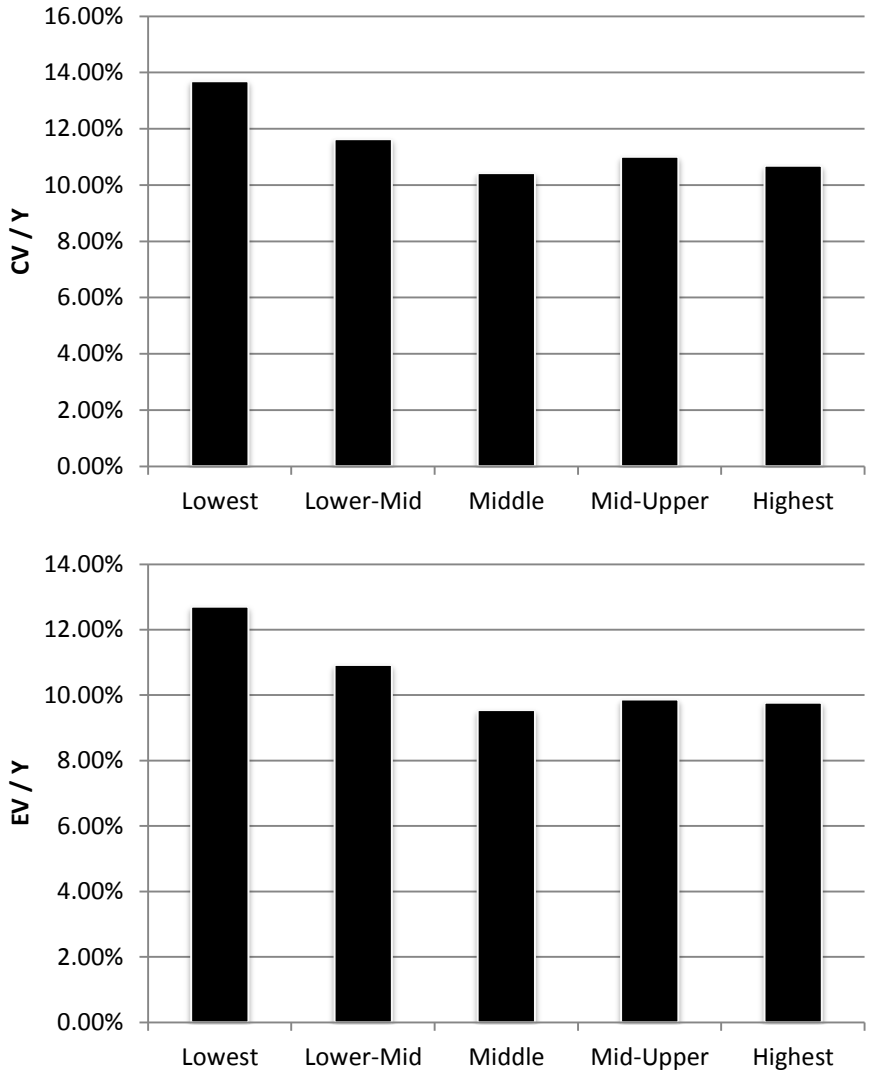
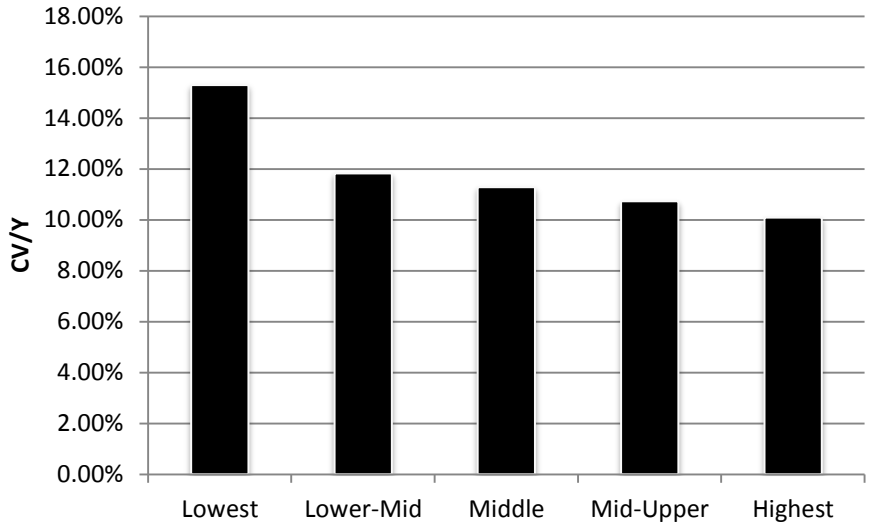


Figure 7 Effect of Existing Indirect Taxes at Mean Net Expenditure Levels: Smoking Households. Households Sorted by ***Equivalised Income***.

Class	y	cv	cv/y	ev	ev/y
Lowest	2,786.00	426.00	15.29%	390.30	14.01%
Lower-Mid	4,877.00	576.70	11.82%	539.20	11.06%
Middle	6,640.00	749.10	11.28%	687.70	10.36%
Mid-Upper	9,666.00	1,037.00	10.72%	951.40	9.84%
Highest	21,550.00	2,172.00	10.08%	1,998.00	9.28%



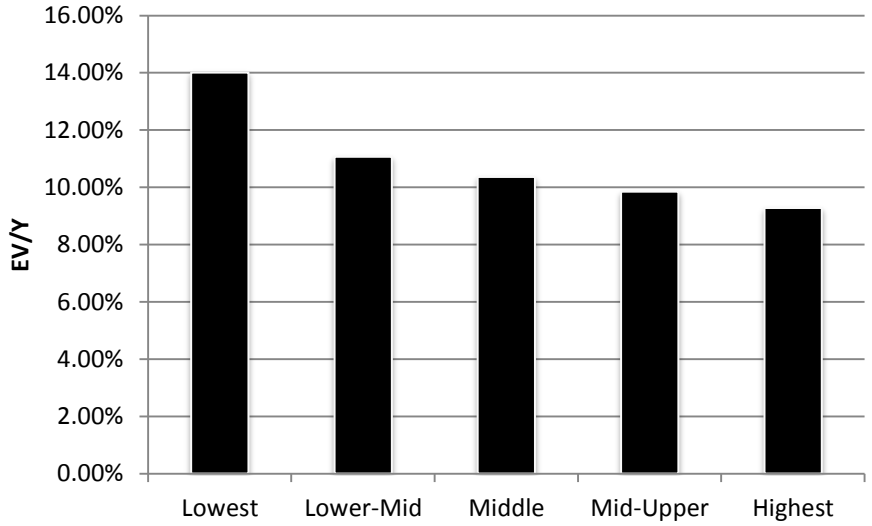


Figure 8 Effect of Existing Tobacco Taxes at Mean Net Expenditure Levels: Smoking Households. Households Sorted by Net Income per Capita.

Class	y	cv	cv/y	ev	ev/y
Lowest	4,562.00	351.40	7.70%	348.10	7.63%
Lower-Mid	6,415.00	336.70	5.25%	337.90	5.27%
Middle	8,566.00	347.10	4.05%	340.70	3.98%
Mid-Upper	10,540.00	431.30	4.09%	419.10	3.98%
Highest	18,910.00	638.10	3.37%	631.60	3.34%

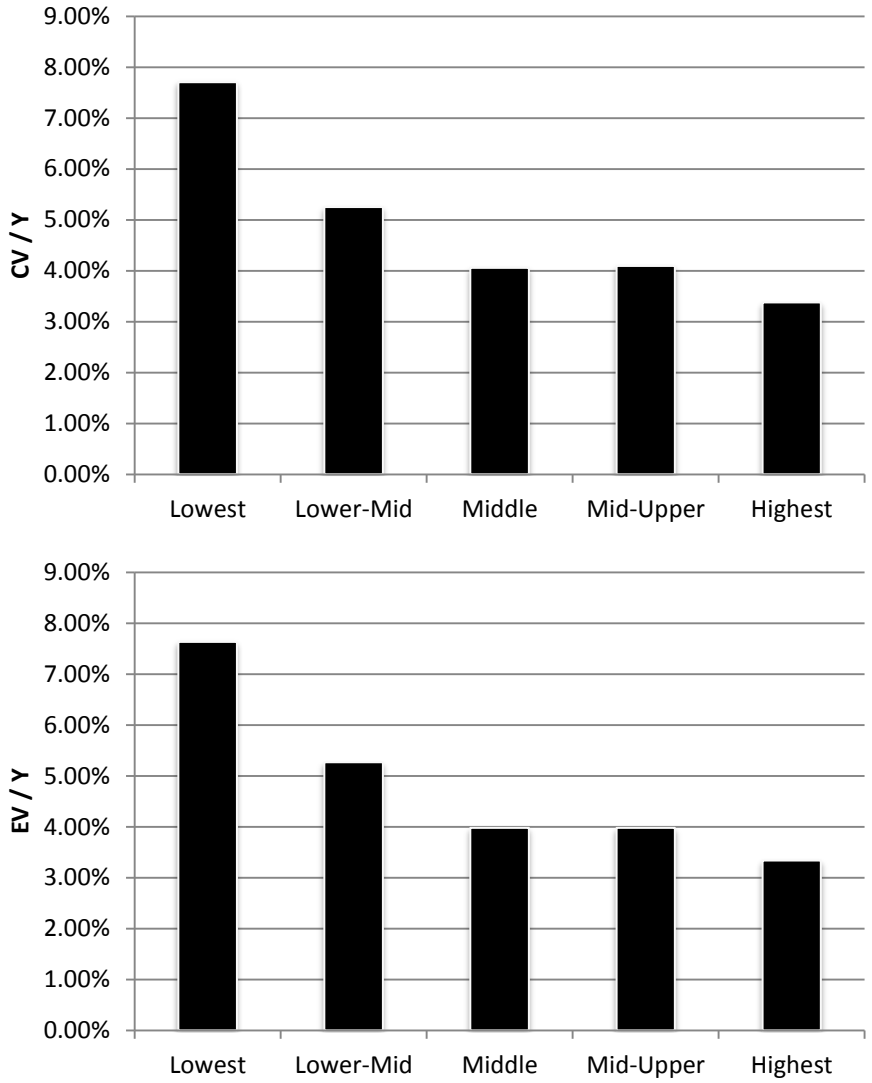
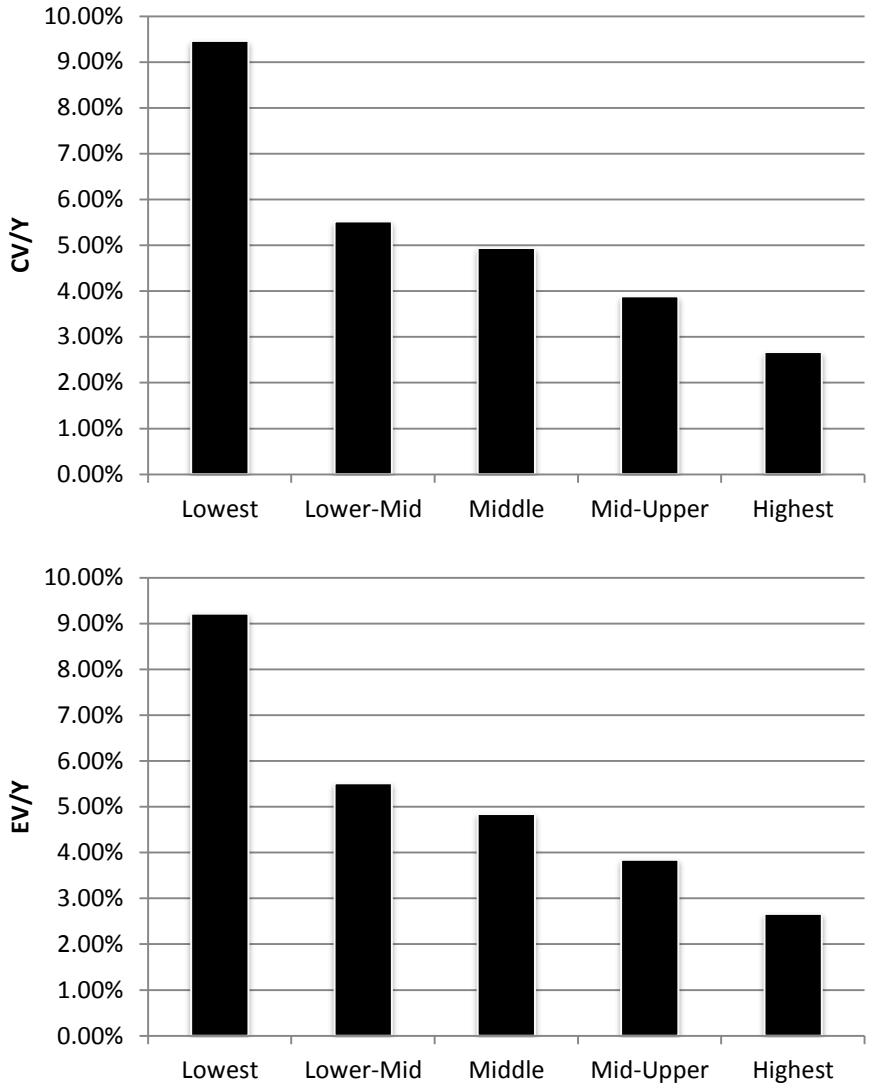


Figure 9 Effect of Existing Tobacco Taxes at Mean Net Expenditure Levels: Smoking Households. Households Sorted by ***Equivalised Income.***

Class	y	cv	cv/y	ev	ev/y
Lowest	2,786.00	263.50	9.46%	256.70	9.21%
Lower-Mid	4,877.00	269.20	5.52%	268.80	5.51%
Middle	6,640.00	327.70	4.94%	321.30	4.84%
Mid-Upper	9,666.00	375.50	3.89%	371.40	3.84%
Highest	21,550.00	575.10	2.67%	573.80	2.66%



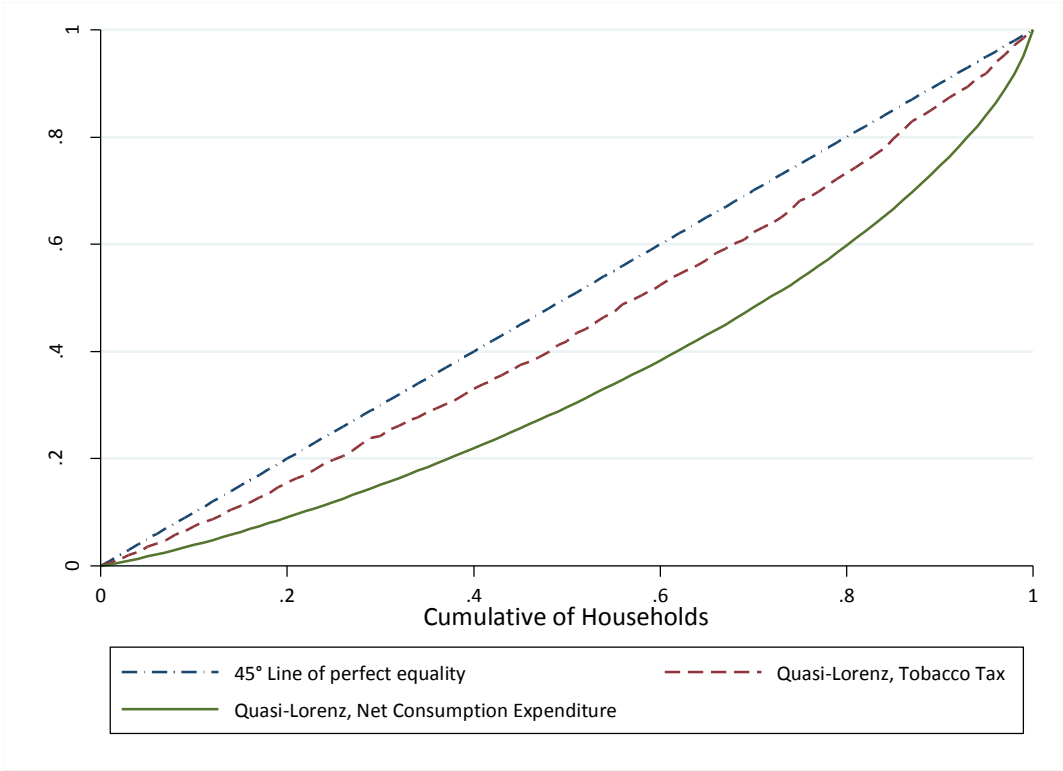


Figure 10 Cumulative Distributions of Net Expenditure and Tobacco Tax in Mexico, 2008**.**

Chapter II

An analysis of adult cigarette demand in Mexico: Evidence from the Mexican Family Life Survey

Introduction

Over the last 30 years the literature of tobacco demand has been mostly concerned with understanding the effect of price. The accumulated empirical evidence from developed and developing countries concludes that the "law of the downward sloping demand curve" applies for tobacco products (Chaloupka 1999). This has the immediate implication that tobacco-control policies, such as taxation, can be applied. However, a small area of research has directed its attention to analysing non-price factors affecting cigarette smoking such as the effect of socio-economic, demographic or psychosocial factors. Thus, the aim of this study is to analyse these using individual-level data.

This analysis centres its attention on Mexico, a developing country which in recent years has implemented a series of tobacco-control policies in order to curb the consumption of tobacco. As such, there is an on-going interest in understanding the demand in order to formulate and implement better government policies. To the best of my knowledge, this is the first study of this kind for Mexico, and one of the few studies from developing countries in which the unit of analysis is the individual and where the focus is on non-price determinants affecting the demand. Other studies analysing the demand for tobacco in Mexico have become available in the literature in recent years and these can be found in Jiménez-Ruiz, de Miera et al. (2008) and in Olivera-Chávez, Cermeño-Bazán et al. (2010). Still, their primarily aim is to analyse the role of price and income. Moreover, these studies used budgetary and aggregate time-series data respectively which provide few insights of smoking behaviour. There are, however, a few studies which have used individual data but they have analysed other specific issues commonly encountered within the economics of smoking literature. These include the compensatory smoking behaviour to a tax increase by Saenz-de-Miera, Thrasher et al. (2010) and the effect of the availability of single cigarettes on quitting behaviour by Thrasher, Villalobos et al. (2009) and Thrasher, Villalobos et al. (2011). Nonetheless, a study in which the aim is to investigate how the cigarette demand is determined using individual data is not yet available.

The methodology employed for analysing the demand for tobacco involved estimating a Two-Part model, usually referred in the literature as the Cragg's model (Cragg 1971) which investigates both factors affecting the decision to smoke and factors concerning the level of consumption. The evidence suggests that both decisions can differ so this is one of the reasons why the Two-Part model is the most common econometric model used in the literature of tobacco demand (Wilkins, Yurekli et al. 2004).

Data for this analysis comes from two rounds of the Mexican Family Life Survey (MxFLS), a representative survey with a rich source of individual information from many topics including cigarette consumption. This data joins the growing number of surveys from developing countries which can be used to provide a better description of smoking behaviour. In addition, it is worth mentioning that given the timing of the survey, data for this analysis will provide relevant insights of smoking behaviour among Mexicans before tobacco-control polices were enacted. Thus, the MxFLS can potentially be used to investigate the impact of such policies.

The remainder of this chapter is structured as follows. The next section provides an account of the methodology employed. In Section 3 data that is used in this study is described. In Section 4 the econometric results are presented and a discussion is provided in Section 6.

Methodology

The objective of this study is to estimate a cigarette demand equation using individual-level data. However, such data is characterized by having a large cluster of zeros denoting "no consumption" or it would denote that a large number of individuals are non-smokers. In fact, information is not observed (missing values), but in practice the information is set to zero. This creates some methodology issues that are worth discussing given that it has attracted a large body of research within the discipline of econometrics.

For instance, the cluster of zeros is too large to be ignored econometrically and so the conventional estimator, OLS, seem unsuitable for the purpose of this study therefore an alternative estimator has to be considered. In order to motivate the discussion, consider the simplest version of a demand model found in the literature, the linear demand model. This model is derived from a standard utility maximization framework under normal assumptions. The econometric specification takes the following functional form:

$$Q_i = \boldsymbol{x}_i'\beta + \varepsilon_i \qquad (1)$$

where Q_i stands for the consumption of cigarettes for individual i, $\boldsymbol{x}_i$ is a vector of individual's characteristics and ε_i is the error term[1]. It is worth noting that given the characteristics of the data at hand, Q_i is observed for some individuals whilst $\boldsymbol{x}_i$ is observed for all of them. This feature of the data is known in the literature as "censoring" therefore data is though to be "censored from below" (or left). This is better represented by:

[1] The expression in (1) can easily be translated into other functional forms such as the linear-log, log-linear and log-log (double log) model [see Wilkins et al. (2004) for a review].

$$Q_i = \begin{cases} Q_i^* \text{ if } Q_i^* > 0 \\ 0 \quad \text{if } Q_i^* \leq 0 \end{cases} \tag{2}$$

where Q_i^* is the latent variable. Censoring of information imposes some econometric difficulties. For example, OLS regressions with censored data are not appropriate because it would lead to biased and inconsistent estimates; there is no guarantee that $E(\varepsilon_i) = 0$ even asymptotically (Gujarati, 2003). An alternative approach which could be taken for estimating (1) is to restrict the analysis only to those observations for non-zero values. In this case, equation (1) can now be represented as:

$$\begin{gathered} Q_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i \\ Q_i = Q_i^* \text{ if } Q_i^* > 0 \\ (Q_i, x_i) \text{ not observed if } Q_i^* \leq 0 \end{gathered} \tag{3}$$

The model in (3) is thought to become truncated and can be estimated simply by OLS. This "restricted" model is usually referred to in the literature as the "smoking intensity demand", a distinction which will be important throughout this study. Conclusions drawn only from using such approach may be misleading due to undesirable properties of the estimates. Long (1997) shows that the error term in this regression does not have zero mean $[E(\varepsilon|Q > 0) \neq 0]$ which leads to inconsistent estimates.

Having a large cluster of zeros, however, is not the only methodological difficulty encountered. An additional characteristic commonly present is that for those individuals reporting consumption, the distribution of the data appears (highly) skewed to the right and exhibits non-constant variance (Tauras 2005). This has the implication that using data in its original structure may lead to inefficient or even inconsistent estimates depending of the econometric model used. Researchers have overcome this problem by applying a suitable variable transformation, usually by taking logarithms (Jones, 2007). Such transformation may particularly improve the efficiency of estimates. However, economic interpretation from such results cannot be given with transformed data therefore an appropriate re-transformation approach should be considered.

In what follows, a revision of the three most common estimators usually considered for censored data is provided. In addition, a short account of the "re-transformation problem" is given together with the estimator actually used to analyse the demand for cigarettes.

Econometrics models for censored data

Censored data may appear to present some methodological difficulties. However, dealing with the censoring of information is about recognizing its importance for understanding consumption. When choosing an econometric estimator, one has to make an assumption of mechanism explaining the zero. Although in practice the nature of this zero may not be entirely known, standard econometric approaches which are conventional in empirical work have attempted to deal with such an issue under different assumptions. The most common econometric approach is the Tobit model though more flexible estimators have emerged over the years such as the Two-Part model or Hurdle model and a closely related one, the Heckman model.

For this analysis, the preference lays on the Two-Part model (2PM hereafter) to estimate the demand for cigarettes in Mexico. The empirical evidence presented in this Chapter shows that this model provides the best fit given the data available. Although the Tobit estimator has been ruled out for this analysis, a discussion of this model has been included since it provides an ideal starting point for introducing the actual model being estimated. The exposition provided below is largely based on standard econometric text-books given that these estimators are conventional methodologies for the problem at hand. The exposition of the Tobit model and Heckman model is mainly based on that provided in Verbeek (2008) whilst the discussion of the 2PM is based on the exposition by Cameron & Trivedi (2005). Given the differences in notation found in different authors, a common notation has been utilised.

The Tobit Model

The analysis of zeros has attracted the attention of researchers, not only for analysing the demand for tobacco, but for a wide range of economic applications. The reason lies on the observation that zeros may represent two different processes. Therefore statistical methods treating these by one distribution, which is in the case of OLS, appear to be limited (Pudney, 1983). The first econometric model to successfully treat the censoring information with two distributions is due to Tobin (1958). This model, commonly knows in the literature as "Tobit" for its resemblance to the Probit model, would specify the demand for cigarettes in terms of an index equation such as:

$$\begin{aligned} Q_i^* &= \boldsymbol{x}_i'\beta + \sigma\varepsilon_i \\ Q_i &= Q_i^* \text{ if } Q_i^* > 0 \\ &= 0 \text{ otherwise,} \end{aligned} \qquad (4)$$

where Q_i^* is the latent dependant variable described before, $\boldsymbol{x_i}$ is a vector of individuals' socio-economic and demographic characteristics affecting the demand for cigarettes, σ is a scale parameter and ε_i is the error term which reflects the unobserved heterogeneity in the utility maximization solution process. The model in (4) is linear in regressors with an additive error that is normally distributive and homoscedastic such that:

$$\sigma\varepsilon \sim NID(0, \sigma^2) \tag{5}$$

The model represented in (4) corresponds to the standard Tobit model (Tobit type I in the literature) where the non-negativity constraint is imposed. In order to estimate the parameters in (4), a Maximum Likelihood (ML) routine is usually applied. The log-likelihood function of the Tobit model can be written as:

$$\log L(\beta, \sigma^2) = \sum_{i \in I_0} \log P\{Q_i = 0\} + \sum_{i \in I_1} \log f(Q_i) \tag{6}$$

where the indexes I_0 and I_1 represent the set of zeros and positive values respectively and $f(\cdot)$ is a specified function. The likelihood function reveals several features of the model that are relevant for choosing an appropriate estimator for modelling the demand for cigarettes. For instance, it is easy to see that the Tobit model decomposes the two processes involved with two different densities. In one hand, the density that represents the probability of $Q_i = 0$ is given by

$$P\{Q_i = 0\} = 1 - \Phi(\boldsymbol{x}'\beta/\sigma) \tag{7}$$

where $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function (CDF). On the other hand, the density representing the distribution of positive Q_i, which is just the truncated (at zero) normal distribution. The conditional expectation is given by:

$$\mathrm{E}\{Q_i | Q_i > 0\} = \boldsymbol{x}'\beta + \sigma \frac{\phi(\boldsymbol{x}'\beta/\sigma)}{\Phi(\boldsymbol{x}'\beta/\sigma)} \tag{8}$$

where $\phi(\cdot)$ is the standard normal probability distribution function (PDF). Technically then, this model accommodates the censoring of the information into a formal statistical model.

The Tobit model, however, relies upon several important assumptions that have been found unsuitable not only for this study but in many applications. First, it is important to point out that with this model the demand for cigarettes is generated by the following process:

$$Q_i = \max(Q_i^*, 0) \tag{9}$$

Therefore, it assumes that the nature of censoring corresponds to a "corner solution"[2]. Empirically, it would imply that 'at current prices and current income, the individual will never consume [cigarettes], and is therefore a corner solution to his or her utility maximization problem' Pudney (1983 p.159). As a result, substantial changes in prices or income could result in positive consumption. This may not hold true given the goods in question. In the literature of tobacco demand, it is more common to assume that the zero arises because the individual's condition of being a non-smoker. This means that the reason for not reporting positive consumption is due to abstention given, for example, the individual's health concerns. Verbeek (2008) pointed out that 'many non-smokers would not smoke even if tobacco were available for free' (p.233). Thus, one of the main limitations of the Tobit model is that it rules out the possibility of a "true zero". In other words, it rules out the possibility that individuals do not consume tobacco purely by choice and not because of the current conditions of the market. Perhaps failing to distinguish corner solutions from "true zeros" is one of the main reasons the Tobit model is usually rejected.

Second, even if the assumption of the "corner solution" is accepted, the structure of the Tobit model is viewed as too restrictive given that this model encompasses the two distributional processes into a single equation. Verbeek (2008) also explains that 'exactly the same variables affecting the probability of a non-zero observation determine the level of a positive observation and, moreover, with the same sign' (p.227). Empirically, this has been found unsatisfactory especially within the context of tobacco consumption. It may be the case that factors determining being a smoker and factors determining the level of consumption may be quite different.

Thirdly, the assumptions on which the model relies for unbiased and consistent ML estimates are too strong to work empirically. It has been stated that the error term in (4) must be homoskedastic and normally distributed. The empirical evidence suggests that these conditions are difficult to meet largely because data, in which this type of regression has been considered, is usually by nature highly skewed. By far the biggest concern is the presence of non-normally distributed errors which, in such a case, calculated estimates are not consistent.

For all the reasons mentioned above, the Tobit estimator is usually rejected in favour of its alternatives. Fundamentally, alternative estimators are models flexible enough in capturing the different determinants involving the process of the probability of encountering positive outcomes and the determinants involving the level of consumption. However, when the process in which the zeros are generated is unknown, most attention is paid to the convenience of alternative

[2] This is the corner-solution model which was previously represented in (3) above.

estimators simply because they rely on weaker distributional assumptions for consistent estimates. The 2PM model presented below, addresses successfully both issues.

The Two-Part Model (2PM)

The underlying assumptions that motivate the use of the Tobit model within the context of tobacco demand appear to be too restrictive. Fundamentally because the empirical evidence favours the view that consumption of cigarettes arises by the individual's choice of being smoker. As a result, there is an interest in disentangling the choice of consumption and the actual positive tobacco consumption which is in fact observed. Thus, an alternative estimator for the Tobit model is usually applied, namely the 2PM. This model provides more flexibility for determining the probability of observing consumption and the observed outcome. As an alternative estimator, cigarette demand is modelled by two separate processes: the first process denoted as "participation" which accounts for the censoring mechanism and the second process denoted as "smoking intensity", which accounts for the outcome or level of cigarette consumption. In its general form, the model can be written as:

$$f(Q|x) = \begin{cases} \Pr[d = 0|x] & \text{if } Q = 0 \\ \Pr[d = 1|x] f(Q|d = 1,\ x) & \text{if } Q > 0 \end{cases} \qquad (10)$$

where $f(\cdot)$ is an specified density function and d is an indicator variable equal to 1 for a cigarette consumer, 0 otherwise. This model is also usually referred in the literature as the Cragg's model (Cragg, 1971) or simply the Hurdle model. The model is appealing for its simplicity in estimation. Usually the participation equation is estimated by means of a Probit model. In turn, the intensity equation can be estimated say, by OLS with the sub-sample of positive values of Q. The expression in (10) can be represented by:

SMOKING PARTICIPATION

$$\begin{aligned} d_i &= \boldsymbol{x}_i'\beta + \varepsilon_{1i} \qquad & \varepsilon_{1i} \sim N[0,1] \\ d_i &= 1 \text{ if } Q_i > 0, \qquad & 0 \quad \text{otherwise.} \end{aligned} \qquad (11)$$

SMOKING INTENSITY

$$\begin{aligned} Q_i &= \boldsymbol{x}_i'\beta + \varepsilon_{2i} \qquad & \varepsilon_{2i} \sim N[0, \sigma^2] \\ Q_i &= Q_i \text{ if } Q_i > 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \qquad (12)$$

One important feature of the 2PM is that it relies on the assumption that ε_{1i} and ε_{2i} are uncorrelated. This means that the demand for cigarettes is identified based on "selection on observables" (Cameron and Trivedi, 2005). There is, however, a discussion on whether correlation between ε_{1i} and ε_{2i} should be allowed which in such a case, a closely related

estimator namely the "heckit" estimator would become relevant. The heckit estimator is a consistent and alternative estimator to the Heckman model (Heckman, 1979) and although is not an efficient estimator it is computationally simpler than the usual Heckman ML estimator. The heckit estimator is also a type of 2PM in which the participation equation is based on a Probit model just as in (11) but the intensity equation is given by:

$$Q_i = \boldsymbol{x}_i'\beta + \rho_{12}\sigma_{12}\lambda_i + \eta_i \qquad (13)$$

where η_i is the error term and an λ_i is an extra regressors added known as the "inverse Mill's ratio". The inverse Mills ratio is given by:

$$\lambda_i = \sigma_{12}\frac{\phi(\boldsymbol{x}_i'\beta)}{\Phi(\boldsymbol{x}_i'\beta)} \qquad (14)$$

but as λ_i is not observable, it is simply replaced by its estimate $\widehat{\lambda_i}$ which is obtained from the binary response model (the Probit model used in the first step). After obtaining an estimate of $\widehat{\lambda_i}$, the model in (12) can be estimated also with the sub-sample of positive values in the data. The resulted estimate for the Mill's ratio $\rho_{12}\sigma_{12} = \sigma_{12}$ is of great interest. ρ_{12} is the correlation coefficient between the error terms in the participation and smoking intensity equation and so if correlations does not exist, $\sigma_{12} = \rho_{12} = 0$, the OLS approach suggested in (12) can be applied (Verbeek, 2008)[3].

However, several issues surrounding the heckit estimator as an alternative to the 2PM are worth mentioning. For example, this estimator is usually motivated by "sample selection" grounds which are a closely related issues to the mechanism explaining the cluster of zeros. So far two mechanisms have been presented; the "corner solution" which has been ruled out, and "abstention" or "choice" which is more in accordance with the existent literature[4], but there is also the possibility that zeros or at least some of those zeros are due to "under-reporting". Under-reporting may arise given that many individuals may feel ashamed to accept having consumed cigarettes given that smoking in some communities is not socially acceptable. Under-reporting is particularly significant in empirical work because when non-responses appear randomly in a sample it may not be such a problem; inferences about the population can be made but this is not longer the case if these responses are not random. Under-reporting then creates a bias in the estimation therefore identification of estimates must account for the sample selection. Statistically, evidence of sample selection bias arises when $\sigma_{12} \neq 0$.

[3] Caution should be taken in applying the correct formula for standard errors.

[4] Pudney (1983) explains additional mechanisms involved when it comes to encountering zeros in data which labelled as "infrequency of purchase". This relates to the idea that for certain goods, individuals did not report consumption given the timing of the survey is too short. Alternatively, he discusses the possibility of observing zeros by "purely involuntary nature". For example, even if an individual is willing and able to acquire certain good s/he may not be able to do so because it is just not available.

Nevertheless, even when sample selection is not a problem of concern, many researchers still see the heckit estimator as an ideal alternative to the 2PM. However, depending of the research question, caution should be taken when choosing the appropriate econometric model given the each estimator produces results with different interpretations. An interesting discussion concerning which estimator should be used in the context of unhealthy goods such as tobacco and alcohol is given by Madden (2008). His analysis pointed out several criteria that should be taken into account before choosing between the 2PM or the heckit estimator. On theoretical grounds, Madden (2008) doubts if the heckit estimator is relevant for analysing, in particular, the demand for tobacco given that the prediction is based on "potential outcomes". This contrasts from the 2PM where prediction is based on "actual outcomes". In the literature of tobacco demand, the main concerned is on the latter. He provides the following conclusion:

> … it seems likely that what we are trying to model is actual smoking, as opposed to potential smoking. It follows that we are interested in the effects of covariates on actual as opposed to potential smoking in which case the two-part model seems more appropriate.
>
> Madden (2008 p. 302)

Moreover, at a more technical level the issue of potential versus actual outcomes relates to the fact that the 2PM is better suited to estimate the "unconditional mean" of Q_i or $\mathrm{E}[Q|x]$ therefore inferences about [unconditional] partial effects, $\Delta[Q]/\Delta[x]$, can be made (Mullahay, 1998). In contrast, the heckit estimator is designed to estimate the "conditional mean" of Q_i or $\mathrm{E}[Q|x, Q > 0]$ and/or to correct for selectivity bias. Thus, "unconditional partial effects" are more difficult to calculate[5]. Nevertheless, even if this difficulty is overcome, it remains an empirical question whether correlation between the two equations is relevant and/or sample selection turns out to be an issue that should be addressed given the problem at hand. Otherwise, the 2PM is the better alternative available to the Tobit model.

Lastly, in empirical applications the practicality of the heckit estimator or the Heckman model in general has been questioned. This has to do with whether the same regressors should be used in the participation and intensity equation or if the exclusion of some variables should be imposed. This is important because under circumstances of no exclusion, adding $\widehat{\lambda_i}$ as an additional regressor may lead to a multicolliniarity problem. Usually, the same regressors are used in both equations to test whether the same factors affecting the probability of being a smoker are the same as those factors affecting the level of consumption, both in sign and statistical

[5] Dow and Norton (2003) discussed in detail this issue. They, in fact, show the appropriate formulae to estimate the unconditional mean and the respective marginal effect and elasticity under the hecktit estimator. However, the issue persists whether the interest is on the conditional mean or unconditional mean.

significance. For the problem at hand, it seems reasonable to follow this approach although there is no reason why it should be so even when collinearity is not suspected. But when a collinearity problem is present, exclusion restrictions must be imposed and so many researchers apply it for a more robust identification. Exclusion is commonly imposed in the intensity equation. An empirical difficulty arises given that there is no clear guidance to which variables should be excluded. Thus, in the absence of clear choices for exclusions, particularly when collinearity persists, on practical grounds the heckit estimator is not an ideal estimator to use.

In summary, from the most common models of censored data this analysis will employ the 2PM for modelling the consumption of cigarettes. This model has been found to be flexible enough to recognize the most plausible mechanism explaining the cluster of zeros which according to the existence literature is by "choice" or "abstention". In addition, given that there is no reason to believe that correlation between the two equations (participation and intensity) and selection bias would be an issue of concern, the heckit estimator will not be considered. Therefore, the methodology will be restricted to estimate a Probit model for participation in the first stage and for the second stage, an alternative procedure to the usual OLS is employed to avoid the re-transformation problem. This is further explained in the next sub-section.

The Modified 2PM

The econometric model used in this study aims to estimate the unconditional mean of Q_i which is represented by $\mathrm{E}[Q|x]$. Given the structure of the model presented in (11), $\mathrm{E}[Q|x]$ is given by:

$$\mathrm{E}[Q|x] = \Phi(x\beta_1)[x\beta_2] \qquad (15)$$

where it uses the statistical property that $\Pr[Q > 0|x] = \Phi(x\beta_1, \varepsilon_1)$ and $\mathrm{E}[Q|Q > 0, x] = x\beta_2 + \mathrm{E}[\varepsilon_2|Q > 0, x] = x\beta_2$. Thus,

$$\mathrm{E}[Q|x] = \Pr[Q > 0|x] \times \mathrm{E}[Q|Q > 0, x] \qquad (16)$$

Under normal circumstances, the model can be estimated in two stages. The first stage corresponds to a Probit model for smoking participation and the second stage may correspond to an OLS regression with the sub-sample of positive observations of cigarette consumption. This addresses the censoring in information. However, the continuous distribution from this sub-sample follows a skewed distribution which would affect the efficiency of OLS estimates. To overcome this problem, many studies have considered applying a suitable transformation of the dependent variable Q_i.

The logarithm transformation is the most commonly used in empirical work. Such transformation appears to deal with some of the major distributional issues encountered such as skewness and kurtosis but not when the log-scale error term ε is heteroscedastic (Manning and Mullahy 2001). For instance, a problem arises when estimating $\mathrm{E}[Q|Q>0,x]$. Under the logarithm transformation what is being estimating is in fact $\mathrm{E}[\ln(Q)\,|Q>0,x]$. Therefore $\mathrm{E}[Q|Q>0,x]$ would be given by:

$$\mathrm{E}[Q|Q>0,x] = exp(x\beta_2 + 0.5\sigma_\varepsilon^2) \tag{17}$$

This re-transformation is possible if the error term is normally distributed and homoscedastic. If the error term is not normally distributed but is homoscedastic (constant variance), $\mathrm{E}[Q|Q>0,x]$ can be obtained by the following re-transformation:

$$\mathrm{E}[Q|Q>0,x] = exp(x\beta_2) \times \mathrm{E}\big(exp(\varepsilon)\big) \tag{18}$$

Alternatively, if the error term is heteroscedastic but normally distributed then:

$$\mathrm{E}[Q|Q>0,X] = exp(x\beta_2)f(x) \tag{19}$$

though it is more problematic to obtain. This is because the presence of heteroscedasticity creates a ***bias*** generated by the variance of ε which depends on some function in x. Thus, the transformation of $\mathrm{E}[\ln(Q)\,|Q>0,x]$ to $\mathrm{E}[Q|Q>0,x]$ should take into account the magnitude of such bias which is possible by calculating the corresponding estimate of $f(x)$. This is the "re-transformation problem". The estimate of $f(x)$ may be represented as:

$$\mathrm{E}(exp(\varepsilon)|x) \quad \text{or} \quad v(\varepsilon|x) \tag{20}$$

In all, whenever a distributional issue arise, it is convenient to apply a generalised linear model (GLM) assuming the log-link relationship (Manning and Mullahy 2001). This is because, instead of estimating $\mathrm{E}[\ln(Q)\,|Q>0,x]$ the model estimates $\ln(\mathrm{E}[Q|Q>0,x])$ and so the conditional mean is given by

$$g[\mathrm{E}(Q)] = \boldsymbol{x}'\beta_2, \;\; \varepsilon_2 \sim F \tag{21}$$

where F is a specified distribution from an exponential family. This procedure then overcomes the re-transformation problem.

The GLM model as the second stage of the 2PM has been used extensively in the context of health care cost and expenditure in the health economics literature (Manning 2006). Within the context of tobacco demand, this procedure has been used by Tauras (2005) and Tauras et al

(2007). The only remaining issue, however, is to identify the exponential family F. This is the mean and variance function for the observed raw-scale variable Q, conditional on x. This function is given by:

$$\text{var}(Q|x) = \alpha[E(Q|x)]^{\lambda} \tag{22}$$

where the parameter of interest is λ. The parameter reveals the most appropriate distribution to use in the GLM model. In order to identify the most appropriate distribution, the parameter of interest is obtained by means of the Park test (Park, 1966). This test involves the estimation of the following equation:

$$\ln\left(\left(Q_i - \widehat{Q_i}\right)^2\right) = \lambda_0 + \lambda_1 \ln\left(\widehat{Q_i}\right) + v_i \tag{23}$$

The estimate of $\widehat{\lambda_1}$ determines which probability function provides the best fit. So for example, a Gaussian distribution is used when $\lambda_1 = 0$ implying that the variance is unrelated to the mean. A Poisson distribution is used when $\lambda_1 = 1$ implying that the variance equals the mean. A Gamma distribution is used when $\lambda_1 = 2$ implying that the variance exceeds the mean and a Ward distribution is used when $\lambda_1 = 3$.

Marginal Effects

In order to investigate the effect of (continuous) explanatory variables on the unconditional mean, partial effects or marginal effects are calculated. These can be obtained by the following expression:

$$\frac{\partial \text{E}[Q]}{\partial x_k} = \frac{\partial(\text{Pr}[Q > 0] \times \text{E}[Q|Q > 0])}{\partial x_k} \tag{24}$$

which can also be written as:

$$\frac{\partial E[Q]}{\partial x_k} = \left(Pr[Q > 0] \times \frac{\partial}{\partial x_k} E[Q|Q > 0]\right) + \left(E[Q|Q > 0] \times \frac{\partial}{\partial x_k} Pr[Q > 0]\right) \tag{25}$$

However, as the majority of covariates included in the model are binary (see next section), their incremental effect is defined as:

$$\begin{aligned} E(Q|x, x_d = 1) - E(Q|x, x_d = 0) \\ = \{\langle Pr(Q > 0|x, x_d = 1) - Pr(Q > 0|x, x_d = 0)\rangle \\ \times E(Q|x, x_d = 0)\} \\ - \{Pr(Q > 0|x, x_d = 0) \times \langle E(Q|x, x_d = 1) - E(Q|x, x_d = 0)\rangle\} \end{aligned} \tag{26}$$

Estimation

The estimation of the 2PM is relatively easy in STATA as in-built commands for the Probit model and GLM are available. The calculation of the marginal effects, however, is more involved. The procedure consists in calculating the discrete change of each binary variable from 0 to 1, holding all variables constant. These finite changes are calculated at means and their respective standard errors are obtained by using the bootstrapped method. The computation were possible using the STATA code produced by Deb, Manning et al. (2010).

Data

General information

The data used in this analysis comes from the Mexican Family Life Survey (MxFLS), a multi-purpose and longitudinal survey from a nationally representative sample of Mexican households and their members. The survey's design is projected to span over a period of ten years[6]. The MxFLS is conducted by research centres and higher education institutions of Mexico and of the United States. The survey follows a similar design to the Indonesian Family Life Survey (IFLS), an on-going longitudinal survey that began in 1993 where its multi-purpose nature was first applied to a developing country.

Given the longitudinal and multi-dimensional vision of the survey, the objectives of the MxFLS are (Rubalcava and Teruel, 2006):

I. Understand the social, economic, demographic and health transitions in Mexico.
II. Provide evidence regarding the effect of migration to the United States.
III. Investigate the effects of behavioural responses to economic changes and well-being of the Mexican population.
IV. Assess the effectiveness of anti-poverty programmes implemented by the Mexican government.

These objectives are achieved by collecting information from a diverse range of topics at community, household and individual level. At community level, it collects qualitative and quantitative information on local infrastructure and services. At household level, data is collected primarily on consumption expenditure patterns. At individual level, information from a broad

[6] Data for the first sweep which constitutes the baseline of the survey has been collected and information for the second and third sweep has been collected as well. Data for the first and second sweep are the only available at the time of writing this paper. Field work for a fourth sweep has been programmed for 2012 and a date for field work for the fifth sweep is yet to be announced.

range of topics which includes usual socio-demographic characteristics, migration history, health care utilizations, biomarkers, etc. Figure 1 shows the complete structure of the survey where it is possible to see the type of information available for analysis. The survey's information has been organized into seven sections plus a "control book" (see below). Only two sections provide information at household level (section 1 and 2) and the remaining sections provide information at individual level.

This analysis uses data both at household and individual level but not at community level. At household level, data from section "Household expenditures" is used to obtain information about consumption expenditure and some dwelling characteristics such as the households' ownership status. At individual level, data is used from the section on "adult information" (3.A and 3.B) which included sub-sections on education history, employment history, adult time allocation, tastes and habits, health status and health insurance. These sub-sections constitute only six out of twenty-two available at individual level in the survey.

Survey's design

The sampling survey design is probabilistic, stratified, multi-staged and by cluster (INEGI 2004). The survey's baseline (first sweep) collects information from 8,440 households and information for 35,677 individuals. This information was collected between April and June of 2002. The second sweep collects information from 8,437 households and information from 38,223 individuals. Data collection spans throughout the period of 2005 and 2007 due to the re-contact process.

Information at household level was provided by one member of the household on behalf of all members belonging to each dwelling. Individual information was obtained by a one-to-one interview with the household member if the individual was at least twelve years of age. For those individuals younger than twelve years of age, a member related to the individual provides the information. For the MxFLS, an individual who is at least 15 years old is considered an adult, so those who meet this requirement are likely to answer the appropriate sections of the questionnaire specifically designed for them.

The MxFLS provides a "control book" which is a separate section answered by one member of the household (usually the head of household) on behalf of all members belonging to the dwelling. This control book collects data on basic socio-demographic information for all individuals in the database. However, as individuals were answering different sections in the survey, some basic information is repeatedly provided. This helped to complement information if missing answers are found across sections or in the control book. At the same time, it was detected that a small percentage of information provided in sub-sections did not match with the

information contained in the control book. Therefore, this analysis uses primarily self-reported information of basic demographic characteristics, with the exception of the information regarding the gender of the individual which is taken directly from the control book. Nevertheless, whenever missing observations for self-reported information are found in the data, these are substituted with the information contained in the control book in order to avoid dropping individuals from the sample.

Another important element to consider in analysing data for the MxFLS is the response rate for each section or sub-section. For example, according to the manual for the baseline, the overall response rate for sections "Adults" is 91% while for "Household Expenditure" is 95% (Rubalcava and Teruel 2006). However, within sections the rate of response varied according to each sub-section analysed. The variability of the response rate is due to the refusal to answer or because an individual belonging to the household was not present at the time of the interview. In a related issue, it is worth mentioning that the survey offers a "proxy book" which is information for individuals in the household who were missing at the time of the survey. The information contained in this book is provided by a member of the household who was able to complete the information on behalf of the missing member. This information can be used to increase the rate of response if there is an interest in minimising the loss of information throughout sub-sections; this information is only available for adult members and no attempt has been made in using such proxies in this analysis. Furthermore, each sub-section is accompanied by its own sampling weight which needs to be used if inferences about the population are to be made. This analysis uses weights to correct for the population mean only for information provided at household level.

Lastly, it is important to point out that data and all appropriate documentation for understanding and handling the information for the MxFLS (manuals, questionnaires and codebooks), is in the public domain and is freely available on the internet with a translation into English for access to a wider audience[7].

Reference Sample

Given the interest of this study, the reference sample being analysed here are those (adult) individuals who completed the sub-section "Tastes and Habits" in the survey. This section contains smoking-related questions which it was found to be useful in determining the individual's smoking status. Thus, with the available information it is possible to identify smokers, former smokers and non-smokers in the sample. Additionally, the questionnaire for this sub-section asks the individual about other lifestyle question such as the kind of drink s/he

[7] For more information visit the web-page: http://www.ennvih-mxfls.org

accompanies with meals at home or in parties and it contains questions to establish whether s/he engaged in physical exercise. All this information is used in the analysis.

Smoking status

The questionnaire starts by determining the individual's smoking state with the following question:

- Do you/did you ever had the habit of smoking cigarettes?

If the individual answers positively, retrospective information about the age of smoking initiation, either in age or initial year is collected:

- How old were you, or what year did you start smoking frequently?

The questionnaire continues by asking about the number of cigarettes smoked when they were at the peak of the addiction or habit:

- By the time you were smoking the most, how many cigarettes did you end up smoking on average per week?

Then, a question on whether cessation as occurred by asking:

- How old were you, or what year did you quit smoking, on a regular basis?

The individual is given the option to respond to the question or indicate that s/he has not quit smoking in a frequent way. If the latter situation applies, information on the number of cigarettes s/he currently smoke and the money value of expenditure is collected.

- Currently how many cigarettes do you smoke on average per week?
- At the present time, how much do you spend on cigarettes per week?

With the above two questions it is possible to separately identify current smokers with ex-smokers. Although the question on the number of cigarettes currently smoked would suffice in determining active smokers, both questions are considered. This is because not all individuals who answered positively on the number of cigarettes necessarily provided information on money spent and vice versa therefore this study defines an "active smoker" as those who answered positively to either question or both. At the same time an individual is considered a "quitter" if answered positively of having been a smoker but not being currently smoking. Lastly, a final question regarding the time the individual has been a smoker is provided:

- If you could put together all the time you have uninterruptedly smoked, how many years would that be? Please do not consider the time you have not smoked.

In all, this subsection provides the number of individuals that have been considered as the useable sample in this analysis. All available data from two sweeps of the MxFLS has been used which correspond to data from the baseline and the first follow-up. Data from the baseline and follow-up correspond to information for 19,804 and 20,606 individuals respectively[8]. It should be noted that although the MxFLS is of longitudinal nature, the sample size has been allowed to "naturally refreshed" by the second sweep in order for the survey not to lose its representativity given the observed rate of attrition [see Rubalcava and Teruel (2008)]. However, as this study will analyse the demand for cigarettes within a cross-sectional framework, any considerations regarding attrition and the short panel available are left for the next chapter.

Dependent and explanatory variables

The dependent variable in the model is the number of cigarettes smoked on average per week. This is a measure of typical consumption as it does not control for differences in the price or quality (Yen and Jones 1996). Factors influencing individual preferences are given by a set of socio-demographic and psychosocial variables which constitute demand shifters and they have been used repeatedly in the tobacco demand literature. The complete description of each explanatory variable used in this analysis is given in the Appendix.

Among socio-demographic variables considered, age in years and its squared value has been added to capture non-linear effect reflecting possible life-cycle patterns in consumption or cohort effects in cigarette consumption (Jones 1989). A proxy for income has been included corresponding to the natural logarithm of monthly consumption per capita which will control for the prevailing economic condition. Furthermore, indicators for gender, urban residence, ethnicity, marital state, education, household role, health insurance, housing tenure and regional dummies have been included.

Concerning gender, it is expected a strong and positive male effect since the majority of active smokers in the data are male. This is consistent with the current rates of smoking prevalence in Mexico and for the wider international experience (Initiative 2004). Thus, this indicator is expected to account for the gender-differentiated smoking habits. An indicator for whether the individual resides in urban areas is also included. Su and Yen (2000) suggest that urbanization make the individual to be subject to a more peer pressure and other metropolitan influences such as advertising. This may help to explain why higher rates of smoking prevalence are observed in urban areas than in rural areas. Race and ethnicity has also been explored in

[8] The control book reports that the number of individuals of at least 15 years of age in the first sweep is 23,306 which mean that this sub-section has an actual response rate of 84.9%.

previous work, particularly those concerning in the US to capture differentiated smoking habits among minorities reflecting cultural differences and tastes (Su and Yen 2000) or to reflect the different means of handling addictions and social stigma (Yen 2005a). This analysis includes an indicator on whether the individual considers himself or herself belonging to an indigenous group. Broadly speaking, this is the only recognised ethnic group in Mexico.

Different indicators for levels of education are included to test the idea that individuals with more education may be more aware of the health risk associated with consumption of unhealthy goods such as tobacco (Jones 1989). Moreover, regional indicators have been included corresponding to the area of residence within Mexico. These indicators are expected to account for differences in the tobacco market which may include regional attitudes towards smoking. However, given the absence of information on prices, these indicators may serve for proxies for regional price differentials as well Su and Yen (2000) and Yen (2005a). In addition, an indicator whether the individual lives in a household which is owned by a member of the household, is included as a proxy for wealth although it may also reflect "time preferences among social groups" or an "indication of the stress of the individual's living environment" (Jones 1989). Lastly, an indicator whether the individual benefits from health care is included. Having health insurance in Mexico is mandatory by law either from either a public or private provider. In this respect, its effect in the analysis should be taken with caution as it may better reflect work status. More specifically, it may better reflect whether an individual is employed in the formal sector as opposed to underreported activities. In fact, health insurance covers all of his or her legal dependants and it does not exclude anyone with medical pre-conditions or smoking status. This indicator, then, may be taken as proxy for employment reflecting lifestyle (Yen, 2005). However, it is important to highlight that not all of those who reported having health insurance coverage obtained it as direct beneficiaries.

Concerning psychosocial variables considered in this analysis, two set of indicators are included. On one hand, a set of indicators that would reflect risky behaviour associated with smoking and on the other hand, a set of indicators that would reflect potential deterrents from smoking or from current consumption of cigarettes. Amongst risk factors, indicators for consumption of alcohol, hot drinks such as coffee, and living with another smoker. Living with another smoker is commonly used in empirical work as it reflects the influence of social interaction on smokers' behaviour within his or her inner group (Jones 1989). Among deterrents from smoking, an indicator for signs of physical exercise is included hypothesizing that it would reduce the individual's consumption of cigarettes if a healthy lifestyle is pursuit. In addition, an indicator of whether the individual's role in the household is being a house maker is considered. The expectation is that when such a role is assumed, young children may be present in the

household. Empirical evidence suggests that living with young children would discourage adults from smoking or at least to moderate tobacco consumption as shown by Aristei and Pieroni (2008) therefore this indicator is intended to capture similar effects.

Descriptive statistics

A table of summary statistics for the entire sample and by smoking status is given in the Appendix for both the first and second sweep. Only a statistical analysis of the sample for the first wave is provided here[9].

In the sample, 44.5% of individuals are males (N = 8,803) and 55.5% are female (N = 11,001). 57.6% live in urban areas (N = 11,401) of whom 56.0% are females and 44.0% are males. The average age in the sample is 37.64 years and no statistical differences in age are found between genders (37.92 years for males and 37.40 years for females, Pr(|T| > |t|) = 0.0341). In terms of marital status, 29.4% of individuals reported being single, 61.2% married or cohabitating, 4.7% being divorced or separated and 4.7% are widowed. In all options of marital status, the proportion of females is found to be greater than males.

In regard to schooling, 8.9% report not having received formal education. 41.2% have attended primary, 26.2% secondary, 14.1% high-school and 9.6% university level. Analysing differences between education and gender, it is found that the proportion of females with no education, primary, secondary and high-school is higher than for males whilst the proportion of individuals with university education is greater for males. The only association between gender and the level of education which is not statistically significant is for secondary (Pr = 0.569). Amongst individuals in the sample, 16.0% (N = 3,178) recognize themselves as belonging to an indigenous group which stratified by gender it is found that 45.2% are males and 54.8% females. Not surprisingly, 61.4% live in rural areas whilst the remaining 38.6% live in urban areas.

Regarding some socio-economic variables considered in this analysis, the sample identifies 72.9% of individuals who are homeowners. The proportion of individuals living in their own house is higher in urban areas (60.8%) compared to those in rural areas (39.2%). The proportion of individuals who reported themselves as house makers is 32.2% and just 1.9% are househusbands. 44.2% of individuals in the sample benefit from health insurance (44.9% are male and 55.1 are female). Breaking down this figure by strata, it is found that 71.9% of individuals living in urban areas benefit from health insurance whilst this figure is 28.1% for those in rural areas. This

[9] Selected means and proportions have been analysed across groups, mainly by gender and stratum. For the case of continuous variables, *t*-tests have been used to analyse differences in means. For the case of indicator variables, statistical associations have been analysed using Chi-square tests.

information reflects one of the characteristics of the Mexican economy in which the "formal sector" is more concentrated in urban areas.

The sample also reveals that 38.1% of individuals consume alcohol either in social gatherings or with meals. The proportion of males who consume alcohol is greater than females (67.2% for males and 32.8% for females) and the association is statistically significant (Pr = 0.0000). Moreover, only 0.3% reported having consumed hot drinks. In addition, 19.6% of individuals reported routinely taking physical exercise, of which 56.6% are male and 43.4% female; the association is statistically significant (P = 0.0000). Also, 29.1% of individuals live with someone who is or has been smoker of whom, 51.7% are females and the remaining 48.3% are males.

The distribution of individuals across the five regions considered is as follows: 20.4% live in the South-East, 19.3% in the Centre, 19.8% in the West, 21.2% live in the North-West and 19.4% live in the North-East. The proportion of individuals living in rural areas is lower for any region reported but not for much. The proportion of an individual in urban areas within regions ranges approximately from 55% to 58%.

The average income measured as the monthly household consumption expenditure per capita is 1,207.25 MXN which for urban areas is 1,489.21 MXN and 829.89 MXN for rural areas Pr(|T| > |t|) = 0.0000. However no statistical differences are found between genders even though income appears greater for males than females (1,233.62 MXN for males and 1,186.20 MXN for females, Pr(|T| > |t|) = 0.4599). Similarly, income in urban areas appears to be greater for males than females, but the difference is not statistically significant (1,455.33 MXN for females and 1,532.48 MXN for males, Pr(|T| > |t|) = 0.4643). Likewise income for males and females is not significantly different in rural areas (818.70 MXN for females and 843.54 MXN for males, Pr(|T| > |t|) = 0.6211).

Description of smokers

The sample shows that 18.3% of individuals (N = 3,631) had or have the habit of smoking cigarettes of whom 72.8% are males and 27.2% females; the association is statistically significant (Pr = 0.0000). From that proportion, 35.7% are identified as quitters (N = 1,296) and 64.3% as active smokers (N = 2,335). As a whole, active smokers account for 11.8% of the baseline of whom 74.2% are males and 25.8% females (Pr = 0.0000). Furthermore, 65.7% of smokers live in urban areas whilst 34.3% are in rural areas (Pr = 0.0000).

The overall age of smokers is 38.04 years of age [38.21 for males and 37.53 for females Pr(|T| > |t|) = 0.3585], and for non-smokers the average age is 37.59 years [37.85 for males and

37.40 for females, Pr(|T| > |t|) = 0.0870]. Among smokers and non-smokers, age is not statistically different between them, Pr(|T| > |t|) = 0.2337. Across strata it was found that smokers appear to be younger in urban areas than in rural areas [36.80 years of age for urban areas and 40.39 for rural areas, Pr(|T| > |t|) = 0.0000].

Data on the age of smoking onset is available for most of those who had or have been smokers of cigarettes; only 18 individuals did not provide information. The mean age at smoking onset is 18.35 years ranging from 5 to 70 years old[10]. Males appear to begin smoking earlier than females. The mean age at smoking onset is 17.40 years for males and 20.90 for females, Pr(|T| > |t|) = 0.0000. It is important to highlight that 54.9% of those who provided this information reported having started smoking before 18 years of age.

Information about the age of smoking cessation has been gathered from 1,473 individuals[11]. However among quitters, only 84.1% provided this information (N = 1,090). The mean age at smoking cessation is 35.72 years of age and statistical differences are found across gender [for male this is 36.96 years and for female it is 32.89 years, Pr(|T| > |t|) = 0.0000].

Cigarette consumption

The survey collects two types of information regarding the consumption of cigarettes[12]. One corresponds to the number of cigarettes smoked at the peak of the addiction and the other one corresponds to the current number of cigarettes consumed which is the dependent variable of interest. For the first case, information is available for 94.16% of individuals who had or have the habit of smoking (N = 3,419). It is reported that at the height of their smoking, respondents smoked an average of 58.58 cigarettes per week. For males it is 64.85 and for females is 41.98 cigarettes and the difference across gender is statistically significant, Pr(|T| > |t|) = 0.0000. The mean number of cigarettes smoked per week for active smokers (current consumption) is 40.58. For males it is 43.95 cigarettes and for females is 30.86, and the difference is statistically significant, Pr(|T| > |t|) = 0.0000. Also, current consumption of cigarettes does not appear to differ by strata; for urban areas it is 39.75 and for rural is 42.14 cigarettes [Pr(|T| > |t|) = 0.2710].

[10] It is conceivable that there are some errors in reporting this information. Extremes values come from individuals who reported the year of smoking onset rather than age.
[11] It has been identified that there are 1,296 quitters in the sample. However, it would appear that some active smokers provided information on the age of smoking cessation which could be taken as the age of the latest attempt to quit.
[12] For those reporting packages of cigarettes, the information is converted into number of cigarettes.

Self-reported information and heaping

The number of cigarettes consumed by current smokers is an element of interest in this analysis. However, given that this information has been self-reported by the respondents in the MxFLS, possible *recall biases* are of concern particularly because they distort the distribution of the observed consumption. Recall biases in this context occurred when the respondent was unable to accurately remember that an *event* (smoking) took place which may lead to either underreporting or over reporting of information. This is called the "memory effect" which occur when 'respondents may not remember precisely how many cigarettes they smoked each day, and therefore report a rounded approximation' (Wang and Heitjan 2008). This phenomenon, usually called as "heaping", can be found in other contexts of empirical work such as job-search duration data (Torelli and Trivellato 1993), age, length of emergency room visits, and others [see (Wang and Heitjan 2008) and the references therein]. Although heaping is a *nuisance* in terms of the quality of data, it is difficult to confirm its presence statistically. It is then when such phenomenon is confirmed, when a method can be suggested to counteract its effect in empirical work.

Perhaps, only a graphical tool can reveal its presence. Figure(s) 2 show the frequency of self-reported consumption of cigarettes for the dataset constructed. Notice that the information has been shown by "cigarettes" and "packages of cigarettes", two possible ways of answers offered to the respondent. In the first case, the majority of respondents revealed that they smoked on average per week less than 20 cigarettes which is less than a package. In the second case, those who reported packages of cigarettes, the majority smoked less than 10 packages. Heaping would be present if certain *spikes* can be detected depending on some multiples (say, multiples of 10 or 20). It is the view of the author that heaping is not present in this case as such multiples are not identifiable. However, it is worth recognizing that when the option of reporting "packages of cigarettes" has been offered (instead of only "cigarettes") this may reflect more accurately consumption expenditure rather than simply consumption.

Consumption expenditure and unit prices

Out of all active smokers, 88.3% (N = 2,062) reported money expenditure on cigarettes. The mean amount spent on a weekly basis is 306.32 MXN and differences across gender are statistically significant [for males it is 319.32 MXN and 264.53 MXN for females, Pr(|T| > |t|) = 0.0000]. Urban smokers spend 312.97 MXN on cigarettes whilst rural smokers spend 294.32 MXN and the difference is not statistically significant, Pr(|T| > |t|) = 0.2377.

Having collected information on consumption expenditure, it opens the possibility of using it to calculate *unit prices*. These are simply obtained by dividing the money spent by the actual

amount of cigarettes smoked. Potentially, this information could be used to calculate the price elasticity of demand which is of interest. This exercise has been conducted, but unfortunately the results are not satisfactory to this end. Unit prices data appear to be *too noisy* for obtaining meaningful results. This can be shown in Figure(s) 3 which displays a scatter plot showing the relationship between the money spent on cigarettes and actual consumption. A non-parametric regression line has been super fitted in the plot in order to detect a possible relationship between these two pieces of information[13]. The first set of plots reveals that the presence of outliers would make the calculation of unit prices unreliable. In general, the amount of money spent on cigarettes does not match consistently with the number of cigarettes consumed. This makes the value of unit prices to go beyond what one would expect to be according with the prices in the market. For instance, in 2002 the average price of filtered cigarettes was around $ 0.6 MXN whilst for unfiltered cigarettes was $ 0.30 MXN. The calculated unit price is well beyond these two figures. Removing outliers from the sample does reveal that, as expected, the more cigarettes consumed, the more money is spent [see lower plots in Figure(s) 3]. Nonetheless, even after this the average unit price does approximate with the observed price in the market. This could be explained by the presence of outliers.

For completeness, calculated unit prices are related by age and by stratum. This relationship is shown in Figure(s) 4 where, instead of presenting a scatter plot, only fitted non-parametric regression lines have been presented. There is a stark difference between the observed relationships in wave 1 to wave 2. In the first case, urban middle-age smokers tend to buy more expensive cigarettes than younger or older smokers do, though the opposite is observed for rural smokers. For wave 2 however, the relationship is clearer both for rural and urban smokers. Unit prices fall with age. In a period of anti-smoking policies, this may simply reflect that, once the habit of smoking is well established, smokers will tend to smoke lower quality cigarettes.

Econometric Results

Model Selection

Several econometric models have been proposed as possible estimators for modelling the demand for cigarettes. However, the Tobit model has not been found to be appropriate for this analysis given the strong assumptions being imposed. This leaves as possible estimators the Heckman model or the 2PM model. Although the preference lays on the use of the latter,

[13] This is the Lowess smoothing available in Stata using the in-built command `lowess`.

preliminary estimates are given from both estimators to test whether correlation between the participation and intensity equation is important to consider. This point is commonly a source of concern.

Another source of concern is the nature of the distribution of the data. From those individuals reporting consumption, data appears to be highly skewed, therefore using it in its original structure may lead to inefficient estimates (Manning, 2006). Researchers have overcome this problem by applying a suitable variable transformation usually by the use of logarithms. Thus, in order to improve the efficiency of estimates this transformation has been applied when estimating different variants of the two-part model[14]. The regressions results are presented in Table 2. This table includes the usual 2PM (assuming independence) and two versions of the Heckman model. These include the selection two-part model or "heckit" and the full Maximum Likelihood Estimation (MLE).

The results seem consistent as most of the coefficients appear statistically significant and with expected signs. However, the centre of attention is on testing whether the disturbances from the participation and intensity equation are correlated. The *de facto* assumption within the 2PM is that correlation does not exist therefore $\rho = 0$. The correlation coefficient is implied to be 0.628 for the case of the heckit estimator. Although a formal correlation test is not directly provided within the model, the standard *t*-test of the Mills' coefficient can be taken as such (Verbeek 2008). The null hypothesis that $\sigma_{12} = 0$ is not rejected. A similar result is encountered using the MLE version of the Heckman model. Notice however, that estimates from this estimator are similar in sign and significance and not substantially different in magnitude. Nevertheless, the important element to look at is the estimated value for rho which is found to be 0.1461. The null hypothesis that $\rho = 0$ (no correlation) is not rejected either (Prob > Chi2 = 0.5351). This supports the use of the 2PM.

The 2PM

Smoking participation

The first part of the 2PM is an estimation of a Probit model for smoking participation which aims to investigate factors influencing the likelihood of being a smoker. All the explanatory variables constructed for this analysis have been used as regressors. For the participation equation (as well as for the intensity equation, see below), the following variables have been omitted: "Marital Status - Single", "Education level - No education" and "Region - Centre". Also,

[14] The final model estimated will use an alternative approach to a logarithm transformation. This is only done for comparison purposes.

indicator variables for two sub-groups have been added which correspond to “attritors” in the first sweep (individuals not successfully followed-up) and “new comers” for the second sweep (individuals only observed by the second sweep). More details will be outlined about such sub-groups in the next chapter.

Probit ML estimates for current smokers are presented in the Table 2 for both sweeps. A test for model specification has been carried out using the Pregibon’s link test which takes the following form[15]:

$$y^* = \delta_0 + \delta_1\left(x\hat{\beta}\right) + \delta_2\left(x\hat{\beta}\right)^2 + v \qquad (27)$$

The test did not show evidence of misspecification as the coefficient for the polynomial is not statistically significant at conventional levels or $\delta_2 = 0$.

Regarding the results for the first sweep, few coefficients appear not to be statistically significant, being these for “Educational level - Secondary”, “Health Insurance”, “Hot drinks”, “Exercise” and for “Region – West” and “Region – North – East”. In terms of continuous variables, age exhibits a parabolic effect where the probability of reporting being a smoker reaches a maximum at an age of 41 or $-[0.0304]/\{2 \times [-0.000367]\} = 41.4$. The results also show that income positively affects the likelihood of being a smoker although the presence of higher polynomials has been detected. Regarding indicator variables, most of the coefficients appear statistical significant and show expected signs. However, at this stage only qualitative interpretations are as follows. For instance, males, living in urban areas or with another smoker and drinking positively affect the likelihood of report being a smoker. In contrast, individuals who belong to an indigenous group or those who are house makers or homeowners are less likely to report being a smoker. In addition, having received an education seems to negatively affect the likelihood of an individual reporting being a smoker relative to the reference individual who has not achieved any level of education although the coefficient for secondary education is not statistically significant. In terms of indicators for regions, those living in the South-East and North-West appear to be less likely than those in the Centre (the area of Mexico City) to be a smoker. Lastly, regarding marital state, all coefficients are positive and highly significant but judging by their magnitude, it would appear that the likelihood of being a smoker differs by each marital state in comparison to those who are single.

Results from the second sweep are consistent with those presented above. In fact, the probability of reporting being a smoker reaches also a maximum at an age of 41. However, two differences are apparent from the results. Being part of an indigenous group does not have an

[15] `linktest` command in STATA.

effect on the likelihood of individuals of being a smoker. Furthermore, secondary education does appear to negatively affect smoking participation. The remaining variables exhibited the same influence as shown for the first sweep.

Smoking intensity

In the second stage of the estimation of the 2PM, a GLM model for the conditional demand has been applied with the sub-sample of smokers. The model assumes that the expected value of Q_i depends on a linear predictor through the logarithm-link function given a set of explanatory variables. The exponential probability distribution for the most appropriate model given the characteristics of the data was determined by applying the Park test as specified in expression (23). The result from this test which are shown in Table 1, suggests that for the 2PM applied to data for the first sweep, the raw-scale variance exceeds the mean (quadratic in the raw-scale prediction) therefore the Gamma distribution provides the better fit. In contrast, for the 2PM applied to data for the second sweep, the raw-scale variance is proportional to the mean therefore the Poisson distribution is used.

Estimates for the smoking intensity equation are given as well in Table 3 for both sweeps. The variable age also exhibits a parabolic effect where, conditional on participation, consumption of cigarettes reaches a maximum at an age of around 58 years.

Controlling for age and income, only specific factors appear to affect the level of cigarette consumption. For instance, neither marital state nor the level of education affect the quantity of cigarettes consumed as none of these coefficients are statistically significant. From the remaining results, it is possible to identify that conditional upon participation, males significantly consume more cigarettes than females. Similarly, smokers living in urban areas appear to consume more that those living in rural areas. In contrast, individuals who identify themselves as indigenous or homeowners significantly consume fewer cigarettes than individuals who are not. Likewise, individuals who engage in physical exercise consume fewer cigarettes than those who do not. Lastly, several regional differences are encountered. Individuals living in the North-West significantly consume more cigarettes than those in the Centre and so those in the North-East and in the West relative to the reference region.

Results from the second sweep reveal that, conditional on participation, consumption of cigarettes reaches a maximum at an age of around 71 years which is considerably higher figure that the one previously obtained. Variables with a positive impact on cigarette consumption include being male and the consumption of hot drinks. Moreover, having higher education

positively affects the level of consumption conditional of being a smoker and individuals living in the West and East consume more cigarettes than those from the Centre.

Effect of indicator variables

In order to provide a more general interpretation of results and given that the majority of covariates included in the model are binary, their effect on the unconditional mean is investigated by calculating partial effects of expression (26) above. The calculated binary effects are first presented for all individuals in the Table 4.

In general, the estimated binary effects confirmed the qualitative effects observed both in the participation and smoking intensity equation. Results for the first wave, for example, reveal a strong male effect. Men are found to be 8.5% more likely to be a smoker than women, all other things being equal. In addition, males appear to be heavy smokers as the overall result show that they consume almost four more cigarettes a week than women. A similarly strong effect is identified for those who live with another smoker. According to the calculated probability they are 20.1% more likely to be a smoker than those who do not. Moreover, the overall binary effect on the unconditional mean show that individuals living with another smoker, smoke seven cigarettes more a week than those who do not. This reflects the risk associated when smoking is present in close interaction with others. Other effects associated with the consumption of cigarettes include residing in urban areas where the largest urban effect is identified on consumption rather than in the likelihood of being smoker. In fact, residents of urban areas smoke almost one cigarette more a week than those in the rural areas. Consumption of alcohol makes and individual to be 6.2% more likely to be a smoker and the overall effect reveal that alcohol increases the level of consumption by around two cigarettes a week. Physical exercise does not seem to affect the likelihood of being a smoker but its overall effect reveals that it significantly decreases the level of consumption by less than one cigarette a week. Among characteristics which significantly affect the likelihood of being a smoker include belonging to an indigenous group, being a house maker or a homeowner. These characteristics make the individual around 2% less likely to be a smoker, all other things being equal. In addition, the overall effect of being an indigenous, house maker or homeowner makes the individual to smoke less by around one cigarette in comparison to those who do not share these characteristics.

Regarding the marital state of the individual, the results show that being divorced or widow make an individual around 8.0% and 6.6% more likely to be a smoker respectively relative to those who are single. This contrasts with those who are married; they are only 1.4% more likely to be a smoker. Those who are divorced or widowed smoke more cigarettes than others, around three and two cigarettes respectively. Also, regarding regional differences the results show that

those residing in the West, North-West and in the North-East smoke around one cigarette more than those residing in the Centre. Conversely, those residing in the South-East smoke around one cigarette less relative to those residing in the Centre.

Focusing only on the overall effect of each binary variable (unconditional level), results from the second sweep appear to differ to those from the first sweep in two aspects. Firstly, it appears that the level of consumption has decreased by the second sweep. In the case of estimates that are statistically significant, the magnitude of the mean is smaller than the magnitude presented for the first sweep. Secondly, whilst only three estimates were not statistically significant in the first model, now the number of insignificant estimates rose to nine. These differences may be explained by changes in the environment that are causing significant changes in smoking behaviour. However, specific factors continue to influence the level of consumption. For example, psychosocial affects appear positively affect substantially the level of consumption. Drinking alcohol or hot-drinks increase consumption by more than two cigarettes a week. Those who live with another smoker, will tend to smoke five cigarettes more a week than those who do not. Conversely, factors that would reduce the consumption of cigarettes by approximately one cigarette include being part of an indigenous group and being a house maker or homeowner.

Given the strong male effect detected, partial effects have been estimated by calculating the increment effect of a discrete change of the binary “Male” on the overall results. Therefore, the unconditional mean has been decomposed by gender. Table 5 and Table 6 show the calculated binary effects by women and men respectively. In general, no systematic differences have been observed by analysing partial effects by women or men. The only conclusion drawn from the results is that men and women only differ on level of smoking intensity.

Discussion

This study has aimed to investigate non-price factors affecting the demand for cigarettes in Mexico. In terms of the results presented, it has been found that amongst the most important demographic factors affecting the consumption of cigarettes are those concerning the marital status of being divorced or widowed, which positively affects the demand. These are consistent risk factors found in previous work as Sheu, Hu et al. (2004) and Tauras (2005). Education seems to negatively affect the demand, a result that has been found consistently in studies from developed and developing nations. However, it is worth mentioning that educations seems to be more relevant in determining smoking participation, just as in the study from China by Mao, Sung et al. (2008) or the study from Greece by Raptou, Mattas et al. (2005) where the findings are

similar to those reported in this analysis. In addition, regional differences have been identified which suggest the presence of differences in habits and attitudes towards smoking within the country. Regional differences were also investigated by Yen (1999), Yen (2005a) and Yen (2005b). The results also reveal that being male and living in urban areas significantly increases both the probability of being a smoker and the level of consumptions. These demographic factors have been previously investigated by Sheu, Hu et al. (2004), Tauras (2005) and Mao, Sung et al. (2008). In contrast, assuming the role of a house maker decreases both the probability of being a smoker and the level of consumptions. Lastly, it has been found that individuals who belong to an ethnic group are less likely to smoke and conditional of smoking, consume fewer cigarettes. This may due to differences in cultural background and stigma; the effect of race and ethnicity on tobacco consumption has been studied by Su and Yen (2000).

Among socio-economic factors which were found to be relevant in determining the demand, being a homeowner appears to decrease the likelihood of being a smoker and conditional of smoking, it decreases consumption; similar result has been observed by Jones (1989). Concerning psychosocial factors affecting the risk of smoking and consumption, it is found that drinking alcohol, being engaged in physical exercise and living with another smoker positively affect the demand for cigarettes. Drinking alcohol has not been previously explored amongst studies of tobacco demand, but the results presented here confirms its association with addictive goods which reveal unhealthy and risky behaviour amongst individuals (Baumert, Ladwig et al. 2010). However, although physical exercise was not found to decrease the likelihood of being a smoker as expected, conditional upon smoking it does reduce the level of smoking therefore having a healthy lifestyle does help to reduce consumption. Overall a mixture of socio-demographic and psychosocial factors has been found to be relevant in determining the demand for cigarettes. In addition, it has been found that by estimating a 2PM, it was possible to determine the direction of influence and statistical significance of each factor on participation and the level of consumption.

There is, however, one important limitations of this study which is that the effect of price is not investigated. This does not mean that its effect is not recognized in determining the demand, but data on prices is simply not available. Further research may look into the effect of price if suitable proxies are found. In addition, the results provided correspond to estimates from pooling all the information from males and females. Recent studies of tobacco demand by Yen (2005a), Yen (2005b) and Bauer, Göhlmann et al. (2007) have explored differences in consumption by stratifying the analysis by gender. Their results suggest that factors affecting demand may differ between them therefore future research would explore this possibility. Lastly, the analysis provided corresponds to a cross-sectional analysis of a sample of individuals and so it was not

possible to take into account the panel structure of the data. This is because the 2PM is not designed to control for unobserved heterogeneity. In the next chapter, the analysis is extended to utilise all the available information contained in the MxFLS with an alternative estimator which analyse factors affecting the demand for cigarettes in a life cycle context.

ANNEX TO CHAPTER II

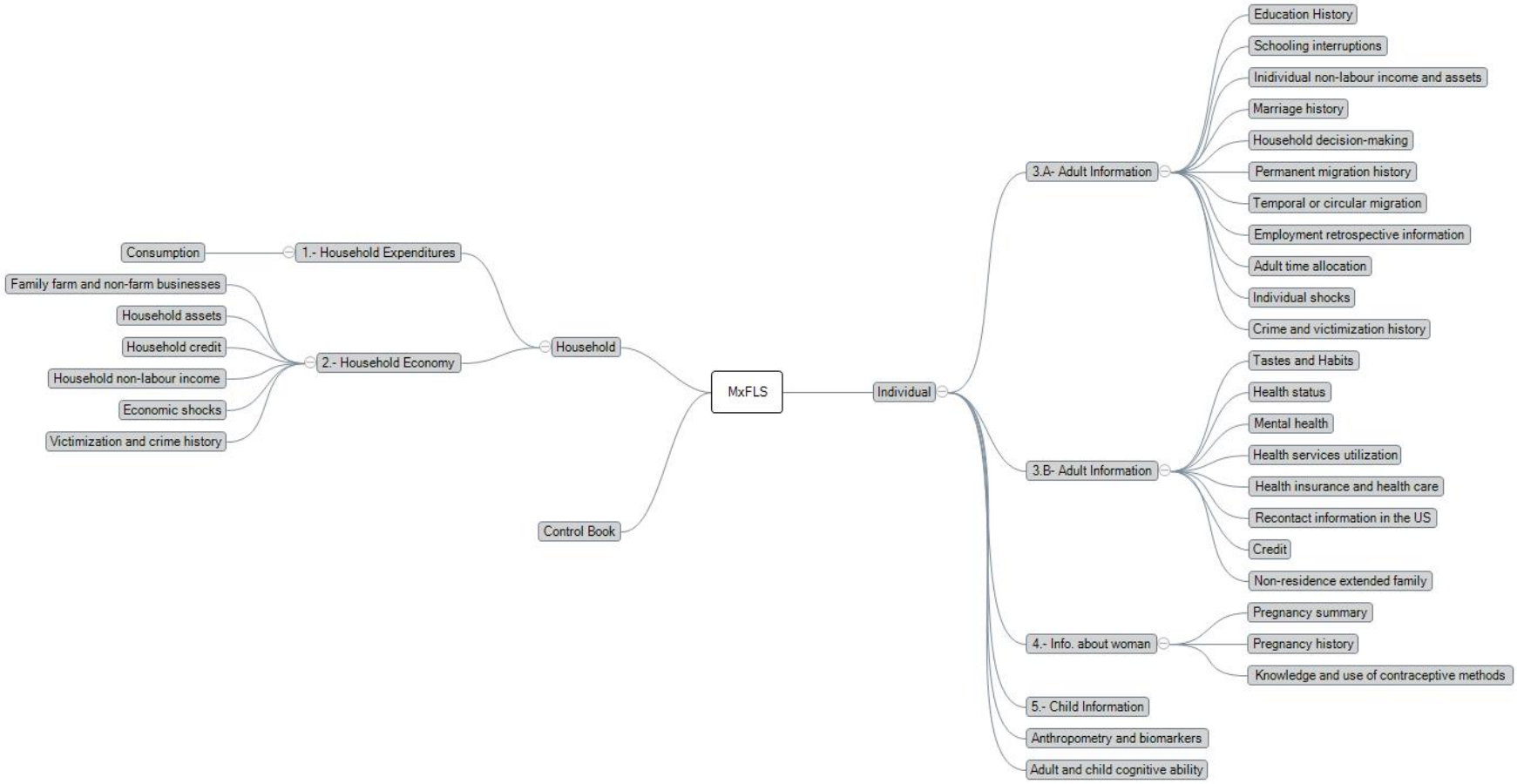
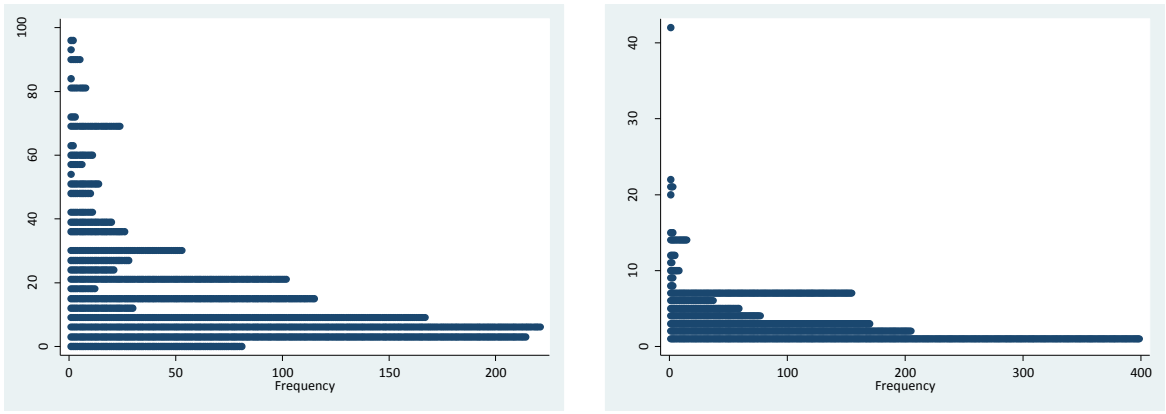
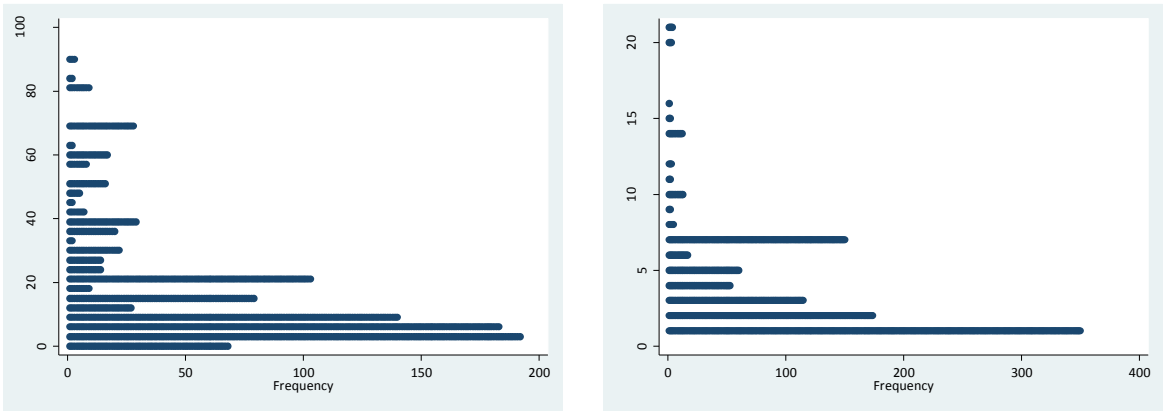


Figure 1. Structure of the MxFLS.

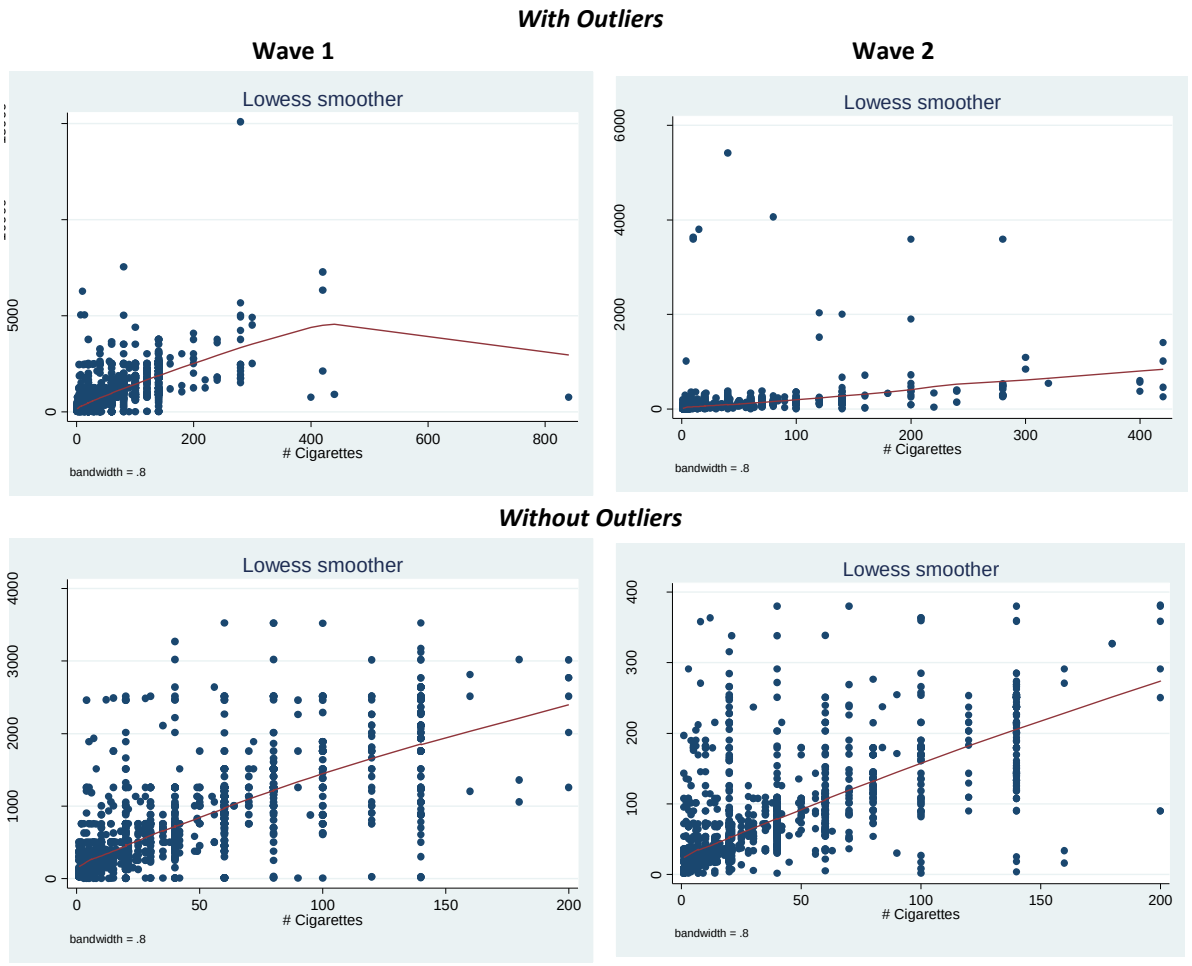


Wave 1

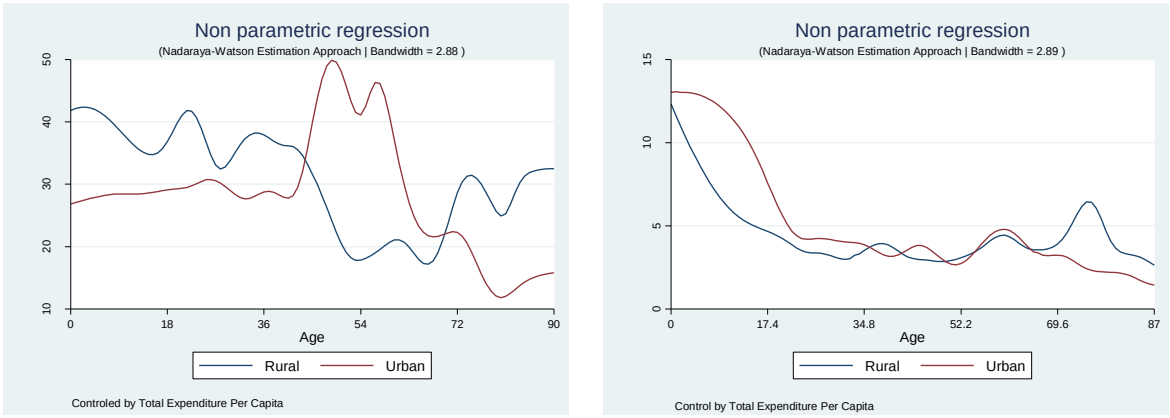


Wave 2

Figure(s) 2 Self reported cigarette consumption, MxFLS 1 & 2.



Figure(s) 3 Money spent and actual consumption of cigarettes, MxFLS 1 & 2.



Figure(s) 4 Calculated unit prices by age and stratum.

Table 1. Results of the (modified) Park Test, MxFLS 1 & 2.				
	Wave 1		**Wave 2**	
VARIABLES	Coefficient	SE	Coefficient	SE
xbeta1	1.662***	(0.151)	1.175***	(0.254)
Constant	1.428**	(0.567)	3.406***	(0.998)
Observations	2335		1967	
Log-likelihood	-19620		-17090	
	Wald tests of linear hypotheses			
	Statistic	Prob > Chi2	Statistic	Prob > Chi2
xbeta1 - 0 = 0	120.60	0.0000	21.49	0.0000
xbeta1 - 1 = 0	19.15	0.0000	0.48	0.4890
xbeta1 - 2 = 0	4.97	0.0257	10.58	0.0011
xbeta1 - 3 = 0	78.09	0.0000	51.79	0.0000

***NOTES**:* Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2. Cigarette consumption data: Estimates from Two-Part and Selection Models.

	Two-Part				Selection Two-Part (Heckit)				Selection MLE±			
	(1)		(2)		(1)		(2)		(1)		(2)	
VARIABLES	d	se	Ln(y)	se	d	se	Ln(y)	se	d	se	Ln(y)	se
Age	0.0304	(0.00499)***	0.0610	(0.00931)***	0.0304	(0.00499)***	0.0805	(0.0188)***	0.0304	(0.00499)***	0.0649	(0.0106)***
Age^2	-0.000367	(5.43e-05)***	-0.000516	(0.000104)***	-0.000367	(5.43e-05)***	-0.000752	(0.000223)***	-0.000366	(5.43e-05)***	-0.000563	(0.000121)***
Income(Ln)	-0.479	(0.0801)***	-0.411	(0.149)***	-0.479	(0.0801)***	-0.716	(0.297)**	-0.480	(0.0801)***	-0.471	(0.169)***
Income^2(Ln)	0.0961	(0.0190)***	0.0804	(0.0361)**	0.0961	(0.0190)***	0.142	(0.0632)**	0.0963	(0.0190)***	0.0924	(0.0393)**
Income^3(Ln)	-0.00470	(0.00122)***	-0.00381	(0.00233)	-0.00470	(0.00122)***	-0.00681	(0.00348)*	-0.00472	(0.00122)***	-0.00440	(0.00245)*
Male	0.645	(0.0376)***	0.457	(0.0715)***	0.645	(0.0376)***	0.874	(0.352)**	0.645	(0.0376)***	0.540	(0.131)***
Indigenous	-0.108	(0.0438)**	-0.398	(0.0835)***	-0.108	(0.0438)**	-0.466	(0.104)***	-0.108	(0.0438)**	-0.411	(0.0849)***
Urban	0.114	(0.0314)***	0.0739	(0.0584)	0.114	(0.0314)***	0.145	(0.0849)*	0.113	(0.0314)***	0.0880	(0.0612)
Married	0.135	(0.0422)***	-0.0682	(0.0730)	0.135	(0.0422)***	0.0157	(0.104)	0.136	(0.0422)***	-0.0517	(0.0759)
Divorced	0.478	(0.0686)***	0.0116	(0.116)	0.478	(0.0686)***	0.313	(0.278)	0.479	(0.0687)***	0.0710	(0.140)
Widowed	0.414	(0.0850)***	-0.0702	(0.162)	0.414	(0.0850)***	0.191	(0.274)	0.413	(0.0850)***	-0.0182	(0.176)
Primary	-0.169	(0.0552)***	-0.103	(0.103)	-0.169	(0.0552)***	-0.212	(0.141)	-0.169	(0.0552)***	-0.124	(0.106)
Secondary	-0.0826	(0.0632)	-0.0852	(0.116)	-0.0826	(0.0632)	-0.141	(0.131)	-0.0833	(0.0632)	-0.0961	(0.117)
High School	-0.242	(0.0711)***	-0.166	(0.130)	-0.242	(0.0711)***	-0.320	(0.186)*	-0.242	(0.0711)***	-0.197	(0.136)
University	-0.304	(0.0738)***	-0.0690	(0.136)	-0.304	(0.0738)***	-0.265	(0.216)	-0.304	(0.0738)***	-0.108	(0.145)
South-East	-0.237	(0.0470)***	-0.117	(0.0895)	-0.237	(0.0470)***	-0.270	(0.157)*	-0.237	(0.0470)***	-0.148	(0.0979)
West	-0.0321	(0.0430)	0.311	(0.0779)***	-0.0321	(0.0430)	0.290	(0.0839)***	-0.0317	(0.0430)	0.306	(0.0778)***
North-West	-0.129	(0.0418)***	0.653	(0.0741)***	-0.129	(0.0418)***	0.571	(0.103)***	-0.129	(0.0418)***	0.637	(0.0769)***
North-East	0.00543	(0.0427)	0.447	(0.0752)***	0.00543	(0.0427)	0.447	(0.0796)***	0.00559	(0.0427)	0.447	(0.0750)***
House maker	-0.164	(0.0448)***	-0.269	(0.0904)***	-0.164	(0.0448)***	-0.380	(0.131)***	-0.164	(0.0448)***	-0.291	(0.0947)***
Homeowner	-0.173	(0.0308)***	-0.117	(0.0547)**	-0.173	(0.0308)***	-0.226	(0.107)**	-0.173	(0.0308)***	-0.139	(0.0614)**
Health insurance	-0.0367	(0.0293)	-0.105	(0.0522)**	-0.0367	(0.0293)	-0.128	(0.0583)**	-0.0369	(0.0293)	-0.110	(0.0523)**

Continues in the next page…

Alcohol	**0.463**	**(0.0289)*** **	**-0.00352**	**(0.0534)**	**0.463**	**(0.0289)*** **	**0.295**	**(0.252)**	**0.463**	**(0.0289)*** **	**0.0554**	**(0.0944)**
Hot drinks	-0.651	(0.346)*	-0.726	(0.832)	-0.651	(0.346)*	-1.171	(0.915)	-0.649	(0.346)*	-0.815	(0.837)
Exercise	-0.0112	(0.0337)	-0.303	(0.0596)***	-0.0112	(0.0337)	-0.310	(0.0632)***	-0.0115	(0.0337)	-0.304	(0.0594)***
Other smoker	1.136	(0.0280)***	-0.0219	(0.0538)	1.136	(0.0280)***	0.726	(0.618)	1.136	(0.0280)***	0.126	(0.203)
Attritors	0.0951	(0.0308)***	-0.0318	(0.0541)	0.0951	(0.0308)***	0.0252	(0.0741)	0.0946	(0.0308)***	-0.0204	(0.0560)
Constant	-2.282	(0.154)***	1.600	(0.288)***	-2.282	(0.154)***	-0.573	(1.811)	-2.282	(0.154)***	1.169	(0.638)*
$\boldsymbol{\rho}$			0.000		0.628						0.1461	(0.1904)
$\boldsymbol{\sigma_2}$					1.376						1.1706	(0.0298)
$\boldsymbol{\sigma_{12} = \rho\sigma_2}$			0.000		0.864	(0.7094412)					0.1710	(0.2264)
Observations	19804		2335						19804			
Log-likelihood	-5393		-3663						-9055			
R-squared			0.168									

NOTES:
Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[±]LR test of independence Equations ($\boldsymbol{\rho = 0}$): Chi2(1) = 0.38 Prob > Chi2 = 0.5351

Table 3. Regression Estimates for the 2PM, MxFLS 1 & 2.					
	Variable	**Wave 1**		**Wave 2**	
		d	y1	d	y2
	Age	0.0304***	0.0657***	0.0242***	0.0442***
		(0.00499)	(0.00838)	(0.00515)	(0.0104)
	Age2	-0.000367***	-0.000568***	-0.000292***	-0.000309***
		(5.43e-05)	(9.22e-05)	(5.51e-05)	(0.000116)
	Income (ln)	-0.479***	-0.172***	-0.491***	-0.396**
		(0.0782)	(0.0510)	(0.101)	(0.179)
	Income^2 (ln)	0.0961***	0.0217***	0.0999***	0.101**
		(0.0183)	(0.00506)	(0.0257)	(0.0459)
	Income^3 (ln)	-0.00470***		-0.00490***	-0.00667**
		(0.00116)		(0.00179)	(0.00315)
	Male	0.645***	0.390***	0.561***	0.254***
		(0.0388)	(0.0634)	(0.0397)	(0.0940)
	Indigenous	-0.108**	-0.448***	-0.0724	-0.319***
		(0.0428)	(0.0801)	(0.0456)	(0.102)
	Urban	0.114***	0.0654	0.146***	-0.0404
		(0.0308)	(0.0567)	(0.0326)	(0.0679)
Marital state	Married	0.135***	-0.0213	0.0990**	-0.0383
		(0.0432)	(0.0714)	(0.0436)	(0.0795)
	Divorced	0.478***	0.126	0.289***	0.0135
		(0.0699)	(0.113)	(0.0751)	(0.119)
	Widowed	0.414***	-0.0703	0.311***	-0.106
		(0.0871)	(0.143)	(0.0880)	(0.161)
Education	Primary	-0.169***	-0.0805	-0.160**	0.0801
		(0.0551)	(0.0904)	(0.0671)	(0.112)
	Secondary	-0.0826	-0.0362	-0.162**	0.167
		(0.0629)	(0.104)	(0.0748)	(0.123)
	High-School	-0.242***	-0.139	-0.196**	0.324**
		(0.0706)	(0.118)	(0.0802)	(0.140)
	University	-0.304***	-0.0866	-0.336***	0.264*
		(0.0739)	(0.116)	(0.0831)	(0.137)
Region	South - East	-0.237***	0.00856	-0.349***	-0.0983
		(0.0454)	(0.0953)	(0.0499)	(0.147)
	West	-0.0321	0.259***	-0.0469	0.164*
		(0.0428)	(0.0693)	(0.0426)	(0.0866)
	North - West	-0.129***	0.572***	-0.144***	0.609***
		(0.0418)	(0.0685)	(0.0443)	(0.0875)
	North - East	0.00543	0.381***	0.00270	0.219**
		(0.0425)	(0.0740)	(0.0457)	(0.0929)

Continue in the next page…

Variable	Wave 1		Wave 2	
	d	y1	d	y2
House maker	-0.164***	-0.167*	-0.250***	-0.0718
	(0.0455)	(0.0878)	(0.0470)	(0.122)
Homeowner	-0.173***	-0.160***	-0.181***	-0.0478
	(0.0307)	(0.0546)	(0.0354)	(0.0683)
Health insurance	-0.0367	-0.0647	-0.0167	-0.106*
	(0.0294)	(0.0523)	(0.0314)	(0.0593)
Alcohol	0.463***	-0.0164	0.522***	-0.0493
	(0.0299)	(0.0511)	(0.0312)	(0.0642)
Hot drinks	-0.651*	0.00313	-0.381	1.131***
	(0.368)	(0.700)	(0.328)	(0.111)
Exercise	-0.0112	-0.257***	0.0882**	-0.220***
	(0.0342)	(0.0588)	(0.0390)	(0.0791)
Other smoker	1.136***	0.0576	1.018***	0.0113
	(0.0284)	(0.0492)	(0.0296)	(0.0628)
Attritors	0.0951***	-0.0411		
	(0.0311)	(0.0531)		
New comers			-0.0321	-0.00494
			(0.0354)	(0.0631)
Constant	-2.282***	1.925***	-2.175***	2.303***
	(0.153)	(0.257)	(0.176)	(0.317)
Observations	19804	2335	20606	1967
Pseudo R-squared	0.249		0.242	
Log-likelihood	-5393	-10726	-4918	-9133

NOTES: Robust standard errors in parentheses / *** p<0.01, ** p<0.05, * p<0.1

d : Probit model of smoking participation

y1 : GLM model with Logarithm-Link & Gamma distribution

y2 : GLM model with Logarithm-Link & Poisson distribution

Table 4. Marginal Effects, All individuals, MxFLS 1 & 2.

	Wave 1			Wave 2		
Variable	**Probability**	**Conditional Level**	**Unconditional Level**	**Probability**	**Conditional Level**	**Unconditional Level**
Male	0.0851***	12.98***	3.899***	0.0578***	9.132***	2.628***
	(0.00555)	(1.966)	(0.234)	(0.00453)	(3.181)	(0.240)
Indigenous	-0.0125***	-13.78***	-1.309***	-0.00661*	-10.82***	-0.743***
	(0.00468)	(2.090)	(0.195)	(0.00400)	(3.076)	(0.208)
Urban	0.0137***	2.355	0.645***	0.0136***	-1.564	0.450***
	(0.00369)	(2.027)	(0.185)	(0.00298)	(2.651)	(0.171)
Married	0.0162***	-0.778	0.540**	0.00927**	-1.475	0.289
	(0.00505)	(2.614)	(0.248)	(0.00401)	(3.081)	(0.211)
Divorced	0.0803***	4.828	3.219***	0.0342***	0.519	1.335***
	(0.0151)	(4.583)	(0.562)	(0.0108)	(4.600)	(0.427)
Widowed	0.0668***	-2.475	2.274***	0.0373***	-3.867	1.256**
	(0.0177)	(4.881)	(0.701)	(0.0129)	(5.595)	(0.530)
Primary	-0.0202***	-2.901	-0.915***	-0.0148**	3.111	-0.427
	(0.00646)	(3.225)	(0.287)	(0.00604)	(4.396)	(0.312)
Secondary	-0.00981	-1.307	-0.438	-0.0145**	6.645	-0.257
	(0.00725)	(3.725)	(0.354)	(0.00629)	(5.054)	(0.361)
High-School	-0.0259***	-4.803	-1.239***	-0.0167***	13.86**	-0.0140
	(0.00656)	(3.896)	(0.308)	(0.00609)	(6.662)	(0.385)
University	-0.0308***	-3.044	-1.307***	-0.0257***	11.17*	-0.479
	(0.00606)	(3.951)	(0.337)	(0.00499)	(6.433)	(0.333)
South - East	-0.0261***	0.312	-0.928***	-0.0280***	-3.628	-1.239***
	(0.00447)	(3.488)	(0.261)	(0.00338)	(5.196)	(0.261)
West	-0.00387	10.21***	0.493**	-0.00436	6.544*	0.129
	(0.00508)	(2.951)	(0.242)	(0.00387)	(3.588)	(0.221)
North - West	-0.0149***	24.34***	0.971***	-0.0128***	28.15***	0.782***
	(0.00455)	(3.413)	(0.249)	(0.00367)	(4.729)	(0.251)
North - East	0.000666	15.48***	0.985***	0.000257	8.971**	0.416
	(0.00523)	(3.394)	(0.310)	(0.00436)	(4.017)	(0.256)
House maker	-0.0192***	-5.722**	-1.053***	-0.0224***	-2.681	-0.981***
	(0.00509)	(2.820)	(0.248)	(0.00395)	(4.437)	(0.232)
Homeowner	-0.0225***	-6.014***	-1.192***	-0.0189***	-1.858	-0.808***
	(0.00425)	(2.137)	(0.217)	(0.00405)	(2.696)	(0.190)
Health insurance	-0.00448	-2.351	-0.309*	-0.00158	-4.062*	-0.244
	(0.00358)	(1.894)	(0.174)	(0.00297)	(2.255)	(0.150)
Alcohol	0.0622***	-0.598	2.226***	0.0583***	-1.909	2.148***
	(0.00445)	(1.869)	(0.189)	(0.00412)	(2.506)	(0.191)
Hot drinks	-0.0479***	0.114	-1.736	-0.0262*	80.39***	2.630**
	(0.0136)	(25.52)	(1.783)	(0.0153)	(13.14)	(1.064)
Exercise	-0.00136	-8.731***	-0.591***	0.00887**	-7.871***	-0.0160
	(0.00414)	(1.874)	(0.196)	(0.00412)	(2.648)	(0.214)
Other smoker	0.201***	2.073	7.445***	0.134***	0.433	5.166***
	(0.00599)	(1.755)	(0.285)	(0.00471)	(2.394)	(0.263)

NOTES: Robust standard errors in parentheses / *** p<0.01, ** p<0.05, * p<0.1
Standard errors obtained by the method of bootstrapped after 100 replications.

Table 5. Marginal Effects, Only Women, MxFLS 1 & 2.

	Wave 1			Wave 2		
Variable	**Probability**	**Conditional Level**	**Unconditional Level**	**Probability**	**Conditional Level**	**Unconditional Level**
Indigenous	-0.00608***	-10.44***	-0.444***	-0.00307*	-8.859***	-0.259***
	(0.00227)	(1.620)	(0.0861)	(0.00184)	(2.546)	(0.0772)
Urban	0.00677***	1.811	0.238***	0.00633***	-1.306	0.178**
	(0.00183)	(1.543)	(0.0669)	(0.00141)	(2.235)	(0.0718)
Married	0.00799***	-0.603	0.210**	0.00434**	-1.222	0.116
	(0.00250)	(2.022)	(0.0985)	(0.00188)	(2.544)	(0.0760)
Divorced	0.0434***	3.722	1.321***	0.0168***	0.431	0.544***
	(0.00882)	(3.524)	(0.280)	(0.00555)	(3.817)	(0.193)
Widowed	0.0355***	-1.924	0.952***	0.0184***	-3.223	0.527**
	(0.0101)	(3.802)	(0.315)	(0.00666)	(4.690)	(0.221)
Primary	-0.0100***	-2.239	-0.341***	-0.00691**	2.607	-0.173
	(0.00323)	(2.485)	(0.113)	(0.00284)	(3.715)	(0.117)
Secondary	-0.00482	-1.016	-0.163	-0.00672**	5.495	-0.115
	(0.00354)	(2.901)	(0.134)	(0.00289)	(4.160)	(0.108)
High-School	-0.0123***	-3.728	-0.445***	-0.00761***	11.49**	-0.0342
	(0.00303)	(3.021)	(0.117)	(0.00272)	(5.566)	(0.129)
University	-0.0145***	-2.367	-0.470***	-0.0114***	9.146*	-0.198*
	(0.00273)	(3.079)	(0.115)	(0.00215)	(5.224)	(0.117)
South - East	-0.0126***	0.242	-0.349***	-0.0127***	-3.009	-0.459***
	(0.00217)	(2.708)	(0.0823)	(0.00159)	(4.290)	(0.0822)
West	-0.00191	7.862***	0.151*	-0.00203	5.434*	0.0338
	(0.00250)	(2.289)	(0.0882)	(0.00180)	(2.961)	(0.0824)
North - West	-0.00725***	18.86***	0.288***	-0.00589***	23.70***	0.242**
	(0.00219)	(2.780)	(0.0844)	(0.00168)	(4.121)	(0.0963)
North - East	0.000330	11.95***	0.321***	0.000120	7.407**	0.138
	(0.00259)	(2.642)	(0.0945)	(0.00204)	(3.272)	(0.0906)
House maker	-0.0102***	-4.687*	-0.409***	-0.0119***	-2.281	-0.419***
	(0.00290)	(2.422)	(0.0968)	(0.00237)	(3.865)	(0.113)
Homeowner	-0.0113***	-4.665***	-0.442***	-0.00905***	-1.545	-0.317***
	(0.00222)	(1.672)	(0.0755)	(0.00202)	(2.241)	(0.0800)
Health insurance	-0.00222	-1.832	-0.110	-0.000740	-3.411*	-0.0855
	(0.00177)	(1.490)	(0.0710)	(0.00139)	(1.919)	(0.0583)
Alcohol	0.0364***	-0.462	1.014***	0.0336***	-1.575	1.043***
	(0.00326)	(1.443)	(0.101)	(0.00309)	(2.039)	(0.127)
Hot drinks	-0.0214***	0.0884	-0.603	-0.0115*	66.91***	0.848**
	(0.00531)	(19.81)	(0.543)	(0.00622)	(12.13)	(0.342)
Exercise	-0.000671	-6.748***	-0.195***	0.00423**	-6.542***	0.0161
	(0.00204)	(1.470)	(0.0709)	(0.00201)	(2.295)	(0.0683)
Other smoker	0.121***	1.605	3.444***	0.0742***	0.359	2.371***
	(0.00495)	(1.362)	(0.215)	(0.00366)	(1.989)	(0.176)

NOTES: Robust standard errors in parentheses / *** p<0.01, ** p<0.05, * p<0.1
Standard errors obtained by the method of bootstrapped after 100 replications.

Table 6. Marginal Effects, Only Men, MxFLS 1 & 2.

	Wave 1			Wave 2		
Variable	**Probability**	**Conditional Level**	**Unconditional Level**	**Probability**	**Conditional Level**	**Unconditional Level**
Indigenous	-0.0243***	-15.18***	-3.255***	-0.0140	-11.53***	-1.936***
	(0.00927)	(2.321)	(0.564)	(0.00854)	(3.291)	(0.494)
Urban	0.0264***	2.580	1.436***	0.0285***	-1.656	0.962**
	(0.00710)	(2.229)	(0.461)	(0.00627)	(2.800)	(0.411)
Married	0.0311***	-0.850	1.108*	0.0194**	-1.567	0.603
	(0.00978)	(2.858)	(0.606)	(0.00842)	(3.277)	(0.476)
Divorced	0.137***	5.286	6.249***	0.0668***	0.551	2.782***
	(0.0235)	(5.024)	(1.214)	(0.0198)	(4.883)	(0.874)
Widowed	0.116***	-2.702	4.209***	0.0727***	-4.099	2.470**
	(0.0283)	(5.326)	(1.495)	(0.0236)	(5.922)	(1.192)
Primary	-0.0387***	-3.174	-2.018**	-0.0310**	3.292	-0.871
	(0.0124)	(3.532)	(0.842)	(0.0127)	(4.641)	(0.872)
Secondary	-0.0189	-1.427	-0.968	-0.0308**	7.061	-0.416
	(0.0141)	(4.064)	(0.910)	(0.0136)	(5.384)	(0.931)
High-School	-0.0518***	-5.246	-2.849***	-0.0360***	14.71**	0.280
	(0.0138)	(4.260)	(0.931)	(0.0136)	(7.075)	(0.983)
University	-0.0626***	-3.322	-2.990***	-0.0572***	11.91*	-0.912
	(0.0133)	(4.310)	(0.804)	(0.0120)	(6.889)	(0.915)
South - East	-0.0515***	0.341	-1.993***	-0.0612***	-3.852	-2.946***
	(0.00910)	(3.810)	(0.697)	(0.00767)	(5.526)	(0.799)
West	-0.00744	11.18***	1.392**	-0.00915	6.945*	0.452
	(0.00981)	(3.238)	(0.640)	(0.00818)	(3.822)	(0.580)
North - West	-0.0290***	26.61***	2.863***	-0.0273***	29.74***	2.420***
	(0.00908)	(3.732)	(0.697)	(0.00800)	(5.041)	(0.660)
North - East	0.00127	16.94***	2.606***	0.000536	9.538**	1.154*
	(0.00999)	(3.732)	(0.724)	(0.00909)	(4.302)	(0.613)
House maker	-0.0352***	-6.133**	-2.324***	-0.0429***	-2.823	-2.079***
	(0.00896)	(2.976)	(0.594)	(0.00692)	(4.638)	(0.641)
Homeowner	-0.0422***	-6.570***	-2.667***	-0.0382***	-1.971	-1.788***
	(0.00779)	(2.335)	(0.438)	(0.00796)	(2.861)	(0.506)
Health insurance	-0.00858	-2.565	-0.728*	-0.00330	-4.296*	-0.644*
	(0.00686)	(2.061)	(0.388)	(0.00621)	(2.379)	(0.383)
Alcohol	0.105***	-0.653	4.058***	0.102***	-2.030	3.901***
	(0.00647)	(2.045)	(0.347)	(0.00595)	(2.677)	(0.479)
Hot drinks	-0.105***	0.125	-4.152	-0.0597	85.25***	7.690***
	(0.0358)	(27.88)	(4.600)	(0.0387)	(13.72)	(2.411)
Exercise	-0.00260	-9.550***	-1.545***	0.0181**	-8.352***	-0.254
	(0.00795)	(2.056)	(0.449)	(0.00827)	(2.785)	(0.493)
Other smoker	0.317***	2.266	12.94***	0.236***	0.459	9.674***
	(0.00880)	(1.918)	(0.591)	(0.00777)	(2.541)	(0.540)

NOTES: Robust standard errors in parentheses / *** p<0.01, ** p<0.05, * p<0.1
Standard errors obtained by the method of bootstrapped after 100 replications.

Chapter III

A panel double-hurdle approach for modelling cigarette demand in Mexico: Evidence from the Mexican Family Life Survey.

Introduction

Many economic studies have shown the usefulness of the Cragg's model, due to Cragg (1971), for modelling tobacco consumption particularly when the unit of analysis are households or individuals. This model, usually known in the literature as the "Double-Hurdle" model (DH hereafter), not only successfully addresses the censoring of information likely to be encountered but it has been found to be a less restrictive estimator than its alternative, the Tobit model. Since its appearance in the literature, the DH has become the standard econometric methodology in studies of similar nature. It allows investigating factors explaining the individual's decision to "participate" in the market (first hurdle) and in which "intensity" or level of consumption (second hurdle). Applications of this model from the tobacco demand literature include Atkinson, Gomulka et al. (1984), Jones (1989) and García and Labeaga (1996).

However, one important characteristic of related studies is that they are usually restricted to cross-sectional analysis. This has the implication that their results are constrained in what they can reveal regarding the dynamics of consumption. This is an important limitation. Labeaga (1999) notes that 'as regards to the individuals' decisions, observing a situation in the past can increase the probability that this situation be repeated in the future' [p.52]. Thus, in the context of smoking behaviour, there is a justification for tobacco consumption to be studied in a life-cycle context, as cited, for example in Labeaga (1993) and Jones and Labeaga (2003).

There are, at least, two reasons for related studies being constrained to a cross-sectional analysis. On the one hand, the absence of panel data has made it impossible for researchers to analyse smoking dynamics which, given the addictive nature of cigarettes, may better explain tobacco consumption. On the other hand, even if panel data is available (which has been mainly sourced from developed countries), the original structure of the DH model was not designed for:

1) Control for individual heterogeneity in preferences.
2) To account for the temporal linkage of consumption (dependence).

In consequence, only static models have been possible to use (Jones and Labeaga 2003). However, given recent advances in econometrics, the "DH framework" has been now extended into a "panel DH framework". This has opened up the possibility to model tobacco consumption using longitudinal datasets which have become more common in recent years (even from developing countries). This means that, unlike the usual static model, the panel framework helps to account for heterogeneity across individuals and dynamic effects. The model to be presented is applied to the consumption of cigarettes in Mexico.

The econometric model estimated in this analysis is an extension of the panel model introduced by Dong and Kaiser (2008). Such a model has two important features:

1) The model assumes a non-zero correlation between the individual-specific error term in the smoking participation decision and the intensity of tobacco consumption.
2) The model integrates the Box-Cox transformation of the dependent variable similar to Jones and Yen (2000) and Moffatt (2005). This has proved to be an effective approach for achieving consistent estimates in the presence of non-normal errors which may be traced from the skewed distribution seen on the data. A skewed distribution is likely to be present in studies of this kind. Evidence suggests that the assumption of normality imposed to the unobserved error term appears to be too strong to work in practice. As it is known in the discipline of econometrics, Maximum Likelihood (ML) estimates may be inconsistent if a departure from normality is encountered.

The source of data use for this analysis also comes from the Mexican Family Life Survey (MxFLS) of 2002-2007. With this data, it has been possible to construct a short panel of individuals. The sample size is 15,036 individuals all of whom were observed in two sweeps. This study restricts the analysis to a balanced panel.

Mexico makes for an interesting case study. Since the year 2002, a series of tobacco-control policies have been enacted in accordance with WHO Framework Convention on Tobacco Control (WHO FCTC). It is expected then that such policies may have started to change the behaviour of smokers and potential smokers. Evidence has been provided in the previous chapter of systematic differences of factors determining the probability of an individual to become a smoker, and factors determining the level of consumption between the period of a pre-reform and post-reform. The results in this Chapter will show that, by controlling for unobserved individual heterogeneity, the prevalence of smoking is explained by both economic and non-economic factors, and the prospects of a potential smoker for heavy smoking are lower in the post-reform period as smokers are expected to smoke less over time.

The remainder of this chapter is structured as follows. In Section 2 a brief background of the model to be estimated is provided. In Section 3, the panel hurdle model is presented and Section 4 data which is used for estimation is described. In Section 5, the results are presented and a discussion of the findings is given in Section 6.

Background

Since the recognition of the health risks associated with cigarette smoking, a body of economics research has appeared in the literature to understand how the demand for tobacco products is determined. In particular, the availability of individual data has provided more insights about smoking behaviour which otherwise is not possible to capture with, say, aggregate or budgetary data (Chaloupka and Warner 2000). Although most of the research has been concerned with investigating the role of "price" in determining cigarette consumption, a growing body of empirical literature has turned its attention in investigating the role of detailed socio-demographic characteristics explaining smoking. There is, however, an important methodological regularity found across studies of tobacco demand which is the reliance on the DH framework to model cigarette consumption. This framework, originally due to Cragg (1971), successfully addresses the problem of censoring. This is a common feature of studies of this kind in which a large cluster of zeroes is encountered, given that tobacco is not consumed by the majority of households or individuals. This is also true for closely related goods such as alcohol or durables. In the previous chapter, some considerations were outlined regarding possible explanations of such characteristics as seen in data. Nevertheless, it should be noted that this framework has been found to be more appropriate when preferences explaining the censoring of information may be due to "corner solutions" (economic reason) or may be due to "abstention" (non-economic reason) (Smith 2002).

One of the well known benefits of using the DH is that it decomposes the mechanism that explains consumption into two different processes: participation and intensity thereby the two hurdles. Moreover, empirical evidence has shown that these two processes may be determined by different factors. This makes the DH superior than its alternative, the Tobit model. There has been, however, one important limitation regarding this model which is the inability to account for individual heterogeneity in preferences (control for unobservable characteristics). The implication is that most studies have been constrained to cross-sectional analysis. This means that it has been difficult to account for any life-cycle considerations of individual behaviour. This is unfortunate given the availability of longitudinal data which has opened up the possibility of investigating some dynamics in consumption which results from purchase carryover, learning behaviour and other factors, and of course, heterogeneity in preferences (Dong, Chung et al. 2004). Thus, the need for the DH to account for the structure of the data, particularly panel structures, has been paramount. Notwithstanding, there are popular empirical strategies available to overcome this limitation. For example, it could be possible to apply the Correlated Random Effects (CRE) framework due to Mundlak (1978) and Chamberlain (1984). Although consistent, it only

constitutes an ad-hoc solution to the problem at hand which may involved a series of computation procedures in order to retrieve the required parameters. There is a vast literature on this topic including Wooldridge (2002) and Cameron and Trivedi (2005).

A recent development on DH models for panel data has been provided by Dong, Chung et al. (2004) and Dong and Kaiser (2008). The framework, which is simply an extension of the Cragg's model to the panel data context, has been applied to model household milk consumption in order to study the phenomenon of short-run consumption, commonly known as "infrequency of purchases". It should be noted, however, that a modified framework of the Cragg's model to accommodate short-run consumption has been previously introduced by Deaton and Irish (1984) with the so-called "P-Tobit" model, albeit it restricts its analysis to a cross-sectional data only.

The first application of the P-Tobit model suggested that the consumption of durables, alcohol and tobacco may be explained by different types of misreporting such as false reporting and/or infrequency of purchases. This derived from, perhaps, problems in the survey's design. In the literature of tobacco demand such issues have been rarely studied. However, it is worth mentioning an early study by Warner (1978) in which he found that data release by the U.S. Department of Health, Education and Welfare (USDHEW) for the years 1972 and 1976 on self-reported cigarette consumption, showed signs of considerable discrepancies with data on tobacco consumption release by the U.S Department of Agriculture. His work suggested that the Surgeon General's Report on Smoking and Health of 1964 may have caused possible increases in underreporting of cigarette consumption in later years. This may be due to false reporting after the release of such report. Although he offers alternative explanations for explaining the observed discrepancies such as changes in the survey's design of the U.S. DHEW, he mainly attributed it to changes in knowledge of and changing attitudes towards smoking. Warner's study may have motivated early suggestions that data of consumption on tobacco collected by surveys may have been affected by such issues. Notwithstanding, misreporting of tobacco consumption has not been a concurrent issue. It would appear then, and as noted in the previous chapter, that tobacco consumption is reported accurately enough. This suggests that the observed zero seen in data is due to abstention. In fact, in a relatively recent study, Miles (2000) found that in the case of the Spanish Expenditure Survey of 1990-91, the probability that a smoking household does not purchase tobacco during the period of the survey is extremely small although his study is more concerned with "stockpiling behaviour" rather than false reporting.

Nonetheless, both Dong, Chung et al. (2004) and Dong and Kaiser (2008) noted that their proposed framework is equally relevant when there is sufficient evidence to suggest that the observed zero may be due to "abstention" or "corner solution". For this reason, the aim of this chapter is to model cigarette consumption under a similar framework. It is expected that such

framework may contribute to increase the current understanding of how the demand for tobacco is determined in a life cycle context.

The econometric model introduced in this analysis, which is labelled "Panel Double-Hurdle model" (PDH hereafter) is well suited to identify the so-called "zero-type" individuals or "non-participants" as they will be referred throughout on this chapter. "Zero-types" are individuals who, for individual-specific rather than circumstantial reasons, would *never* smoke. The model, to be introduced in the next sub-section, has been extended in two important ways. Firstly, it has the advantage of taking into account the temporal linkage amongst individuals by introducing "dependence". In other words, the model assumes a non-zero correlation between the individual-specific error term in the smoking participation decision and the intensity or level of tobacco consumption. Dependence is an important element of the PDH as it recognises the individual's smoking status cycle which may be explained by individual heterogeneity in preferences. Not accounting for dependence would represent a source of inconsistency of the estimates as it would be the case if panel data is used within a cross-sectional framework. However, the extension of the panel data requires care, because the outcome of the smoking participation equation (first hurdle), that is, the determination of whether an individual is of the zero-type, must apply for every time period. Switching in and out of the zero-type is ruled out. In contrast, the outcome of the smoking intensity equation (second hurdle), that is, the amount actually consumed in any time period, is determined at the level of individual observations. Thus, individuals classified as "non-participants" must consume zero in every time period.

Secondly, in order to address the skewed distribution seen in data, the dependent variable is transformed using the Box-Cox transformation. Such transformation has been considered in the context of the cross-sectional DH by Jones and Yen (2000)[1] and by Moffatt (2005). The model introduced by Jones and Yen (2000), which is labelled "Box-Cox Double-Hurdle" (BCDH hereafter) is a generalisation of the DH which is fully parametric and nests a wide range of limited dependent models (models for censored data) such as the usual DH (with no transformation), Heckman model with lognormal and the Two-Part model (previously estimated). In turn, Moffatt (2005) introduces the BCDH with dependence and the BCDH for interval data. The Box-Cox transformation within censoring models has also been considered by Chaze (2005) who introduces his own generalisation of the BCDH distinguishable from the one presented by Jones and Yen (2000). In short, the main feature of the model is that it integrates the transformation of

[1] To my knowledge, the Box-Cox double hurdle model first appeared in the literature in an article by Yen (1993). Perhaps the generalisation of the model was officially established in a working paper published by the Institute of Fiscal Studies by Jones & Yen dated in 1994 which was finally published by a journal in 2000. However, at least five times the model has been used before the year of 2000 by Steven T. Yen. Nonetheless, the key citation for this model will be Jones and Yen (2000).

the dependable variable within the model, in order for the assumption of normality to hold, so that consistent maximum likelihood (ML) estimates can be obtained.

In the tobacco demand literature, the BCDH has been applied by Aristei and Pieroni (2008) which used it to model tobacco consumption expenditure using budgetary data from Italy. The BCDH has also been applied in other contexts of the economics of smoking literature, by Yen and Jones (1996) and in a follow-on study by Aristei and Pieroni (2009). In these two cases, the focus of interest is on investigating the influence of addiction on quitting behaviour rather than the impact of addiction on the level of consumption as in the traditional demand analysis. Nevertheless, all these studies have been of cross-sectional nature. Of course, the application of the BCDH has not been exclusive to smoking. Recent work includes the demand for food away from home by Bai, Wahl et al. (2010) and land conservation investments by Genanew and Alemu (2012); see Table I for further examples[2].

In addition, addressing the need of applying a suitable transformation of the dependent variable has produced a considerable amount of research. In particular, several versions of the DH model which aims to relax the bivariate normality assumption imposed to the error term have been previously proposed in the literature. One of such variant includes the use of the *inverse hyperbolic sine* where an example of this given by Yen and Huang (1996). Alternatively, the *copula method* has been suggested by Smith (2001). Nevertheless, the Box-Cox transformation has been widely accepted as the most efficient approach to address the long tails seen in data.

Econometric model

In this section three variants of the Cragg's model are presented to model cigarette consumption. The DH model is first introduced which gives the foundation for the BCDH and its panel estimator. The essence of the hurdle model is that it assumes that two hurdles need to be passed in order for a positive consumption to be observed. The first hurdle determines whether the respondent is a "participant"; if they fall at the first hurdle, they are a "non-participant", and their consumption will necessarily be zero. If the first hurdle is passed, a "second hurdle" is encountered, and whether this is crossed determines whether the consumption is zero or positive.

[2] Unfortunately the list is not exhaustive as some journals restrict access.

The double-hurdle model

The basic form of the hurdle model was introduced by Cragg (1971) and its statistical properties are well established in the literature; see Pudney (1989) and Smith (2002). This model has been used extensively in empirical work to model tobacco consumption. A representative example is given by García and Labeaga (1996) which defines the two hurdles and the way in which they interact in determining observed consumption as follows:

$$\begin{aligned} d_i^* &= \boldsymbol{x}_{1i}'\beta_1 + \varepsilon_{1i} \\ y_i^* &= \boldsymbol{x}_{2i}'\beta_2 + \varepsilon_{2i} \end{aligned} \tag{1}$$

where d_i^* denotes whether individual i is a smoker or reports positive consumption of tobacco (latent participation variable) therefore:

$$\begin{aligned} d_i &= 1 \text{ if } d_i^* > 0 \\ d_i &= 0 \text{ otherwise} \end{aligned} \tag{2}$$

y_i^* is the latent dependent variable such as that:

$$\begin{aligned} y_i &= y_i^* \text{ if } y_i^* > 0 \text{ and } d_i^* > 0 \\ y_i &= 0 \text{ otherwise} \end{aligned} \tag{3}$$

where $y_i$is the observe consumption of tobacco. $\boldsymbol{x}_i'$ is a vector of individuals' characteristics (i.e socio-economic and/or demographic), β_i is a vector of parameters to be estimated and ε_i is the error term.

As the model produces two error terms, these are assumed to follow a (bivariate) normal distribution with zero mean and constant variance such that:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right\} \tag{4}$$

The general likelihood function of the model may be written as (Yen 2005a):

$$\begin{aligned} L = \prod_{y=0} \{1 - \psi(x_1'\beta_1, x_2'\beta_2/\sigma; \rho)\} \\ \times \prod_{y>0} \left\{ \frac{1}{\sigma} \phi[(y - x_2'\beta_2)/\sigma] \Phi\left[\frac{x_1'\beta_1 + \rho\,(y - x_2'\beta_2)/\sigma}{(1-\rho^2)^{1/2}} \right] \right\} \end{aligned} \tag{5}$$

where $\psi(\cdot,\cdot;\cdot)$ is the bivariate standard normal cumulative distribution function, $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function (CDF) and $\phi(\cdot)$ is the standard normal probability distribution function (PDF). However, following the convention in the field the

assumption of independence will be imposed[3] which for estimations purposes this helps to simplify the log-likelihood function to:

$$\log L = \sum_{y=0} \ln\left[1 - \Phi(x_1'\beta_1)\Phi\left(\frac{x_2'\beta_2}{\sigma}\right)\right] + \sum_{y>0} \ln\left[\Phi(x_1'\beta_1)\frac{1}{\sigma}\phi\left(\frac{y - x_2'\beta_2}{\sigma}\right)\right] \qquad (6)$$

The box-cox double-hurdle model

The dependent variable of interest shows a strong positive skewed distribution. This is a problem because in this class of model, consistency in estimation depends crucially on the assumption of normality in the error terms. To address this issue, Jones and Yen (2000) has shown that by internalising the Box-Cox transformation of the dependent variable within the model it is possible to correct for non-normal errors. This transformation depends on estimating the optimal transformation parameter, λ, defined by Box and Cox (1964) as:

$$y^T = \{y^\lambda - 1\}/\lambda, \qquad 0 < \lambda < 1 \qquad (7)$$

Notice the restriction imposed to λ which is expected not to take extreme cases such as zero ($\lambda = 0$) or one ($\lambda = 1$). In the former case, it would take the transformation of y_i into logarithms which suggest that the model employed in Chapter 2 is sufficient for identifications of parameters. In contrast, the latter case would make the transformation unnecessary therefore the DH presented in (4) would be the ideal model to estimate. The intermediate case or $\lambda = 0.5$ would take the transformation of y into square roots. This suggests that it would be feasible to interpret the reported consumption of cigarettes as a discrete distribution which supports the idea of using count-data models (Yen and Jones 1996). Nevertheless it is expected that the parameter λ to lie somewhere between the limits of 0 and 1.

Given the Box-Cox transformation, the observed consumption is modified as:

$$\begin{aligned} y_i^T &= y_i^{*T} \quad \text{if} \quad d_i = 1 \\ y_i^T &= -\frac{1}{\lambda} \quad \text{if} \quad d_i = 0 \end{aligned} \qquad (8)$$

where d_i has been defined in (2). It should be noted that the lower limit of the transformed variable is now $-1/\lambda$ and not zero. The general likelihood function of the model is written as (Jones and Yen 2000):

[3] $\text{corr}(\varepsilon_2, \varepsilon_1) = \rho = 0$

$$L = \prod_{y=0}\left\{1 - \psi\left(x_1'\beta_1, \frac{x_2'\beta_2 + 1/\lambda}{\sigma}, \rho\right)\right\} \times \prod_{y>0}\left\{\frac{1}{\sigma}\phi[(y^T - x_2'\beta_2)/\sigma]\Phi\left[\frac{x_1'\beta_1 + (\rho/\sigma)(y^T - x_2'\beta_2)}{(1-\rho^2)^{1/2}}\right] y_i^{\lambda-1}\right\} \quad (9)$$

Similarly to the DH model presented above independence will also be assumed in order to ease the estimation of the parameters of interest. In such case, the log-likelihood function of (9) with independence is written as (Moffatt 2005):

$$\log L = \sum_{y=0} \ln\left[1 - \Phi(x_1'\beta)\Phi\left(\frac{x_2'\beta + 1/\lambda}{\sigma}\right)\right] + \sum_{y>0} \ln\left[\Phi(x_1'\beta) y_i^{\lambda-1}\frac{1}{\sigma}\phi\left(\frac{y_i^T - x_2'\beta}{\sigma}\right)\right] \quad (10)$$

The panel hurdle model: controlling for unobserved heterogeneity

The basic form of the panel hurdle model has been previously introduced by Dong, Chung et al. (2004) and Dong and Kaiser (2008). The central feature of the model is that the first hurdle has only one outcome per respondent, and that outcome is applied to all observation for that respondent. If respondent i falls at the first hurdle ($d_i = 0$), then all the observations on y for respondent i must be zero ($y_{it} = 0, t = 1, \dots, T$). The model defines the two hurdles and observed consumption as:

FIRST HURDLE

$$\begin{aligned} &d_i^* = z_i'\alpha + \varepsilon_{1i} \\ &d_1 = 1 \text{ if } d_i^* > 0;\ d_i = 0 \text{ otherwise} \\ &\varepsilon_{1i} \sim N(0,1) \end{aligned} \quad (11)$$

SECOND HURDLE

$$\begin{aligned} &y_{it}^{**} = x_{it}'\beta + u_i + \varepsilon_{2it} \\ &y_{it}^{*} = \max(y_{it}^{**}, 0) \\ &\varepsilon_{2it} \sim N(0, \sigma^2) \\ &u_i \sim N(0, \sigma_u^2) \end{aligned} \quad (12)$$

OBSERVED CONSUMPTION

$$y_{it} = d_i y_{it}^* \quad i = 1, \dots, n \quad t = 1, \dots, T \quad (13)$$

where y_{it} is the consumption of cigarettes of individual i in time period t. $\boldsymbol{z}_i$ is a vector of characteristics of individual i relevant in explaining whether the first hurdle is passed and $\boldsymbol{x}_{it}$ is a vector of characteristics of individual i and/or time period t that determine the consumption level in that time period. It should also be noted that there is a respondent-specific random effect term (u_i) appearing in the second hurdle that allows between-subject heterogeneity.

Similarly to the cross-sectional version of the DH, it is intended to apply the Box-Cox transformation to y_{it}^{**} which in such case, the second hurdle in (12) becomes:

$$\begin{aligned} y_{it}^{**T} &= \frac{y_{it}^{**\lambda} - 1}{\lambda} = x_{it}'\beta + u_i + \varepsilon_{2it} \\ y_{it}^{*T} &= \max\left(y_{it}^{**T}, -\frac{1}{\lambda}\right) \end{aligned} \tag{14}$$

And the rule for arriving at the observed value (y_i^T) becomes:

$$\begin{aligned} y_{it}^{T} &= y_{it}^{*T} \text{ if } d_i = 1 \\ y_{it}^{T} &= -\frac{1}{\lambda} \text{ if } d_i = 0 \end{aligned} \tag{15}$$

In contrast to models presented in (6) and (10) in which it is assumed that the error terms between the two hurdles are uncorrelated, the panel hurdle relaxes this assumption. This means that individual i's idiosyncratic propensity to pass the first hurdle is represented by the error term ε_{1i}; his/her idiosyncratic propensity to consume, conditional on passing the first hurdle, is represented by u_i. It is between these two terms that correlation is introduced:

$$\text{corr}(\varepsilon_1, u) = \text{P} \neq 0 \tag{16}$$

The correlation parameter P is incorporated in the estimation as follows. Given that $\text{corr}(\varepsilon_1, u) = \text{P}$, from (11), ε_1 may be represented as (supressing the i subscript):

$$\varepsilon_1 = \text{P}\frac{u}{\sigma_u} + \sqrt{1 - \text{P}^2}\,\xi \tag{17}$$

where $\xi \sim N(0,1)$ and $\xi \perp u$. The requirement for passing the first hurdle becomes:

$$d_i = 1 \text{ if } \xi > -\frac{z_i'\alpha + \text{P}\frac{u}{\sigma_u}}{\sqrt{1 - \text{P}^2}} \tag{18}$$

and derived from this the probability of passing the first hurdle (conditional on u) becomes:

$$\Phi\left(\frac{z_i'\alpha + \mathrm{P}\frac{u}{\sigma_u}}{\sqrt{1-\mathrm{P}^2}}\right) \tag{19}$$

The log-likelihood function may be written as:

$$\begin{aligned}\log L = \int_{-\infty}^{\infty} \Bigg\{ & \Phi\left(\frac{z_i'\alpha + \mathrm{P}\frac{u}{\sigma_u}}{\sqrt{1-\mathrm{P}^2}}\right)(L|d_i = 1, u) \\ & + \left[1 - \Phi\left(\frac{z_i'\alpha + \mathrm{P}\frac{u}{\sigma_u}}{\sqrt{1-\mathrm{P}^2}}\right)\right](L|d_i = 0)\Bigg\} f(u;\sigma_u)du\end{aligned} \tag{20}$$

where:

$$(L|d_i = 1, u) = \prod_{t=1}^{T}\left[1 - \Phi\left(\frac{x_{it}'\beta + (1/\lambda) + u_i}{\sigma}\right)\right]^{i\in I_0}\left[y_{it}^{\lambda-1}\frac{1}{\sigma}\phi\left(\frac{y_{it}^{T} - x_{it}'\beta - u_i}{\sigma}\right)\right]^{i\in I_1} \tag{21}$$

and

$$(L|d_i = 0) = \prod_{t=1}^{T}\left[1 - \Phi\left(\frac{x_{it}'\beta + u_i}{\sigma}\right)\right]^{i\in I_0}\left[\frac{1}{\sigma}\phi\left(\frac{y_{it} - x_{it}'\beta - u_i}{\sigma}\right)\right]^{i\in I_1} \tag{22}$$

Estimation

The estimation is performed using the method of Maximum Simulated Likelihood [MSL, (Train 2003)]. This requires the use of Halton Draws, which, when converted to normality, represents simulated realisation of the random effect term u. In the model with dependence, in accordance with (19), the simulated values also appear in the probability of passing the first hurdle. Maximisation of the simulated likelihood function is also performed using the ML routine in STATA. The programme required suitable starting values which are obtained using estimates from a pooled Box-Cox hurdle model.

The parameters of the BCDH [model in (10)] are obtained using the ML routine in STATA following the written programme by Moffatt (2005). For comparison purposes, estimates from the pool double hurdle model with no transformation (DH) [expression in (6)] are also presented. The DH has been estimated using the user-written command `dhurdle` by Fennema and Sinning (2007) assuming independence. Alternatively, estimates of the DH may be obtained by using the user-written command `craggit` by Burke (2009). However results from using `craggit` may differ somewhat given the structure of the log-likelihood function being assumed.

Data

The data set used for this analysis is from the Mexican Family Life Survey (MxFLS), a multi-purpose and longitudinal survey from a nationally representative sample of Mexican households and their members[4]. Particularities of the survey's design have been described in more detail in the previous chapter. For completeness relevant information is also provided here. The data used is a short panel of individuals obtained from two sweeps of the MxFLS. Data for the first sweep, which constitutes the baseline for the survey, was collected between March and July of 2002. The collection of information from the second sweep or "follow-up" was collected starting in the year 2005. The follow-up spanned the years 2005 and 2007 due to the re-contact process. The overall number of individuals belonging to all households surveyed during the first sweep is 35,677 and for the second sweep is 38,223. As the sample of interest corresponds only to adults of at least 15 years old of age who successfully completed the section "Tastes & Habits" of the questionnaire, the usable sample is 19,785 and 20,606 individuals in the first and second sweep respectively.

The rate of attrition relative to the first wave is 24.08% corresponding to 4,769 individuals. However, in order for the survey not to lose its representativeness, the survey's design "naturally" updated the sample by the second sweep. This involved adding individuals who by birth or by marriage are considered members of the home. The number of "new-comers" is 5,571 individuals or 27.03% of the sample relative to the second sweep. Thus, data available for analysis correspond to an unbalanced panel with a sample of 40,410 individuals of whom 15,036 are successfully followed-up. The inclusion of new-comers in the sample would compensate for the loss of information by those not followed-up otherwise results may be confounded by the presence of attrition bias [see Rubalcava and Teruel (2008) for details].

For all adult respondents, several questions regarding their smoking status and, if appropriate, their smoking history, were asked. Those who have the habit of smoking and have not quit smoking in a frequent way (therefore active smokers) were asked the question: 'Currently how many cigarettes do you smoke on average per week?' This helped to create the two independent variables needed for all variants of the Cragg's model estimated. For the smoking participation equation, a dichotomous variable equal to 1 if any positive consumption on cigarettes was reported and equal to 0 otherwise. In turn, for the smoking intensity equation the dependent variable correspond to the continuous measure of the number of cigarettes smoked on average per week. It is worth noting that in 49.1% of cases (in both sweeps), information on

[4] All data and appropriate documentation for understanding and handling the information for the MxFLS (manuals, questionnaires and codebooks), is in the public domain and it is freely available on the internet with a translation into English for access to a wider audience. For more information visit the web-page: http://www.ennvih-mxfls.org

the number of cigarettes smoked was reported as “number of packages of cigarettes of 20 units each” rather than just the number of cigarettes. Nevertheless this information has been transformed into number of cigarettes only.

Characteristics of the data

A table of descriptive statistics is given in the Appendix; some important features of the data are described here. For the first sweep, the sample is stratified between men (44.4%) and women (55.6%). The number of smokers is 2,335 corresponding to 11.8% of the sample where 74.2% are men and 25.8% are women. For the second sweep, the sample is stratified between men (43.9%) and women (56.1%). The number of smokers is 1,967 corresponding to 9.5% of the sample where 75.8% are men and 24.2% are women. Similar characteristics of the data is observed if the focus is only on those who were successfully followed-up which, as previously noted, correspond to 15,036 individuals. Within this sub-sample, 41.3% are men and 58.7% are women. And, in the first sweep, the number of smokers is 1,616 or 10.7% of whom 73.3% are males and 26.7% are females. By the second sweep, the number of smokers has fallen to 1,463 individuals (almost 10% decrease) of whom 74.6% are males and 25.4% are females.

Among the sub-sample of attritors (*N*=4,769; 54.3% men, 45.7% women), 719 individuals were smokers (15.1%) of whom 76.1% are men and 23.9% are women. Among the sub-sample of “new comers” (*N*=5,571; 50.9% men, 49.1% women), 504 individuals are smokers (9.0%) of whom 78.9% are men and 21.1% are women.

The data revealed marked changes amongst individuals regarding their smoking status over time. The number of smokers decreased by almost 16% even though the number of observations in the sample increased by the second sweep. This can be seen in Table 2 which summarises the overall number of individuals with their respective smoking status by the time of the survey for the entire dataset available.

Changes in smoking status between the two sweeps have been observed also for those who were successfully followed-up. To better summarise these changes, Figure 1 shows the number of observed individuals and their respective smoking status throughout the two sweeps. This figure shows, for example, that in the first sweep, 12,480 individuals reported that did not have the habit of smoking. By the second sweep, this number is reduced to 11,831 individuals. This means that 649 individuals have become smokers at some point between the two sweeps although only 481 have been categorised as current smokers by the second sweep.

Furthermore, in the first sweep 2,556 reported that have or had the habit of smoking but only 1,616 individuals were current smokers at the time. However, by the second sweep, 769 had quit and so the remaining 857 individuals were smokers during the duration of the survey. In addition, the survey reports the presence of 940 individuals who had the habit of smoking but did not report being smokers in the first sweep therefore they are former smokers (quitters). However, the data also shows that, by the second sweep, 135 of them have returned to active smoking.

Figure 1 reveals then, that the number of current smokers observed in the second sweep is the result of a mixed group of smokers. Thus, among 1,463 of current smokers identified, 847 of them have smoked throughout both sweeps, 135 were quitters but returned to smoking by the second sweep and 481 individuals who did not have the habit of smoking but decided to smoke and it was reported by the second sweep.

Participants and non-participants

For the panel hurdle model the focus is on two special types of participants which are invariably found within those who where followed-up. On the one hand, there are individuals who never had the habit of smoking over the period of the survey therefore they are of zero-type or simply "no-participants". On the other hand, there are individuals who had the habit of smoking in both sweeps therefore they are "participants". Tentatively, the number of participants is 847 or 5.6% of the sample of whom 74.1% are men and 25.9% are women. In turn, the number of non-participants is 11,831 of whom 33.0% are men and 67.0% are women. These figures, however, underestimate the number of participants and non-participants alike given the changing of smoking status reported in the survey. Thereby, irrespective of the switching in and out behaviour, it is possible to label 2,232 individuals for each sweep as participants even though consumption of cigarettes may zero at either period for some of them (but not in both). This figure is in fact, the actual number of participants considered in the panel hurdle model. Stratifying the number of participants by gender it is found that 73.9% are men and 26.1% are women. In turn, the number of non-participants is 12,804 of whom 35.7% are men and 64.3% are women. A table of summary statistics for this special sub-sample is also found in the Appendix.

The distributions of data on cigarette consumption

One important characteristic of the data is that it is extremely censored, with only 11.8% and 9.6% of individuals reporting consumption of cigarettes in the first and second sweep respectively. Thus, the observed large cluster of zeroes cannot be ignored econometrically which

makes the double-hurdle framework suitable for this analysis. However, an additional characteristic seen in the data which is of concern is that the distribution of reported consumption of cigarettes appeared to be non-normally distributed. Although this is common in studies of this kind, this makes it necessary to undertake a suitable variable transformation in order to obtain consistent ML estimates. In the previous chapter, a transformation of the dependent variable was avoided by estimating a GLM model with a log-link function which internalises the logarithm transformation. As this model is not well suited for panel data, the proposed methodology addresses non-normal distributions using a more general framework: the Box-Cox transformation.

Preliminary examination of the data revealed a positively skewed distribution explained by the spread of values reported in the sample. For example, the truncated sample of the first sweep, the average consumption of cigarettes smoked per week was 40.5 (SD 49.71) ranging from 1 to 440. The average figure is larger than the median of 20 and the estimated value for kurtosis (kurtosis = 14.24) is well above its expected value of 3 for a normal distribution. Similarly, the truncated sample of the second sweep, the average consumption of cigarette smoked per week was 41.9 (SD 52.82) ranging from 1 to 420. Similar average figures are also identified for some sub-samples mentioned throughout this analysis corresponding to attritors, new comers and those followed-up. The data reveals that amongst attritors the average number of cigarettes smoked among all smokers was 36.4 (SD 45.43) ranging from 1 to 420. Among new-comers the average consumption of cigarettes was 36.7 (SD 45.28) ranging from 1 to 400. Lastly, amongst those individuals who were followed-up the weekly average consumption of cigarettes was 42.4 (SD 51.40) ranging from 1 to 440 in the first sweep. By the second sweep, the weekly average consumption of cigarettes reported increased to 43.7 (SD 55.08) ranging from 1 to 420.

To see the distribution of the truncated samples, kernel density estimation plots are presented in Figure(s) 2. For a better appreciation of the data, a normal density plot has been superimposed to the actual density. It is worth noticing the presence of the long tail to the right of the distribution which is due to the small number of extreme values reported. The data reveals that, out of all smokers, 44 individuals in the first sweep (1.9%) and 47 individuals in the second sweep (2.4%) can be considered heavy smokers having reported to consume more than 140 cigarettes on average per week. This the usual cut-off value for heavy smoking (Baumert, Ladwig et al. 2010). In the other extreme, 81 individuals in the first sweep (3.5%) and 68 individuals in the second sweep (3.4%) reported having consumed only 1 cigarette on average per week. Thus, the majority of smokers in the sample can be taken as either light or moderate smokers.

Given the observed skewed distribution of reported cigarette consumption, this analysis will address this by considering the transformation of the dependent variable using the Box-Cox

transformation presented in equation (7) above. From all the models estimated, the optimal Box-Cox parameter (λ) has been found to be around 0.7 (see below) which is significantly different from 0 for a logarithm transformation and significantly different from 1 which may have suggested to ignore this issue altogether. To see the effect of the transformation on the distribution, Figure(s) 3 reproduces the information presented in Figure(s) 2 but now with transformed data. These figures as well present information only for the truncated samples. Figure(s) 3 clearly shows that the Box-Cox transformation improves the skewedness seen in the distribution of raw data. However, it is worth noting that although the model would yield consistent estimates, the transformation of the dependent variable does not seem to bring the distribution into a normal distribution. For the first sweep for example, a common statistical test for skewedness and kurtosis rejected normality at conventional levels of significance[5]. In fact, the average figure of y^T for the truncated sample is 16.9 (SD 15.9) which is still larger than the median of 10.8, although the estimated value for kurtosis is now closer to its expected value of 3 for a normal distribution (kurtosis = 7.3). All this means that perhaps in the margin, the Box-Cox transformation may just be providing the distribution needed to achieve consistency given the characteristics of the data at hand. This result is better than its alternative of ignoring this issue which is still a common approach used by researchers in empirical work.

Explanatory variables

Explanatory variables considered during the model selection procedure include socio-economic, demographic and psycho-social factors which are commonly used in the literature of tobacco demand and these are believed to influence individual preferences. Such variables have been defined and discussed in the previous chapter.

It is worth noting that most of the explanatory variables included in the model are of a dichotomous nature with only two continuous variables: income and age. Their effect in each hurdle was analysed in the 2PM used in the previous chapter. However, given that the number of participants differs for the panel model, their effect in each hurdle has also been found to be different. To see this, univariate non-parametric regressions have been employed to investigate the relationship between smoking participation and consumption of cigarettes on age and income respectively. Figure(s) 4 and 6 show how the prevalence of smoking depends on age and income. In turn, Figure(s) 5 and 7 show the relationship of consumption of cigarettes against age and income respectively with the truncated sample. In all figures, the curve fitted through the scatter plot is a "smooth via local polynomials" available in STATA with the user's written command `lpoly` of Gutierrez, Linhart et al. (2003). This smoother 'involves fitting the response to a

[5] `sktest` command in STATA.

polynomial form of the regressor via locally weighted least squares' (Ibid). To increase the readability of the curves, the scatterplot has been supressed and bandwidth has been increased to a double its default value. Lastly, two versions of each plot are presented. On the figure of the left, the degree of the polynomial smooth has been set to zero corresponding to the "local mean smoothing". On the figure of the right, the degree of the polynomial has been set to which visually the curve fits the curve best. Confidence intervals plots have also been added.

From Figure(s) 4 and 5, it is possible to see that age shows a clear (inverted) U-shape tendency on both smoking prevalence and consumption of cigarettes which suggests that this variable should enter in each hurdle with a polynomial. In turn, Figure(s) 6 and 7 show that income has a strong positive effect on both the prevalence of smoking and consumption. This means that variable income should enter in each hurdle as linear.

Results

Two set of results are included in this analysis. The first set includes estimates from the DH and BCDH model introduced in the third section (equations 6 & 10) which has been estimated by pooling all the information available. Pooled regressions or "population-average models" have been estimated first with the unbalanced panel and then with the balanced panel. In all cases, an indicator variable for observations corresponding to the second sweep, "wave 2", has been added to allow the intercept to vary over time. This helps to investigate changes on the probability of smoking participation and of changes in the level of cigarette consumption over the period of the survey. In addition, following the recommendation by Jones (2007) indicators for individuals who were not followed-up, "attritors" and "new comers", have been included in regressions with the unbalanced panel to investigate possible systematic differences between these two sub-samples on the overall results. All pool regressions have been augmented with robust standard errors which have been obtained by the clustering of observations within individuals; this helps to correct for the correlation in individual errors (Cameron and Trivedi 2005). Furthermore, as the error terms are likely to be correlated over time for a given individual, one additional set of pool regressions have been estimated which include the *Mundlak* framework due to Mundlak (1978). This involved adding within-individuals means of the time varying variables [see Jones (2007)]. Preliminary work showed that the identification of parameters for all models were not an issue given that the same set of explanatory variables were used in each equation (or hurdle) such that $\boldsymbol{x}_1' = \boldsymbol{x}_2'$. This has been the same approach followed in the 2PM estimated in the previous chapter.

The second set of results presented in this analysis includes the estimation of the panel hurdle model (equation 20) also introduced in the third section of this chapter. It should be emphasised, however, that the regression results from this model are different from the pool regressions discussed above in three main aspects. Firstly, the estimation is not restricted to the observed number of participants for each sweep. Rather, the panel hurdle model take into consideration the irrevocable and immutable condition of being a smoker therefore an equal number of participants have been added in each sweep.

Secondly, when estimating the panel model, exclusion restrictions have been imposed for a more robust identification. Imposing the same variables in each hurdle may not be efficient as convergence of the log-likelihood function takes considerable amount of time. Although using the same set of explanatory variables in each hurdle allows investigating the direction of influence of each variable used, which is of interest, albeit at the expense of over fitting the model. Nevertheless, the underlying reason for the exclusion restriction being imposed is the identification both of the "non-economical hurdle" and the "economical hurdle". Ideally, variables reflecting the individual's social, psychological or ethical characteristics may be included in the first hurdle (Dong and Kaiser 2008). This may better explain the probability of an individual reporting being a smoker ("non-economical hurdle"). In contrast, socio-economic variables such as income may be included in the second hurdle which may explain the individual's level of cigarette consumption (conditional to being a smoker) which is then referred as the "economical hurdle". In a series of trials, the panel double hurdle was first estimated with all explanatory variables available in each hurdle. The estimation results, which are not presented here, do show the presence of such hurdles. Consequent trials involved adding only relevant variables to each equation mostly when these are found to be appropriate or statistically significant. This procedure allowed identifying "age", "male", "indigenous", "education" and "other smoker" as ideal variables to include in the first hurdle. All the variables available for this research have been included in the second equation including an indicator variable for "wave 2" which allows the intercept to vary overtime.

Thirdly, the variable "age" has been added to control for a possible cohort effect. In this respect, this variable corresponds to the individual's age at the time of the first wave which corresponds to the year of 2002. Lastly, estimates from the panel hurdle model are accompanied with estimates from the DH and BCDH which have been obtained using the same data. These are included to draw qualitatively comparisons with estimates of the population-average models which do not account for the panel structure of the data.

Pool regression results

Estimates obtained from estimating the DH are presented in Tables 3 & 4 which corresponds to pooled regressions with an unbalanced panel. The DH has also been estimated with the balanced panel and estimates are presented in Table 7. Similarly, the BCDH has been estimated with the unbalanced panel (Tables 5 & 6) and with the balanced panel (Table 8). These regressions assume that there is a single cross-section and does not account for the panel structure of the data. These results are presented in this analysis so make comparison with those presented in the previous chapter.

There are several important findings worth mentioning at this stage from pool regressions presented in Tables 3-6. For instance, the *Mundlak* framework revealed the importance of taking into account possible correlated individual effects. Despite the fact that robust standard errors are computed, these appeared to be biased. The identified bias is downwards which overestimates the effect of the panel structure. This can be seen by comparing the magnitude of the standard errors of both the DH and BCDH which are, in general, smaller than the standard errors with *Mundlak*. This led to the wrong computation of *p*-values which made some coefficients statistically significant. Thus, interpretations of results are only given based on estimates obtained utilising the *Mundlak* framework (Tables 3, 5, 7 & 8).

As previously noted, all pool regressions of the DH and BCDH have been estimated with the balanced panel data as well (Tables 7 & 8). There was, however, some statistical evidence that the presence of attrition bias would confound the results. In the previous chapter, estimates of the 2PM for the first sweep showed evidence that "attritors" were more likely to report being a smoker than those who were followed-up. There was no evidence though that their level of cigarette consumption was systematically different from the remaining sample of smokers. And yet, it is possible to see from Table 3 that, when using the entire data available (unbalanced panel), the statistical evidence that "atrittors" are more likely to be smokers is weak. In fact, this is only true for the DH (with *Mundlak*) but no evidence of possible bias is detected on either hurdle under the BCDH (Table 5)[6]. One explanation could be the presence of "new comers" in the sample by the second sweep. But the results from the 2PM (of the second sweep) showed that "new comers" are neither more likely to be smoker nor smoke in different intensity in comparison to the rest of smokers in the sample who were followed-up. It is perhaps then, the pooling of information and the use of such data under a different statistical framework which may explain

[6] Preliminary work did show that attritors are more likely to be smokers by estimating the DH and BCDH which is a similar finding of the result from the 2PM of the previous chapter. However, this result only holds from estimates from the first sweep.

this finding. Thus, there is no reason to believe that attrition may confound the results and, on this evidence, interpretations of results are given only with estimates obtained with the balanced panel (Tables 7 & 8).

Some commentaries of results are as follows. At least when it comes to the first hurdle or the smoking participation equation, estimated coefficients in both the DH and BCDH are of the same sign and, in general, of the same statistically significance. However, the magnitudes of the standard errors in the BCDH are smaller than those presented for the DH. This suggests that by estimating the BCDH some efficiency has been gained. Nevertheless, from the estimates presented it is possible to see that participation is explained, in general, by individuals' characteristics just as the wider empirical evidence suggests. For example, explanatory variables with positive and strong significant effects include being male, living with another smoker and living in the West region. The negative sign on the coefficient for the "wave 2" dummy suggests that the probability of being a smoker decreased by the second sweep.

A similar pattern of results is observed for the smoking intensity equation (second hurdle). Estimates for the DH are in general in accordance with those obtained for the BCDH in sign and of the same statistically significance. The level of cigarette consumption is identified by a variety of demand shifters. There is seems to be one effect unique to the balanced panel. This has to do with how "income" should enter on either hurdle. In the previous chapter, the effect of higher polynomials was observed. Under the DH or BCDH with unbalanced panel, only the intensity equation showed signs of a quadratic term (Tables 3-6). As we can see from Table 7 or 8, the effect of polynomial has disappeared.

Panel results

Estimates for the panel hurdle model are presented in Table 9. The table also contains estimates from pool regressions using the DH and BCDH models for comparison. However, there are two additional reasons for obtaining and presenting these estimates. Firstly, depending of the model specification, estimates from pool regressions are used as suitable starting values for estimation of the ML function otherwise convergence would not be possible[7]. Secondly, it is important to draw qualitative comparisons with estimates from models which do not account for the structure of the data (DH and BCDH) with those models that do (panel hurdle model). Given the imposition of exclusion restrictions, commentaries of results are given by equation. The attention is on the estimates obtained for the panel hurdle model.

[7] Estimates from the DH are used as starting values for the BCDH. In turn, estimates of the BCDH are used as starting values for the panel hurdle model.

FIRST HURDLE

The results show a strong gender effect. Men are found to be more likely than women to report being a smoker or, as in the jargon of the double-hurdle framework, of passing the first hurdle. Ethnicity seems to have no effect on participation. Individuals who belong to an indigenous group are just as likely to be non-smokers. Furthermore, negative signs are shown on the coefficients for all levels of education considered on the participation equation. Notwithstanding, the statistical significance is strong only on the coefficients for "primary" and "university". For the DH and BCDH, education does not appear to have an effect on smoking participation.

Results from the variable reflecting risky behaviour shows the expected result. Living with "another smoker" shows a strong effect on participation. This shows the effect of close social interaction with another smoker which reinforces the habit of smoking.

SECOND HURDLE

The observed effects of socio-demographic variables on consumption are as follows. Conditional to being a participant, males smoke in greater intensity than female smokers. This result not only confirms that gender plays a dominant factor in determining smoking behaviour but, as a consequence, the demand for cigarettes. Individuals who belong to an indigenous group smoke fewer cigarettes than those who do not. Moreover, individuals who live in urban areas appear to smoke more than those living in rural areas. In terms of marital status, it is found that individuals who are divorced or widowed smoke more than those who are single.

Among the effect of socio-economic variables, it is observed that education does not play a role on the level of consumption. This result contrasts from those obtained in the DH and BCDH in which education has a (demand) shifter effect. In other words, conditional upon smoking, having an education makes smokers to smoke less. In addition, income has a strong effect in the intensity equation as expected, and the same effect is observed for the DH and BCDH. The effect of this variable shows a nonlinear pattern with respect to the consumption of cigarettes.

The indicator variable for consumption of alcohol exerts a positive and strong effect on the second hurdle. This means that drinking alcohol makes a smoker to consume more cigarettes relative to smokers who do not. The same effect is identified both in the DH and BCDH. Similarly, a smoker who lives with another smoker consumes more cigarettes than those who do not. In consequence, the evidence presented suggests that current smokers are more likely to become heavier smokers if they live with another smoker.

Other variables with strong statistical effect include the indicator variables for two regions within Mexico. Although regional results have been mixed throughout, panel estimates confirm two regularities. The first one is that individuals from the South-East smoke less relative to those who are from the Centre (area of Mexico City). In contrast, individuals from the North-West smoke with greater intensity than those in the Centre. In the absence of prices, regional indicators may capture some differences on the tobacco market. It seems likely, however, that these indicators may also capture social attitudes towards smoking especially from such regions of the country.

The results also show that, conditional to being a smoker, an individual who is a "house maker" smokes fewer cigarettes and individuals who live in a house which is owned by a member of the same household appear to smoke less relative to those who do not. In addition, the negative sign on the coefficient for the "wave 2" dummy suggests that consumption of cigarettes has decreased by the second sweep.

There is one additional parameter worth discussing. The correlation coefficient, ρ, which is usually ignored in cross-sectional analysis. This parameter has been found to be of magnitude of -0.362. Firstly, the coefficient is of negative sign which means that an individual who has an unusually high propensity to be a smoker, is expected to smoke a smaller quantity if they are a smoker; an individual who has an unusually low propensity to be a smoker, is expected to smoke a larger quantity if they are a smoker.

Discussion

In the first set of results provided, a DH and BCDH model has been employed to model the demand for cigarettes. These models are part of the standard methodology for this kind of analysis where there is zero mass problem. In contrast with the 2PM employed in the previous chapter, these models have been used to take advantage of the entire dataset available. However just as the 2PM, the data has been taken as cross-sectional and the models do not take into account its panel structure. This exercise has been carried out to investigate robust results in regards to factors driving the likelihood of an individual to becoming a smoker and factors driving an individual to smoke in a greater or lesser intensity. Recalling that the 2PM only allowed investigating these factors by each sweep (wave 1 or wave 2), in this chapter, the DH and BCDH allowed to pool both sweeps.

Using pool hurdle models has not substantially changed the overall findings encountered from using a single cross-sectional. However, key variables have emerged as the most important

factors influencing the demand for cigarettes. For example, the findings of this analysis suggest that, indeed, smoking follows an age profile. In particular, at a younger age, the probability of becoming a smoker is high but declines with time perhaps as individuals become more aware of the health consequences of smoking. However, the hypothesis that individuals become wiser as they become older [therefore less likely to become a smoker (Yen 2005a)] is partially supported due to the inflection point encountered. Furthermore, the findings suggest that education has an effect on the level of consumption. This result appears consistent as the evidence suggests that 'education may improve the cognitive skills of an individual regarding the health risk associated with cigarette smoking' (Ibid). Although the role of education has been well-documented in studies of tobacco demand, a particular finding is that, in the long run, education only influences participation and not intensity. The findings also suggest that cigarette consumption is largely influenced by gender. In Mexico, as in many other countries, the rate of smoking prevalence amongst men is more than twice that of women. It is perhaps not surprising that men are both more likely to be smokers and smoke more cigarettes. It was not possible, however, to investigate whether there are differences in smoking behaviour between men and women as found by Yen (2005a), Yen (2005b) and Bauer, Göhlmann et al. (2007). In addition, the effect of ethnicity on consumption appears consistent with the evidence provided from 2PM model of the previous chapter in which it was suggested that being part of an indigenous group has little or no effect on the probability of being a smoker, only on consumption. This is an interesting result given that the wider literature usually shows significant effects on both hurdles concerning minorities, albeit not always with the same sign [for an example see Su and Yen (2000)].

Another relevant socio-economic factor encountered was that homeowners and house makers smoke fewer cigarettes. For the first case, this finding has been consistent throughout the literature [see for an example Atkinson, Gomulka et al. (1984)] which may reflect the individual's condition of being "rent-free" which otherwise may be a source of psychological stress and therefore the need to smoke in greater intensity.

The effect of risky behaviour has appeared to be relevant in explaining cigarette consumption. For instance, alcohol consumption encourages smokers to consume more cigarettes. The association between smoking and drinking is well-known in the literature therefore this result confirms the complementary relationship (Aristei and Pieroni 2008). Furthermore, being in close contact with other smokers has a positive effect on participation and intensity. This was also observed by Jones (1989).

The standard double hurdle model has been extended to account for the panel structure of the data. Given the addictive nature of cigarettes, to analyse consumption under a cross-sectional framework implies that life-cycle considerations such as preferences, learning behaviour and

purchase carryover are explicitly ignored. However, the results produce few differences if compared with pool regression estimates. Also, the panel model introduced here incorporates the Box-Cox transformation of the dependable variable to address the skewed distribution seen in data. This causes the unobserved error to be non-normally distributed which is a source of inconsistency in the estimates. The panel estimator appears to provide further insights regarding the consumption of cigarettes in Mexico in recent years.

It is worth emphasising the estimates from the panel hurdle model are "heterogeneity free" as it has been possible to control for unobserved individuals' characteristics. This has revealed (observable) characteristics that make an individual more likely to be a smoker and what make smokers, smoke to a greater or lesser intensity.

An important result obtained from the estimates of the panel model is that the quantity of cigarettes consumed by smokers gets reduced over time. This is reflected by the negative sign of the coefficient of "wave 2". Given that all other determinants of smoking behaviour have been included, an interpretation of this effect is that any change in consumption between wave 1 and wave 2 can only be explained in terms of taste changes resulting from anti-smoking legislation undertaken in Mexico since 2002. Data used in this analysis corresponds to a time of pre-tobacco-control policies (wave 1) to a time of post-tobacco-control policies (wave 2). Although the majority of such policies have been targeted at reducing smoking prevalence, the findings may suggest that such policies are now affecting the level of consumption.

There are some important limitations to this analysis. Firstly, a variable usually considered in similar studies, price, has been omitted in all models estimated. As noted in the previous chapter, the MxFLS does not specifically collect information on prices of cigarettes. Moreover, it has not been possible to find suitable proxies that would match the individual and their area of residence. Secondly, given the statistical complexity of the BCDH and panel hurdle model, it has not been possible to provide a decomposition of the unconditional mean, given certain individuals' characteristics. This is usually done by calculating marginal effects which, in the presence of indicator variables, involves calculating changes in the probability and the unconditional and conditional level of consumption that results from a finite change in each variable. Thus, this study has been restricted to analyse the direction of influence of each variable on smoking participation and the level of cigarette consumption. Perhaps for more insightful policy prescriptions further research may take these considerations into account. Lastly, the estimation of the panel model brings the possibility of analysing smoking behaviour within a life-cycle. However, at the time of writing this chapter, only data from two waves of the MxFLS have been used. Thus, life-cycle effects are only detected to a modest degree. This issue is, perhaps, left for future research.

ANNEX TO CHAPTER III

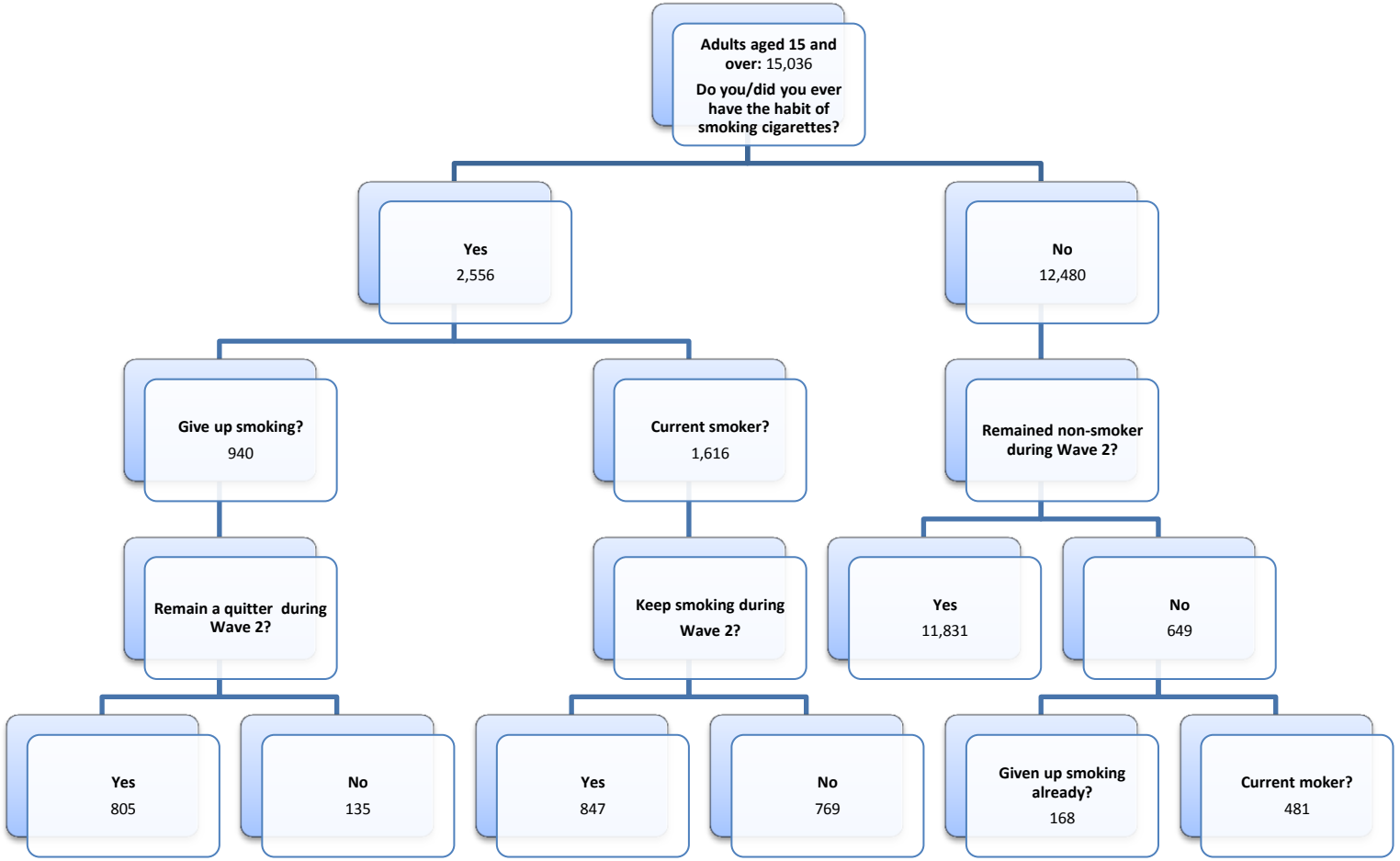
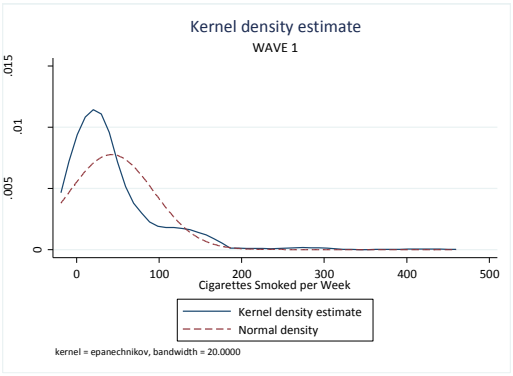
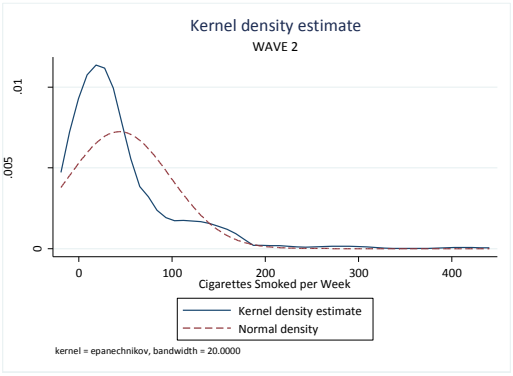
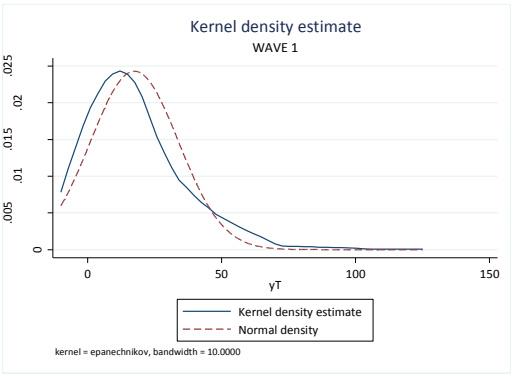


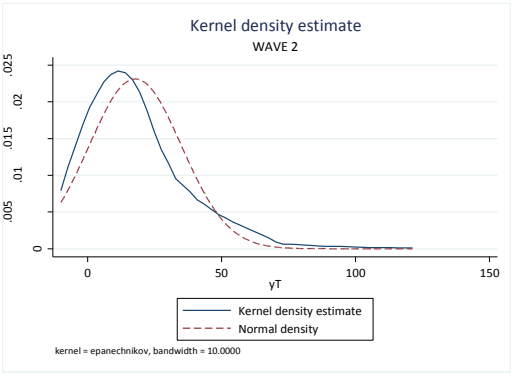
Figure 1 Observed Transition of Smoking Status, MxFLS 1 & 2.



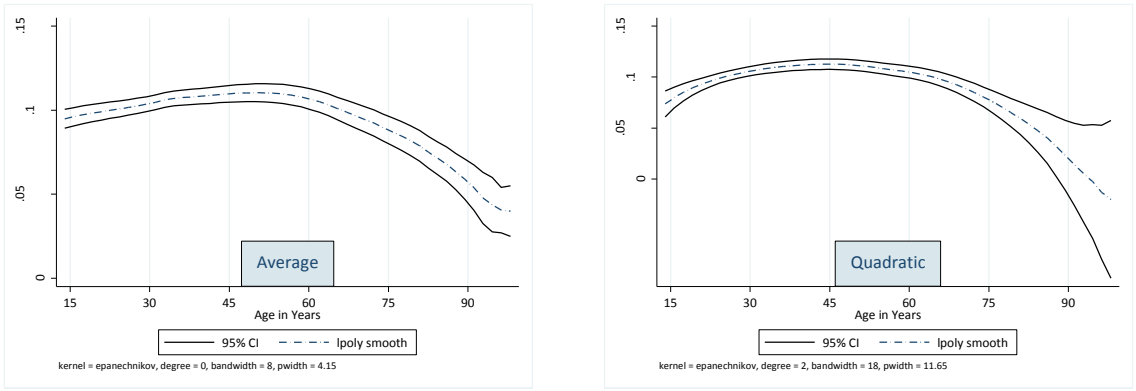


Figure(s) 2. Kernel Density Estimate of Cigarettes smoked per Week, MxFLS 1 & 2.





Figure(s) 3 Kernel Density Estimate of Cigarette smoked per Week using Box-Cox Transform ($\boldsymbol{y^T}$) with $\lambda = \mathbf{0.7}$, MxFLS 1 & 2.



Figure(s) 4 Local Polynomial Regression of Dichotomous Variable representing Smoking Participation, Against Age (Balanced Panel) MxFLS 1 & 2.

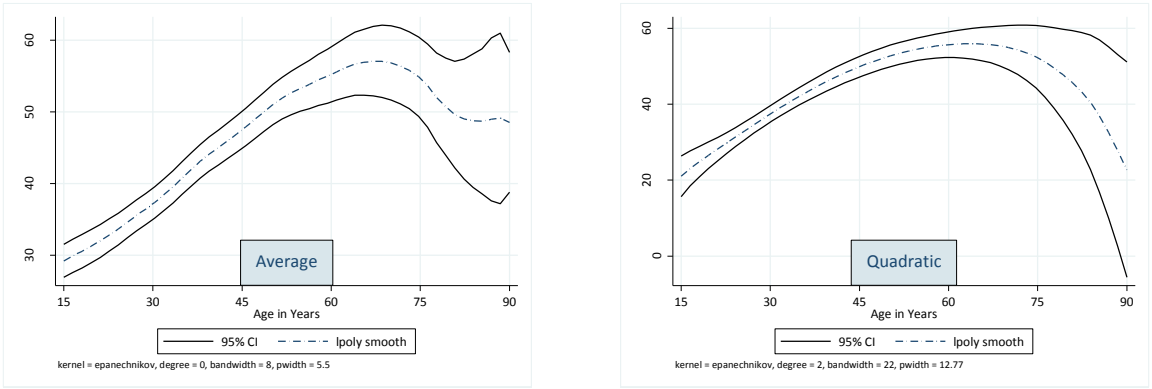


Figure 5 Local Polynomial Regression of Cig. Consumption Against Age, Smokers Only (Balanced Panel), MxFLS 1 & 2.

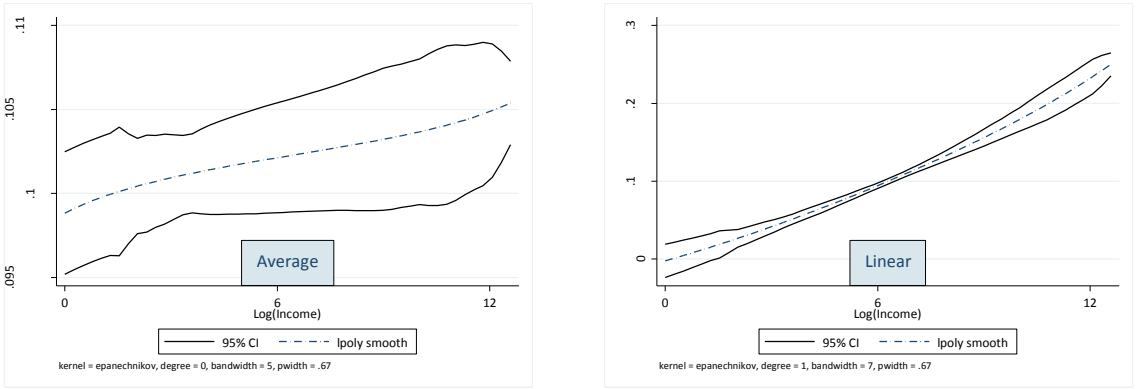


Figure 6 Local Polynomial Regression of Dichotomous Variable representing Smoking Participation, Against Income (Balanced Panel), MxFLS 1 & 2.

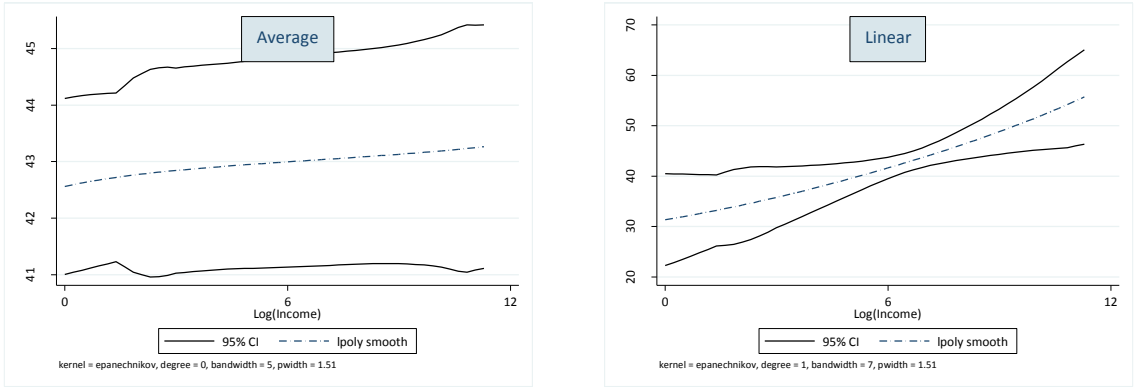


Figure 7 Local Polynomial Regression of Cig. Consumption Against Income, Smokers Only (Balanced Panel), MxFLS 1 & 2.

Table 1. Applications of the Box-Cox Double-Hurdle Model.

Reference	*Application*
Yen (1993)	Food away from home (US)
Yen (1994)	Alcohol consumption (US)
Yen (1995)	Alcohol consumption (US)
Yen, Dellenbarger et al. (1995)	Crawfish consumption (US)
Yen and Jones (1996)	Smoking (UK)
Burton, Dorsett et al. (1996)	Meat (UK)
Jones and Yen (2000)	Beef consumption (US)
Chaze (2005)	Health expenditure (Switzerland)
Moffatt (2005)	Loan default (UK)
Martínez-Espiñeira (2006)	Wildlife valuation (Canada)
Aristei and Pieroni (2008)	Tobacco consumption (Italy)
Aristei and Pieroni (2009)	Smoking (Italy)
Keelan, Henchion et al. (2009)	Food away from home (Ireland)
Bai, Wahl et al. (2010)	Food away from home (China)
Genanew and Alemu (2012)	Land productivity (Ethiopia)

Table 2. Distribution of observations according with the individual's smoking status, MxFLS 1 & 2.

	Wave 1	*Wave 2*	*All sample*
Have or had the habit of smoking	3,631	3,840	7,471
Active Smokers	2,335	1,967	4,302
Ex-smokers	1,296	1,873	3,169

Table 3. Double-Hurdle Model – Population Average, MxFLS 1 & 2.				
	Double-Hurdle with Mundlak			
VARIABLES	D		Y	
Age	-0.0930*	(0.0519)	3.153***	(0.369)
Age^2	0.000868	(0.000547)	-0.0340***	(0.00418)
Income	-0.210	(0.436)	-20.40***	(3.091)
Income^2 (ln)	0.0584	(0.0524)	2.427***	(0.304)
Male	2.309***	(0.328)	37.07***	(3.228)
Indigenous	0.00605	(0.375)	-9.747**	(4.154)
Urban	0.719***	(0.264)	5.173*	(3.074)
Married	-0.787	(0.567)	8.763***	(2.611)
Divorced	-0.219	(0.582)	26.35***	(4.677)
Widowed	-0.0508	(0.475)	20.04***	(6.486)
Primary	-0.175	(0.302)	-14.94***	(5.585)
Secondary	0.938	(0.625)	-18.36***	(5.945)
High-School	0.495	(0.622)	-22.53***	(6.238)
University	5.336	(12.24)	-32.00***	(6.343)
South-East	-0.542*	(0.322)	-16.59***	(4.132)
West	1.026***	(0.297)	-5.718*	(3.016)
North-West	0.476*	(0.281)	-3.472	(3.173)
North-East	2.297*	(1.240)	-3.624	(3.012)
House maker	-0.205	(0.231)	-10.99***	(4.076)
Homeowner	0.0224	(0.210)	-14.25***	(2.260)
Health Insurance	0.700*	(0.362)	-6.324***	(2.064)
Alcohol	0.153	(0.252)	34.74***	(2.340)
Hot Drinks	-3.188	(2.575)	50.24	(122.2)
Exercise	0.712	(0.592)	-3.675*	(2.087)
Other Smoker	0.647***	(0.198)	77.13***	(2.534)
Wave 2	-0.529***	(0.173)	-5.345***	(1.728)
Attritors	0.558*	(0.323)	2.696	(2.462)
New Comers	0.476	(0.441)	-3.779	(2.911)
Constant	-168.2***	(14.35)	0.486	(1.678)
σ_ε	79.81***	(1.984)		
Observations	40410			
Log-likelihood	-30986			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Double-Hurdle Model – Population Average, MxFLS 1 & 2.				
	Double-Hurdle			
VARIABLES	D		Y	
Age	-0.0933**	(0.0438)	3.330***	(0.356)
Age^2	0.000913**	(0.000461)	-0.0358***	(0.00401)
Income	-0.204	(0.311)	-16.60***	(2.388)
Income^2 (ln)	0.0495	(0.0386)	1.940***	(0.233)
Male	2.207***	(0.275)	36.97***	(2.960)
Indigenous	0.0583	(0.331)	-10.96***	(3.879)
Urban	0.750***	(0.234)	5.642**	(2.690)
Married	-0.648	(0.425)	7.927***	(2.624)
Divorced	-0.200	(0.503)	26.23***	(4.767)
Widowed	-0.0307	(0.444)	20.00***	(6.449)
Primary	-0.187	(0.248)	-13.78***	(5.182)
Secondary	0.932**	(0.467)	-16.82***	(5.526)
High-School	0.731	(0.492)	-20.85***	(5.759)
University	2.641	(1.681)	-29.22***	(5.916)
South-East	-0.600**	(0.298)	-16.01***	(3.979)
West	0.998***	(0.301)	-5.497*	(3.061)
North-West	0.553**	(0.258)	-3.154	(3.093)
North-East	1.934***	(0.615)	-3.062	(3.033)
House maker	-0.289	(0.218)	-9.792**	(3.964)
Homeowner	-0.0258	(0.198)	-14.01***	(2.239)
Health Insurance	0.596**	(0.270)	-5.640***	(2.064)
Alcohol	0.166	(0.252)	34.40***	(2.401)
Hot Drinks	-4.127**	(1.890)	101.1	(167.1)
Exercise	0.638*	(0.366)	-3.245	(2.045)
Other Smoker	0.594***	(0.184)	77.03***	(2.555)
Wave 2	-0.436***	(0.143)	-5.405***	(1.716)
Attritors	0.528	(0.322)	3.449	(2.426)
New Comers	0.332	(0.421)	-3.385	(2.959)
Constant	0.701	(1.331)	-177.2***	(12.38)
σ_ε	79.78***	(1.980)		
Observations	40410			
Log-likelihood	-30998			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 5. Box-Cox Double-Hurdle Model – Population Average, MxFLS 1 & 2.				
	Box-Cox Double Hurdle with Mundlak			
VARIABLES	D		Y	
Age	-0.0703*	(0.0363)	1.265***	(0.147)
Age^2	0.000676*	(0.000375)	-0.0138***	(0.00165)
Income	-0.137	(0.290)	-8.211***	(1.234)
Income^2 (ln)	0.0457	(0.0356)	0.951***	(0.120)
Male	1.980***	(0.287)	13.41***	(1.461)
Indigenous	-0.0520	(0.268)	-3.425**	(1.632)
Urban	0.661***	(0.188)	1.475	(1.134)
Married	-0.563*	(0.289)	3.735***	(1.045)
Divorced	-0.147	(0.366)	10.69***	(1.943)
Widowed	-0.0103	(0.353)	8.540***	(2.673)
Primary	-0.177	(0.222)	-5.908***	(2.183)
Secondary	0.723*	(0.418)	-7.925***	(2.358)
High-School	0.485	(0.469)	-9.635***	(2.488)
University	2.121	(1.946)	-13.70***	(2.519)
South-East	-0.484**	(0.238)	-6.394***	(1.708)
West	0.838***	(0.261)	-2.826**	(1.219)
North-West	0.421*	(0.255)	-2.172*	(1.268)
North-East	1.604***	(0.595)	-2.284*	(1.193)
House maker	-0.261	(0.179)	-3.317*	(1.724)
Homeowner	-0.0111	(0.165)	-5.583***	(0.922)
Health Insurance	0.467**	(0.216)	-2.519***	(0.826)
Alcohol	0.177	(0.231)	13.85***	(1.074)
Hot Drinks	-3.338**	(1.299)	36.60	(47.56)
Exercise	0.541**	(0.267)	-1.467*	(0.869)
Other Smoker	0.636***	(0.152)	30.39***	(1.196)
Wave 2	-0.472***	(0.123)	-1.880**	(0.771)
Attritors	0.436	(0.273)	1.043	(0.991)
New Comers	0.379	(0.306)	-1.717	(1.204)
Constant	-0.208	(1.101)	-62.61***	(5.834)
σ_ε	31.08***	(0.907)		
λ	0.736***	(0.00928)		
Observations	40410			
Log-likelihood	-30572			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Box-Cox Double-Hurdle Model – Population Average, MxFLS 1 & 2.				
	Box-Cox Double Hurdle			
VARIABLES	D		Y	
Age	-0.0778**	(0.0362)	1.341***	(0.146)
Age^2	0.000778**	(0.000378)	-0.0146***	(0.00165)
Income	-0.161	(0.280)	-6.668***	(1.008)
Income^2 (ln)	0.0423	(0.0363)	0.759***	(0.101)
Male	1.923***	(0.258)	13.59***	(1.336)
Indigenous	-0.0125	(0.257)	-3.862**	(1.583)
Urban	0.669***	(0.197)	1.820	(1.132)
Married	-0.529*	(0.276)	3.410***	(1.051)
Divorced	-0.168	(0.362)	10.67***	(1.956)
Widowed	-0.0192	(0.363)	8.431***	(2.721)
Primary	-0.196	(0.216)	-5.414**	(2.143)
Secondary	0.771**	(0.382)	-7.345***	(2.289)
High-School	0.656	(0.452)	-9.005***	(2.396)
University	2.331	(1.541)	-12.65***	(2.433)
South-East	-0.509**	(0.229)	-6.297***	(1.626)
West	0.852***	(0.248)	-2.755**	(1.224)
North-West	0.481**	(0.243)	-2.061*	(1.252)
North-East	1.520***	(0.472)	-2.062*	(1.197)
House maker	-0.312*	(0.175)	-3.031*	(1.643)
Homeowner	-0.0364	(0.161)	-5.487***	(0.915)
Health Insurance	0.430**	(0.206)	-2.263***	(0.833)
Alcohol	0.159	(0.219)	13.81***	(1.029)
Hot Drinks	-3.489***	(1.156)	38.57	(42.67)
Exercise	0.531**	(0.259)	-1.279	(0.847)
Other Smoker	0.609***	(0.148)	30.38***	(1.173)
Wave 2	-0.412***	(0.114)	-1.898**	(0.742)
Attritors	0.408	(0.264)	1.405	(0.981)
New Comers	0.288	(0.298)	-1.553	(1.198)
Constant	0.221	(1.083)	-66.97***	(5.212)
σ_ε	31.11***	(0.896)		
λ	0.736***	(0.00929)		
Observations	40410			
Log-likelihood	-30584			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Population average with *Mundlak* (Balanced Panel), MxFLS-1 & 2.				
	Double Hurdle with Mundlak			
VARIABLES	D		Y	
Age	-0.131*	(0.0686)	3.170***	(0.457)
Age^2	0.00113*	(0.000681)	-0.0324***	(0.00513)
Income	0.324**	(0.130)	3.462	(2.152)
Income^2 (ln)	-0.00901	(0.0147)	0.504***	(0.134)
Male	2.376***	(0.339)	42.28***	(3.749)
Indigenous	0.131	(0.392)	-11.48***	(4.418)
Urban	0.582*	(0.310)	9.712***	(3.200)
Married	-1.962**	(0.995)	5.606*	(3.299)
Divorced	-1.259	(1.082)	25.63***	(5.551)
Widowed	-0.938	(0.902)	18.94**	(7.614)
Primary	0.181	(0.315)	-21.10***	(6.790)
Secondary	1.861**	(0.893)	-23.50***	(7.288)
High-School	1.079	(0.666)	-25.06***	(7.598)
University	6.423***	(1.074)	-34.36***	(7.737)
South-East	-0.444	(0.362)	-20.41***	(4.793)
West	1.240***	(0.290)	-3.857	(3.838)
North-West	0.302	(0.323)	1.882	(3.985)
North-East	5.846**	(2.465)	-1.087	(3.996)
House maker	-0.0870	(0.240)	-13.99***	(3.967)
Homeowner	0.226	(0.278)	-14.65***	(2.894)
Health Insurance	1.033***	(0.309)	-8.445***	(2.455)
Alcohol	0.104	(0.272)	36.15***	(2.723)
Hot Drinks	-2.128*	(1.269)	17.74	(48.89)
Exercise	0.983	(0.742)	-3.667	(2.634)
Other Smoker	0.785***	(0.221)	78.88***	(3.091)
Wave 2	-0.602***	(0.169)	-5.424***	(1.690)
Constant	1.604	(2.569)	-250.2***	(16.99)
σ_ε	83.54***	(2.482)		
Observations	30072			
Log-likelihood	-22386			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 8. Population average with *Mundlak* (Balanced Panel), MxFLS-1 & 2.				
	Box-Cox Double Hurdle with Mundlak			
VARIABLES	D		Y	
Age	-0.111*	(0.0629)	1.250***	(0.188)
Age^2	0.000965	(0.000627)	-0.0129***	(0.00210)
Income	0.297**	(0.124)	1.296	(0.844)
Income^2 (ln)	-0.00750	(0.0123)	0.199***	(0.0528)
Male	2.197***	(0.308)	16.22***	(1.584)
Indigenous	0.0851	(0.336)	-4.124**	(1.797)
Urban	0.567**	(0.273)	3.635***	(1.278)
Married	-1.713*	(0.879)	2.538*	(1.307)
Divorced	-1.029	(0.966)	10.38***	(2.255)
Widowed	-0.818	(0.799)	8.294***	(3.104)
Primary	0.178	(0.276)	-8.849***	(2.718)
Secondary	1.648**	(0.706)	-9.991***	(2.908)
High-School	1.012*	(0.602)	-10.68***	(3.031)
University	7.510***	(2.008)	-14.61***	(3.089)
South-East	-0.392	(0.319)	-8.205***	(1.971)
West	1.178***	(0.265)	-2.116	(1.512)
North-West	0.322	(0.300)	0.0161	(1.606)
North-East	5.786	(3.600)	-1.424	(1.582)
House maker	-0.0844	(0.215)	-5.432***	(1.573)
Homeowner	0.204	(0.232)	-5.843***	(1.135)
Health Insurance	0.898***	(0.270)	-3.426***	(0.964)
Alcohol	0.0734	(0.231)	14.69***	(1.070)
Hot Drinks	-1.956	(1.336)	5.529	(20.44)
Exercise	0.820	(0.585)	-1.381	(1.051)
Other Smoker	0.759***	(0.197)	31.30***	(1.357)
Wave 2	-0.568***	(0.146)	-2.117***	(0.674)
Constant	0.921	(2.403)	-96.66***	(7.022)
σ_ε	32.65***	(1.115)		
λ	0.738***	(0.0111)		
Observations	30072			
Log-likelihood	-22095			

NOTE: Robust standard errors in parentheses adjusted for clustering to allow for repeated observations.
*** p<0.01, ** p<0.05, * p<0.1

Table 9. Variants of the Double-Hurdle Model (Balanced Panel), MxFLS 1 & 2.

	DH				BCDH				Panel-Hurdle with Box-Cox Transformation			
VARIABLES	*D*		*Y*		*D*		*Y*		*D*		*Y*	
Age	-0.112	(0.0553)**	2.775	(0.404)***	-0.076	(0.0465)	1.210	(0.237)***	-0.048	(0.0149)***	1.487	(0.248)***
Age^2	0.001	(0.000553)*	-0.031	(0.00458)***	0.001	(0.000514)	-0.013	(0.00306)***	0.000354	(0.000174)**	-0.013	(0.00313)***
Income			-50.140	(8.815)***			-19.940	(3.536)***			-11.220	(2.065)***
Income^2 (ln)			10.550	(2.023)***			4.215	(0.820)***			2.361	(0.489)***
Income^3 (ln)			-0.528	(0.132)***			-0.212	(0.0539)***			-0.118	(0.0320)***
Male	5.642	(0.668)***	45.130	(3.507)***	0.925	(0.151)***	16.350	(1.529)***	0.464	(0.103)***	13.990	(1.902)***
Indigenous	-1.002	(0.288)***	-6.770	(3.522)*	-0.153	(0.178)	-3.087	(1.711)*	0.051	(0.114)	-5.038	(1.904)***
Urban			9.829	(2.630)***			3.735	(1.003)***			3.558	(0.850)***
Married			3.216	(3.256)			0.890	(1.233)			0.515	(0.997)
Divorced			26.730	(5.209)***			10.450	(2.045)***			7.344	(1.655)***
Widowed			23.070	(6.675)***			8.813	(2.667)***			5.968	(2.228)***
Primary	-0.251	(0.345)	-15.940	(5.472)***	-0.215	(0.294)	-4.885	(3.905)	-0.383	(0.141)***	1.472	(2.329)
Secondary	3.935	(2.181)*	-19.590	(6.004)***	10.410	(1.967)	-13.250	(4.262)***	-0.175	(0.157)	0.253	(2.581)
High School	1.594	(1.394)	-24.300	(6.458)***	0.956	(0.580)*	-13.950	(4.286)***	-0.181	(0.172)	-0.272	(2.777)
University	3.870	(4.228)	-35.070	(6.711)***	10.200	(3.965)	-19.710	(4.416)***	-0.451	(0.172)***	0.285	(2.998)
South-East			-23.150	(3.793)***			-8.978	(1.497)***			-7.248	(1.351)***
West			2.326	(3.365)			0.588	(1.317)			1.162	(1.186)
North-West			3.039	(3.480)			0.965	(1.355)			2.846	(1.179)**
North-East			7.896	(3.613)**			2.714	(1.383)**			3.918	(1.245)***
House maker			-15.400	(3.446)***			-5.876	(1.346)***			-4.570	(1.090)***
Homeowner			-14.090	(2.491)***			-5.297	(0.948)***			-4.432	(0.791)***

Continue Next Page…

Health insurance			-5.443	(2.300)**			-1.951	(0.877)**			-1.038	(0.697)
Alcohol			35.060	(2.403)***			14.010	(0.926)***			11.000	(0.748)***
Hot drinks			-12.820	(30.46)			-8.373	(11.63)			-12.220	(7.697)
Exercise			-2.414	(2.499)			-0.925	(0.959)			0.131	(0.814)
Other smoker	0.332	(0.247)	82.280	(3.027)***	0.548	(0.135)***	29.760	(1.619)***	1.133	(0.0691)***	13.570	(0.946)***
Wave 2			-8.569	(1.444)***			-3.319	(0.574)***			-2.756	(0.528)***
Constant	3.140	(1.416)**	-182.200	(17.12)***	1.499	(1.188)	-63.940	(9.498)***	0.777	(0.368)**	-61.310	(6.658)***
σ_u									17.670	(1.024)***		
σ_ε	83.950	(2.453)***			30.660	(1.259)***			17.776ª	(0.6479151)***		
λ					0.732	(0.0113)***			0.695	(0.00946)***		
P									-0.362ª	(0.0630541)***		
Observations	30072				30072				30072			
Log-likelihood	-22417				-22119				-21384			

NOTE: Robust standard errors in parentheses. Standard errors for (DH) and (BCDH) are adjusted for clustering to allow for repeated observations.

ª Standard errors calculated with the Delta Method.

*** p<0.01, ** p<0.05, * p<0.1

Conclusions

This thesis has explored the distributional effect of the tobacco excise and investigated factors that determined smoking participation and level of consumption in Mexico. Several conclusions that can be drawn from this thesis and these are presented here. Further research and policy implications are also presented.

The first objective of this thesis has been to assess the fairness of the tobacco tax. This research has revealed that, indeed, this tax can be deeming as unfair or *regressive*. Two different measures of vertical equity have shown that the tax imposed on tobacco hit poor households harder than richer ones. The impact can be quite substantial. For households at the lower end of the income distribution, the impact of the tax can represent up to 7.6% of their consumption expenditure whilst for households at the higher end, it can represent up to 3.4%. Although the regressivity has been suggested in the past by studies which have analysed patterns of tobacco consumption, this study has provided a formal evidence of this.

The two measurements of vertical equity also contradict the conclusions drawn from the official reports of the incidence of taxation and public spending. These reports have suggested that the tax was progressive given that the share of fiscal revenue obtained from richer households is substantially larger than the share from poorer households. This study, however, has found that the methodology employed in the official reports fails to deliver a clear picture of the true distributional effect of this tax. However, the issue can be traced from the wording of the law that dictates the way in which the analysis of tax incidence must be conducted. This study, in fact, finds the wording of the law unfortunate. It is worth noting that the same assumptions used by the official reports in calculating the burden of this tax have been followed in this analysis. In consequence, a similar assessment should reach other indirect taxes valid at the federal level. Therefore, as long as indirect taxes are concerned, a more comprehensive methodology should be applied. In order to achieve this, the law that mandates the fiscal incidence reports should be reformed to allow more flexibility in the methodology employed. It could also mandate non-government organisations to carry out such studies as it is for the case for measuring poverty. This would provide a more objective assessment of the distributional impact of taxes in Mexico.

The second objective of this thesis has been to investigate the role of non-price factors determining the demand for tobacco. To this end, a two-part model has been applied to individual-level data on cigarette consumption. This model allows investigating the two decisions that explain consumption; whether to smoke and, upon smoking, how much to consume. However, in terms on the overall effects, the findings suggest that consumption of cigarette is determined by demographic characteristics such as gender and marital state and by psychosocial factors such as the drinking of alcohol and the close interaction with another smoker. These are the most relevant factors which positively affect consumption. This means that tobacco-control

policies should aim to target specific groups in society which would imply that current policies may have mixed effects. In addition, the findings suggest that education continues to be an important socio-economic factor especially when it comes to the decision to smoke. Therefore, anti-smoking campaign should take this into consideration for the design of better policies. Lastly, regional differences were presented. The findings reveal different attitudes towards smoking across regions within Mexico. This suggests that the use of taxation as a national tobacco-control policy is insufficient. Therefore, state governments should implement tobacco controls appropriate to their own territory.

The third objective of this thesis is to estimate the Cragg's model within a panel data context. Therefore, the demand for cigarettes using individual-level data has been extended from a static framework to a dynamic framework. The panel model introduced successfully addresses the censoring in information encountered and controls for the unobserved heterogeneity. In addition, it incorporates the Box-Cox transformation to contend with the long tail seen in data. A number of significant effects are found within the two consumption decision equations. However, one important finding is that potential smokers will tend to smoke less intensely. This is perhaps explained by the series of tobacco-control policies undertaken in Mexico since the early 2000's. The findings suggest that such policies are having a positive effect on the reduction of consumption. Thus, the continuation of more effective policies and enforcement should be the government's top priority.

Bibliography

2001. Art. 25. *Ley de Ingresos de la Federación para el Ejercicio Fiscal de 2001.* Mexico: Diario Oficial de la Federacion (DOF): 31/12/2000.

ANDREWS, R. L. & FRANKE, G. R. 1991. The Determinants of Cigarette Consumption: A Meta-Analysis. *Journal of Public Policy & Marketing,* 10**,** 81-100.

ARISTEI, D. & PIERONI, L. 2008. A double-hurdle approach to modelling tobacco consumption in Italy. *Applied Economics,* 40**,** 2463-2476.

ARISTEI, D. & PIERONI, L. 2009. Addiction, social interactions and gender differences in cigarette consumption. *Empirica,* 36**,** 245-272.

ATKINSON, A. B., GOMULKA, J., STERN, N. H., ECONOMIC & SOCIAL RESEARCH, C. 1984. *Household expenditure on tobacco, 1970-1980 : expenditure from the Family Expenditure Survey,* London, London School of Economics and Political Science].

BAI, J., WAHL, T. I., LOHMAR, B. T. & HUANG, J. 2010. Food away from home in Beijing: Effects of wealth, time and "free" meals. *China Economic Review,* 21**,** 432-441.

BAUER, T., GÖHLMANN, S. & SINNING, M. 2007. Gender differences in smoking behavior. *Health Economics,* 16**,** 895-909.

BAUMERT, J., LADWIG, K.-H., RUF, E., MEISINGER, C., DÖRING, A., WICHMANN, H.-E. & INVESTIGATORS, F. T. K. 2010. Determinants of Heavy Cigarette Smoking: Are There Differences in Men and Women? Results From the Population-Based MONICA/KORA Augsburg Surveys. *Nicotine & Tobacco Research,* 12**,** 1220-1227.

BISHOP, J. A., LIU, H. & MENG, Q. 2007. Are Chinese smokers sensitive to price? *China Economic Review,* 18**,** 113-121.

BOX, G. E. P. & COX, D. R. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological),* 26**,** 211-252.

BURKE, W. J. 2009. Fitting and interpreting Cragg's tobit alternative using Stata. *Stata Journal,* 9**,** 584-592.

BUTTENHEIM, A. M., WONG, R., GOLDMAN, N. & PEBLEY, A. R. 2010. Does social status predict adult smoking and obesity? Results from the 2000 Mexican National Health Survey. *Global public health,* 5**,** 413-26.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics methods and applications,* Cambridge; New York, Cambridge University Press.

CEFP 2009. Simulador de incidencia del impuesto al valor agregado - Nota metodologica. LX Legislatura - Camara de Diputados, Mexico.

CEFP 2009. Simulador de incidencia del impuesto especial sobre produccion y servicios en bienes y servicios no petroleros - Nota metodologica. LXI Legislatura - Camara de Diputados, Mexico.

CHALOUPKA, F. J. 1999. Macro-social influences: The effects of prices and tobacco-control policies on the demand for tobacco products. *Nicotine & Tobacco Research,* 1**,** S77-S81.

CHALOUPKA, F. J., HU, T.-W., WARNER, K. E., JACOBS, R. & YUREKLI, A. 2000. Chapter 10 Taxation of tobacco products. *In:* CHALOUPKA, F. J., JHA, P. & HEALTH, W. C. O. T. A. (eds.) *Tobacco control in developing countries.* Oxford: Oxford University Press.

CHALOUPKA, F. J. & WARNER, K. E. 2000. Chapter 29 The economics of smoking. *In:* ANTHONY, J. C. & JOSEPH, P. N. (eds.) *Handbook of Health Economics.* Elsevier.

CHAMBERLAIN, G. 1984. Chapter 22 Panel data. *In:* ZVI, G. & MICHAEL, D. I. (eds.) *Handbook of Econometrics.* Elsevier.

CHAZE, J.-P. 2005. Assessing household health expenditure with Box–Cox censoring models. *Health Economics,* 14**,** 893-907.

CNOSSEN, S. 2005. Economics and politics of excise taxation. *In:* CNOSSEN, S. (ed.) *Theory and Practice of Excise Taxation - Smoking, Drinking, Gambling, Polluting, and Driving. .* Oxforf University Press.

CNOSSEN, S. & SMART, M. 2005. Taxation of Tobacco. *In:* CNOSSEN, S. (ed.) *Theory and Practice of Excise Taxation.* Oxford University Press.

COADY, D. 2005. Indirect tax and public pricing reforms. *In:* COUDOUEL, A. P., STEFANO (ed.) *Analyzing the distributional impact of reforms* World Bank.

CONEVAL. 2008. *Programa de Calculo de la Medicion de la Pobreza - Stata Version* [Online]. Mexic City. Available: http://medusa.coneval.gob.mx/cmsconeval/rw/resource/STATA.zip?download=true [Accessed March 24 2011].

CORNWELL, A. & CREEDY, J. 1997. Measuring the Welfare Effects of Tax Changes Using the LES: An Application to a Carbon Tax. *Empirical Economics,* 22**,** 589-613.

CRAGG, J. G. 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica,* 39**,** 829-844.

CREEDY, J. 1998. Measuring the Welfare Effects of Price Changes: A Convenient Parametric Approach. *Australian Economic Papers,* 37**,** 137-151.

CREEDY, J. & DIXON, R. 1999. The Distributional Effects of Monopoly. *Australian Economic Papers,* 38**,** 223-237.

CREEDY, J. & FOSTER, J. 2002. TaxLab. 1.2.9 ed.: Department of Economics, University of Melbourne.

DEATON, A. & IRISH, M. 1984. Statistical models for zero expenditures in household budgets. *Journal of Public Economics,* 23**,** 59-80.

DEB, P., MANNING, W. G. & NORTON, E. 2010. Sample programs for ASHE June 2010 at Cornell University Ithaca.

DECOSTER, A. 2005. How progressive are indirect taxes in Russia? *The Economics of Transition,* 13**,** 705-729.

DONG, D., CHUNG, C. & KAISER, H. M. 2004. Modelling milk purchasing behaviour with a panel data double-hurdle model. *Applied Economics,* 36**,** 769 - 779.

DONG, D. & KAISER, H. M. 2008. Studying household purchasing and nonpurchasing behaviour for a frequently consumed commodity: two models. *Applied Economics,* 40**,** 1941-1951.

DUCLOS, J.-Y. & ARAAR, A. 2006. *Poverty and Equity Measurement, Policy and Estimation with DAD* New York, Springer & International Development Research Centre.

EBDON, C. 2005. Tax systems and structures. *In:* ROBBINS, D. (ed.) *Handbook of Public Sector Economics.* CRC Press, Taylor and Francis Group.

FENNEMA, J. A. & SINNING, M. 2007. Estimating double-hurdle models with dependent errors and heteroskedasticity. *German Stata Users' Group Meetings 2007.* Stata Users Group.

FRANCO-MARINA, F. & LAZCANO-PONCE, E. 2010. Tendencias del tabaquismo en adultos en México entre 1988 y 2008. *Salud Pública de México,* 52**,** S108-S119.

FREY, B. S. 2005. Excise taxes: economics, politics, and psychology. *In:* CNOSSEN, S. (ed.) *Theory and Practice of Excise Taxation - Smoking, Drinking, Gambling, Polluting, and Driving.* Oxforf University Press.

FRISCH, R. 1959. A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with Many Sectors. *Econometrica,* 27**,** 177-196.

GALLET, C. A. & LIST, J. A. 2003. Cigarette demand: a meta-analysis of elasticities. *Health Economics,* 12**,** 821-835.

GARCÍA, J. & LABEAGA, J. M. 1996. ALTERNATIVE APPROACHES TO MODELLING ZERO EXPENDITURE: AN APPLICATION TO SPANISH DEMAND FOR TOBACCO*. *Oxford Bulletin of Economics and Statistics,* 58**,** 489-506.

GENANEW, B. W. & ALEMU, M. 2012. Investments in Land Conservation in the Ethiopian Highlands: A Household Plot-level Analysis of the Roles of Poverty, Tenure Security, and Market Incentives. *International Journal of Economics and Finance,* Vol 4**,** 32-50.

GUTIERREZ, R. G., LINHART, J. M. & PITBLADO, J. S. 2003. From the help desk: Local polynomial regression and Stata plugins. *Stata Journal,* 3**,** 412-419.

HAUGHTON, J. & KHANDKER, S. R. 2009. *Handbook on Poverty and Inequality,* Washington, D.C., World Bank.

HERNÁNDEZ-ÁVILA, M., RODRÍGUEZ-AJENJO, C. J., GARCÍA-HANDAL, K. M., IBÁÑEZ-HERNÁNDEZ, N. A. & MARTÍNEZ-RUIZ, M. J. 2007. Perspectivas para el control del tabaquismo en México: reflexiones sobre las políticas actuales y acciones futuras. *Salud Pública de México,* 49**,** s302-s311.

IFS 2009. *Dimensions of tax design : the Mirrlees Review,* Oxford, Oxford University Press.

INEGI 2004. Sample Design - Description of MxFLS Baseline Sample. Instituto Nacional de Estadística Geografía e Informática, CIDE & UIA.

INEGI 2009. *La Industria Automotriz in Mexico: Series Estadisticas Sectoriales,* Mexico City, INEGI.

INITIATIVE, W. H. O. T. F. 2004. *Building Blocks for Tobacco Control - A Handbook Tools for Advancing Tobacco Control in the 21st Century* [Online]. Geneva: World Health Organization. Available: http://public.eblib.com/EBLPublic/PublicView.do?ptiID=284680.

JHA, P. & CHALOUPKA, F. J. 2000. The economics of global tobacco control. *BMJ,* 321**,** 358-361.

JHA, P., CHALOUPKA, F. J. & YUREKLI, A. 2003. The rationale for government intervention in the tobacco market. *In:* MUSGROVE, P. (ed.) *Health economics in development.* Washington, DC: World Bank.

JIMÉNEZ-RUIZ, J. A., DE MIERA, B. S., REYNALES-SHIGEMATSU, L. M., WATERS, H. R. & HERNÁNDEZ-ÁVILA, M. 2008. The impact of taxation on tobacco consumption in Mexico. *Tobacco Control,* 17**,** 105-110.

JONES, A. M. 1989. A Double-Hurdle Model of Cigarette Consumption. *Journal of Applied Econometrics,* 4**,** 23-39.

JONES, A. M. 2007. *Applied econometrics for health economists : a practical guide,* Abingdon, Radcliffe.

JONES, A. M. & LABEAGA, J. M. 2003. Individual heterogeneity and censoring in panel data estimates of tobacco expenditure. *Journal of Applied Econometrics,* 18**,** 157-177.

JONES, A. M. & YEN, S. T. 2000. A Box–Cox Double-hurdle Model. *The Manchester School,* 68**,** 203-221.

KESSELMAN, J. R. & CHEUNG, R. 2004. Tax incidence, progressivity, and inequality in Canada. *Canadian Tax Journal / Revue Fiscale Canadienne,* 52**,** 709-789.

KING, M. A. 1983. Welfare analysis of tax reforms using household data. *Journal of Public Economics,* 21**,** 183-214.

KURI-MORALES, P. A., CORTÉS-RAMÍREZ, M. & CRAVIOTO-QUINTANA, P. 2005. Prevalencia y factores asociados a la venta de cigarros a menores de edad en el Distrito Federal. *Salud Pública de México,* 47**,** 402-412.

KURI-MORALES, P. A., CORTÉS-RAMÍREZ, M., CRAVIOTO-QUINTANA, P., JESÚS HOY, M. & GONZÁLEZ-ROLDÁN, J. F. 2006. Políticas, vigilancia y acceso de los menores de edad a los cigarros en la Ciudad de México. *Salud Pública de México,* 48**,** s99-s106.

KURI-MORALES, P. A., CRAVIOTO, P., HOY, M. J. & TAPIA-CONYER, R. 2000. Assessment of cigarette sales to minors in Mexico. *Tobacco Control,* 9**,** 436-437.

LABEAGA, J. M. 1993. Individual behaviour and tobacco consumption: A panel data approach. *Health Economics,* 2**,** 103-112.

LABEAGA, J. M. 1999. A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics,* 93**,** 49-72.

LIGHTWOOD, J., COLLINS, D., LAPSLEY, H. & NOVOTNY, T. E. 2000. Chapter 4 Estimating the costs of tobacco use. *In:* JHA, P. & CHALOUPKA, F. J. (eds.) *Tobacco Control in Developing Countries.* Oxford: Oxford University Press.

LONG, J. S. 1997. *Regression models for categorical and limited dependent variables,* Thousand Oaks, Sage Publications.

MADDEN, D. 2008. Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics,* 27**,** 300-307.

MANNING, W. G. 2006. Dealing with skewed data on costs and expenditure. *In:* JONES, A. (ed.) *The Elgar companion to health economics.* Cheltenham [u.a.]: Elgar.

MANNING, W. G. & MULLAHY, J. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics,* 20**,** 461-494.

MAO, Z., SUNG, H.-Y., HU, T.-W. & YANG, G. 2008. The Demand for Cigarettes in China. *In:* HU, T.-W. (ed.) *Tobacco Control Policy in China: economics and health.* New Jersey: World Scientific.

MIERA-JUÁREZ, B. S. D., JIMÉNEZ-RUIZ, J. A., REYNALES-SHIGEMATSU, L. M., LAZCANO-PONCE, E. & HERNÁNDEZ-ÁVILA, M. 2007. El consumo de tabaco en los hogares mexicanos, 1994-2005. *Salud Pública de México,* 49**,** s263-s269.

MILES, D. 2000. The Probability That a Smoker Does Not Purchase Tobacco: A Note. *Oxford Bulletin of Economics and Statistics,* 62**,** 647-656.

MILLER, G. J. 2005. Government Fiscal Policy Impacts. *In:* ROBBINS, D. (ed.) *Handbook of Public Sector Economics.* New York, USA: CRC Press, Taylor and Francis Group.

MOFFATT, P. G. 2005. Hurdle Models of Loan Default. *The Journal of the Operational Research Society,* 56**,** 1063-1071.

MULLAHY, J. 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics,* 17**,** 247-281.

MUNDLAK, Y. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica,* 46**,** 69-85.

OLIVERA-CHÁVEZ, R. I., CERMEÑO-BAZÁN, R., MIERA-JUÁREZ, B. S. D., JIMÉNEZ-RUIZ, J. A. & REYNALES-SHIGEMATSU, L. M. 2010. El efecto del precio del tabaco sobre el consumo: un análisis de datos agregados para México. *Salud Pública de México,* 52**,** S197-S205.

PECK, R. M. 2004. Tool 6: Equity Issues, Tobacco, and the Poor. *In:* YUREKLI, A. & BEYER, J. D. (eds.) *Economics of Tobacco Toolkit.* Washington DC: World Bank.

PISINGER, C. & GODTFREDSEN, N. S. 2007. Is There a Health Benefit of Reduced Tobacco Consumption? A Systematic Review. *Nicotine & Tobacco Research,* 9**,** 631-646.

PUDNEY, S. 1989. *Modelling individual choice : the econometrics of corners, kinks and holes,* Oxford [u.a.], Blackwell.

RAMSEY, F. P. 1927. A Contribution to the Theory of Taxation. *The Economic Journal,* 37**,** 47-61.

RAPTOU, E., MATTAS, K., TSAKIRIDOU, E. & KATRAKILIDIS, C. 2005. Factors Affecting Cigarette Demand. *International Advances in Economic Research,* 11**,** 275-290.

REMLER, D. K. 2004. Poor Smokers, Poor Quitters, and Cigarette Tax Regressivity. *American Journal of Public Health,* 94**,** 225-229.

REYNALES-SHIGEMATSU, L. M., CAMPUZANO-RINCÓN, J. C., SESMA-VÁSQUEZ, S., JUÁREZ-MÁRQUEZ, S. A., VALDÉS-SALGADO, R., LAZCANO-PONCE, E. & HERNÁNDEZ-ÁVILA, M. 2006. Costs of Medical Care for Acute Myocardial Infarction Attributable to Tobacco Consumption. *Archives of medical research,* 37**,** 871-879.

REYNALES-SHIGEMATSU, L. M., RODRÍGUEZ-BOLAÑOS, R. D. L. Á., JIMÉNEZ, J. A., JUÁREZ-MÁRQUEZ, S. A., CASTRO-RÍOS, A. & HERNÁNDEZ-ÁVILA, M. 2006. Costos de la atención médica atribuibles al consumo de tabaco en el Instituto Mexicano del Seguro Social. *Salud Pública de México,* 48**,** s48-s64.

ROSEN, H. 2001. *Public Finance,* Boston, McGraw-Hill/Irwin.

RUBALCAVA, L. & TERUEL, G. 2006. User's Guide for the Mexican Family Life Survey First Wave.

RUBALCAVA, L. & TERUEL, G. 2008. User's Guide for the Mexican Family Life Survey Second Wave.

SAENZ DE MIERA, B., JIMÉNEZ RUIZ, J. A. & REYNALES SHIGEMATSU, L. M. 2007. La economia del tabaco en Mexico. Instituto Nacional de Salud Public.

SAENZ-DE-MIERA, B., THRASHER, J. F., CHALOUPKA, F. J., WATERS, H. R., HERNANDEZ-AVILA, M. & FONG, G. T. 2010. Self-reported price of cigarettes, consumption and compensatory behaviours in a cohort of Mexican smokers before and after a cigarette tax increase. *Tobacco Control,* 19**,** 481-487.

SAHN, D. E. & YOUNGER, S. D. 2003. Estimating the Incidence of Indirect Taxes in Developing Countries. *In:* BOURGUIGNON, F. P. D. S., LUIZ A. (ed.) *The impact of economic policies on poverty and income distribution: evaluation techniques and tools.* World Bank.

SESMA-VÁZQUEZ, S., CAMPUZANO-RINCÓN, J. C., CARREÓN-RODRÍGUEZ, V. G., KNAUL, F., LÓPEZ-ANTUÑANO, F. J. & HERNÁNDEZ-AVILA, M. 2002. El comportamiento de la demanda de tabaco en México: 1992-1998. *Salud Pública de México,* 44**,** s82-s92.

SHCP 2000-2010. Distribucion del Pago de Impuestos y Recepcion del Gasto Publico. Ciudad de Mexico: Secretaria de Hacienda y Credito Publico.

SHEU, M.-L., HU, T.-W., KEELER, T. E., ONG, M. & SUNG, H.-Y. 2004. The effect of a major cigarette price change on smoking behavior in california: a zero-inflated negative binomial model. *Health Economics,* 13**,** 781-791.

SMITH, M. 2001. Specification of Non-Normal Double-Hurdle Models. *International Congress on Modelling and Simulation.* The Australian National University.

SMITH, M. 2002. On specifying double-hurdle models. *In:* ULLAH, A., WAN, A. T. K. & CHATURVEDI, A. (eds.) *Handbook Of Applied Econometrics And Statistical Inference.* New York: Marcel Dekker.

SU, S.-J. B. & YEN, S. T. 2000. A censored system of cigarette and alcohol consumption. *Applied Economics,* 32**,** 729-737.

TAURAS, J., POWELL, L., CHALOUPKA, F. & ROSS, H. 2007. The demand for smokeless tobacco among male high school students in the United States: the impact of taxes, prices and policies. *Applied Economics,* 39**,** 31-41.

TAURAS, J. A. 2005. An Empirical Analysis of Adult Cigarette Demand. *Eastern Economic Journal,* 31**,** 361-375.

THRASHER, J., VILLALOBOS, V., BARNOYA, J., SANSORES, R. & O'CONNOR, R. 2011. Consumption of single cigarettes and quitting behavior: A longitudinal analysis of Mexican smokers. *BMC Public Health,* 11**,** 134.

THRASHER, J. F., VILLALOBOS, V., DORANTES-ALONSO, A., ARILLO-SANTILLÁN, E., CUMMINGS, K. M., O'CONNOR, R. & FONG, G. T. 2009. Does the availability of single cigarettes promote or inhibit cigarette consumption? Perceptions, prevalence and correlates of single cigarette use among adult Mexican smokers. *Tobacco Control,* 18**,** 431-437.

TORELLI, N. & TRIVELLATO, U. 1993. Modelling inaccuracies in job-search duration data. *Journal of Econometrics,* 59**,** 187-211.

TRAIN, K. E. 2003. *Discrete choice methods with simulation,* Cambridge [u.a.], Cambridge Univ. Press.

VALDÉS-SALGADO, R., LAZCANO-PONCE, E. C. & HERNÁNDEZ-AVILA, M. 2005. Current panorama of tobacco consumption and control measures in Mexico. *Prevention and Control,* 1**,** 319-327.

VALDÉS-SALGADO, R., REYNALES-SHIGEMATSU, L. M., LAZCANO-PONCE, E. & HERNÁNDEZ-ÁVILA, M. 2007. Antes y después del Convenio Marco en México: una comparación desde la Encuesta sobre Tabaquismo en Jóvenes 2003 y 2006. *Salud Pública de México,* 49**,** s155-s169.

VALDÉS-SALGADO, R., THRASHER, J., SÁNCHEZ-ZAMORANO, L. M., LAZCANO-PONCE, E., REYNALES-SHIGEMATSU, L. M., MENESES-GONZÁLEZ, F. & HERNÁNDEZ-ÁVILA, M. 2006. Los retos del convenio marco para el control del tabaco en México: un diagnóstico a partir de la Encuesta sobre Tabaquismo en Jóvenes. *Salud Pública de México,* 48**,** s5-s16.

VÁZQUEZ-SEGOVIA, L. A., SESMA-VÁZQUEZ, S. & HERNÁNDEZ-AVILA, M. 2002. El consumo de tabaco en los hogares en México: resultados de la Encuesta de Ingresos y Gastos de los Hogares,1984-2000. *Salud Pública de México,* 44**,** s76-s81.

VERBEEK, M. 2008. *A guide to modern econometrics,* Chichester, John Wiley & Sons.

WANG, H. & HEITJAN, D. F. 2008. Modeling heaping in self-reported cigarette counts. *Statistics in Medicine,* 27**,** 3789-3804.

WARNER, K. E. 1978. Possible Increases in the Underreporting of Cigarette Consumption. *Journal of the American Statistical Association,* 73**,** 314-318.

WARNER, K. E. 2000. The economics of tobacco: myths and realities. *Tobacco Control,* 9**,** 78-89.

WARNER, K. E., CHALOUPKA, F. J., COOK, P. J., MANNING, W. G., NEWHOUSE, J. P., NOVOTNY, T. E., SCHELLING, T. C. & TOWNSEND, J. 1995. Criteria for determining an optimal cigarette tax: the economist's perspective. *Tobacco Control,* 4**,** 380.

WATERS, H. R., SAENZ-DE-MIERA, B., ROSS, H. & REYNALES SHIGEMATSU, L. M. 2010. La Economia del Tabaco y los Impuestos al Tabaco en Mexico. *Union Internacional contra la Tuberculosis y Enfermedades Respiratorias.* Paris.

WHO 2003. WHO Framework Convention on Tobacco Control. Geneva, Switzerland: WHO Document Production Services.

WHO 2011. Systematic review of the link between tobacco and poverty Geneva: World Health Organization.

WILKINS, N., YUREKLI, A. & HU, T.-W. 2004. Tool 3. Demand Analysis - Economic Analysis of Tobacco Demand. *In:* YUREKLI, A. & BEYER, J. D. (eds.) *Economics of Tobacco Toolkit.* Washington DC: World Bank.

WOOLDRIDGE, J. M. 2002. *Econometric analysis of cross section and panel data,* Cambridge, Mass. [u.a.], MIT Press.

YEN, S. T. 1993. Working Wives and Food Away from Home: The Box-Cox Double Hurdle Model. *American Journal of Agricultural Economics,* 75**,** 884-895.

YEN, S. T. 1999. Gaussian versus count-data hurdle models: cigarette consumption by women in the US. *Applied Economics Letters,* 6**,** 73-76.

YEN, S. T. 2005a. Zero observations and gender differences in cigarette consumption. *Applied Economics,* 37**,** 1839 - 1849.

YEN, S. T. 2005b. A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations. *American Journal of Agricultural Economics,* 87**,** 453-466.

YEN, S. T. & HUANG, C. L. 1996. Household Demand for Finfish: A Generalized Double-Hurdle Model. *Journal of Agricultural and Resource Economics,* 21**,** 220-234.

YEN, S. T. & JONES, A. M. 1996. Individual cigarette consumption and addiction: A flexible limited dependent variable approach. *Health Economics,* 5**,** 105-117.

YUREKLI, A. 2004. Tool 4: Design and Administer Tobacco Taxes. *In:* YUREKLI, A. & BEYER, J. D. (eds.) *Economics of Tobacco Toolkit.* Washington DC: World Back.

Appendix

Appendices to Chapter 1

Relevance of excises in public finances

In 2008, total federal revenue reached 23.4% of the country's GDP of which 63.1% comes from taxes or from "non-oil origins". The distinction between oil and non-oil origins is important given the reliance of Mexico on oil sales to complete the government's budget[1]. Federal revenue from the main taxes (income tax, VAT, excises and taxes on imports) accounts for 9.1% of the country's GDP.

Revenue derived from taxing alcohol and tobacco accounts for a small proportion of the total non-oil federal revenue. On average, between 2000 and 2008 this proportion has been around 3.5% and has remained fairly constant throughout this period (See Figure 1). In 2008, the Mexican Government raised 49,283.9 MXN millions (nominal) for IEPS (excluding motor fuel), from which taxing tobacco represents 50.2%, alcohol 46.7% and others 3.1%. Nevertheless, the importance of the IEPS in the Mexican tax system remains modest if compared with the revenue coming from income tax or the VAT which are the main source of non-oil fiscal revenue (See Figure 2).

Fuel

Fuel is also in the list of goods that are subject to an excise tax. Correctly speaking though, it is not a tax; rather it is the difference between the price that consumer pays in the petrol station (which includes the commission that a distributor receives) and the producer price. In Mexico, the State-owned oil company *Petróleos Mexicanos* (PEMEX) is the sole producer whose responsibility is to extract and refine oil, as well as gas, and distribute all the varieties of fuel needed for national consumption. Its objective is to meet the national demand at low prices, a policy that has been followed since its creation in 1938. Its infrastructure capacity to refine oil, however, is not sufficient to meet the national demand, so that PEMEX imports the quantity of fuel which it is not able to refine. For example, in 2008 almost 43% of internal consumption on petrol and 18% of internal consumption on diesel was imported (INEGI 2009). As a consequence, the producer price of fuel varies according with the fluctuations of the spot oil-price which is traded internationally, while the consumer price set by the government is pre-adjusted and therefore remains constant. PEMEX has the responsibility to report the IEPS revenue from petrol and diesel to fiscal authorities; Figure 3 shows the evolution of revenue in the last nine years.

[1] This distinction can also be thought as "fiscal" and "non-fiscal" origin.

Making the price dependable on production costs attracted the public debate during most of 2002, especially in areas around the border with the United States[2]. From 2001 through to 2003 the price of oil remained unchanged which benefited the public finances, but by the end of 2005, the price of oil began to increase and reached historic highs, making PEMEX incur losses which had to be compensated with revenue gained from other sources. As a result, it is considered that from 2006, consumers benefited from low fuel prices as a consequence of a subsidy being applied in the final price. Foreseeing the negative impact on the public finances, motor fuel suffered periodically price increases from 2006. Even though the highest increases on fuel-related products occurred in 2007 and 2008, these increases do not stop the government providing 217,609.1 MXN millions (nominal) in subsidising motor fuel for consumers in 2008.

Structure of excises in Mexico

Alcohol

The structure of excises on alcohol consists in levying at a rate which increases as the lelvel of alcohol increases. For the 2008 tax schedule, drinks with alcohol content up to 14° G.L. are taxed at 25%. Drinks with alcohol content of more than 14° G.L. and up to 20° G.L. are taxed at 30% and drinks with alcohol content of more than 20° G.L. are taxed at 50%. Beer, as a particular drink, which is specifically defined in the IEPS code, is subject to a 25% excise irrespective of alcohol content. This levy is paid by the producer and apart from beer, bottles must wear special labels issued by fiscal authorities to provide the consumer with a sign that the product has met with its fiscal obligation, and hence it is commonly seen as a sign of authenticity. Lastly, the level and structure of excise tax on alcohol has remained practically unchanged in recent years, the most significant reform was implemented in 2002 where it was decided to move from imposing a specific tax or unit tax to an *ad valorem* tax.

[2] Specific areas neighbouring the border with the U.S. have been subject to a special policy in order to prevent cross-border sales, but the policy is not always automatic.

Derivation of equation (15)

The demand function that derives from the LES can be expressed as (where the subscripts have been dropped for convenience):

$$x(p,y) = \gamma + \frac{\beta}{p}\left(y - \sum_{j=1}^{n} p\gamma\right) \quad (1)$$

This can be rewritten as:

$$px = p\gamma + \beta\left(y - \sum_{j} p\gamma\right) \quad (2)$$

Taking natural logarithm in both sides of the equation and differentiating with respect to y yields:

$$e_i = \frac{y\beta}{p\gamma + \beta\left(y - \sum_j p\gamma\right)} = \frac{y\beta}{px} \quad (3)$$

Taking once more the natural logarithms in both sides of the equation but differentiating with respect to p yields:

$$1 + e_{ii} = \frac{p\gamma(1-\beta)}{p\gamma + \beta\left(y - \sum_j p\gamma\right)} = \frac{p\gamma(1-\beta)}{px} \quad (4)$$

After some re-arrangements this yield:

$$p\gamma = \frac{px(1+e_{ii})}{1-\beta} = \frac{yw(1+e_{ii})}{1-\beta} \quad (5)$$

which is the expression shown as equation (15).

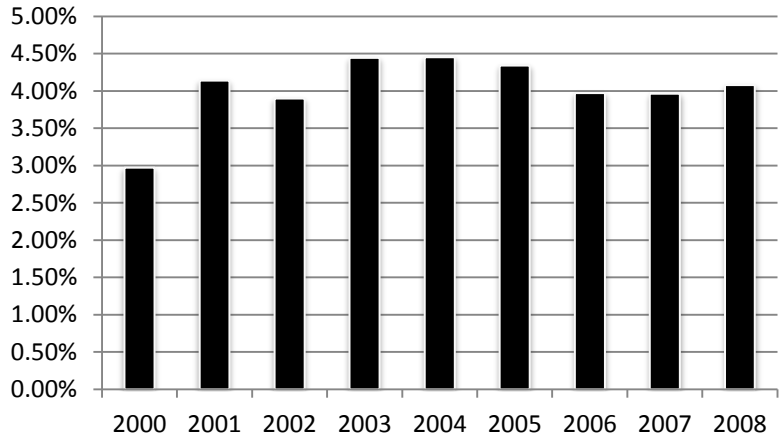


Figure A1 IEPS as a Proportion of Total Non-Oil Fiscal Revenue, Mexico 2000-2008.
Note: (Excluding motor fuel). Data source: CEFP (2011).

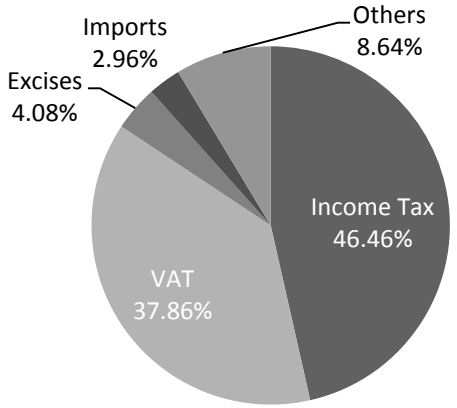


Figure A2 Total Non-Oil Fiscal Revenue and its Components, Mexico 2008.
Data source: CEFP (2011).

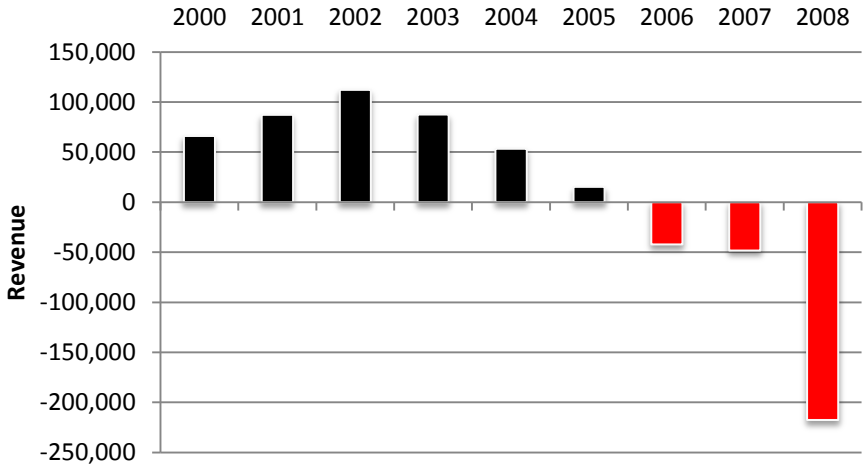


Figure A3 IEPS Oil Fiscal Revenue, Mexico 2000-2008.
Note: Revenue in millions of MXN in nominal terms.
Data source: CEFP (2011).

Table 1 Commodity Aggregates and their Taxable Items, ENIGH 2008.

Commodity Groups	ENIGH-2008 Code	Number of Items with taxable goods				
		Exempt	VAT 0 %	VAT 15%	Excise	Not Classified
Food	A001-A222, A242-A247	-	210	18	-	-
Alcohol	A223-A238	-	-	16	16	-
Tobacco	A239-A241	-	-	3	3	-
Rent	----------	-	-	-	-	-
Household and Serv.	G002-G022	6	2	12	-	1
Household Furniture	K001-K036	-	-	36	-	-
Household Glass	I001-I026	-	-	26	-	-
Cleaning Items	C001-C024	-	-	24	-	-
Clothing and Shoes	H001-H122, H136	-	-	123	-	-
Personal Care	D001-D026, H132	-	-	27	-	-
Health	J001-J072	13	41	18	-	-
Education and Leisure	E001-E033, H134-H135, L001-L029, N003	15	2	50	1	-
Accessories	H123-H131, H133	-	-	10	-	-
Communication	F001- F009	-	-	9	-	-
Transport	M001-M018, F010-F017	1	-	22	-	3
Public Transport	B001-B007	7	-	-	-	-
Others	N001-N002, N006-N016, T901-T914	4	1	16	-	6

Table 2 Survey's Items and their Statutory Tax Rate, ENIGH 2008.

Commodity Aggregate and its Components	ENIGH Code	VAT Rate %	Excise Tax %
Alcohol			
Cognac and Brandy	A223	15	50
Beer	A224	15	25
***Anis* (Liquor)**	A225	15	50
Sherry	A226	15	50
Fruit Liquor	A227	15	50
Aguamiel, Pulque, Tlaquiche	A228	15	50
Aguardiente, Alcohol (sugar cane),	A229	15	50
Ron (*Anejo*, White, with Lemon)	A230	15	50
Rompope	A231	15	50
Cider (White and Red)	A232	15	25
Tequila (Anejo, Blue and White)	A233	15	50
Wine (White, Red, Pink)	A234	15	50
Vodka	A235	15	50
Whisky	A236	15	50
Alcoholic Drinks	A237	15	25
Other: Champagne, etc	A238	15	50
Tobacco			
Cigarettes	A239	15	150
Cigars	A240	15	150
Rolling Tobacco	A241	15	150

Table 3 Different Welfare Indicators by per Capita Income Quintiles, Mexico 2008.

	Household					Per Capita			
Quintile	No. Households	Gross Income	Net Income	Gross Expenditure	Net Expenditure	Gross Income	Net Income	Gross Expenditure	Net Expenditure
Lowest	5,859	3,728.20	3,187.43	4,612.15	4,071.38	809.76	652.33	1,027.20	869.76
Lower-Mid	6,023	6,407.63	5,771.17	6,369.25	5,732.78	1,538.55	1,328.68	1,564.30	1,354.44
Middle	5,866	9,288.29	8,630.70	7,974.54	7,316.95	2,353.02	2,116.30	2,110.47	1,873.75
Mid-Upper	5,767	13,100.79	12,282.13	10,383.83	9,565.17	3,745.57	3,439.55	3,096.66	2,790.64
Highest	5,953	28,660.50	27,419.42	19,083.93	17,842.84	10,926.26	10,416.09	7,410.80	6,900.63
Overall	29,468	12,236.85	11,457.95	9,684.60	8,905.69	3,874.56	3,590.52	3,041.83	2,757.80

Table 4 Different Welfare Indicators by per Capita Expenditure Quintiles, Mexico 2008.

	Household					Per Capita			
Quintile	No. of Obs.	Gross Income	Net Income	Gross Expenditure	Net Expenditure	Gross Income	Net Income	Gross Expenditure	Net Expenditure
Lowest	6,187	5,338.69	4,792.65	3,961.17	3,415.13	1,107.68	946.98	833.58	672.88
Lower-Mid	5,829	7,553.37	6,955.87	6,057.77	5,460.27	1,736.42	1,548.30	1,419.46	1,231.34
Middle	5,850	9,590.07	8,852.29	7,946.64	7,208.86	2,469.09	2,215.63	2,062.12	1,808.65
Mid-Upper	5,692	12,846.72	12,160.07	10,289.90	9,603.24	3,761.16	3,499.01	3,022.94	2,760.79
Highest	5,910	25,858.84	24,532.11	20,170.30	18,843.57	10,299.99	9,744.16	7,872.27	7,316.44
Overall	29,468	12,236.85	11,457.95	9,684.60	8,905.69	3,874.56	3,590.52	3,041.83	2,757.80

Data Source: ENIGH, 2008.

Table 5 Frisch Parameters for Each Total Expenditure Group.

Expenditure Group	Midpoint Expenditure	Frisch Parameter
1	1,110.48	-28.163
2	1,817.62	-21.703
3	2,305.91	-17.899
4	2,723.82	-15.153
5	3,083.56	-13.087
6	3,414.23	-11.481
7	3,742.50	-10.200
8	4,065.39	-9.156
9	4,374.69	-8.292
10	4,680.94	-7.565
11	5,002.86	-6.946
12	5,334.34	-6.413
13	5,658.13	-5.950
14	6,013.00	-5.544
15	6,382.55	-5.186
16	6,767.53	-4.867
17	7,166.25	-4.583
18	7,589.11	-4.327
19	8,057.73	-4.096
20	8,605.36	-3.887
21	9,182.19	-3.696
22	9,815.74	-3.521
23	10,569.70	-3.361
24	11,457.90	-3.214
25	12,542.00	-2.937
26	14,049.30	-2.625
27	16,025.20	-2.399
28	18,749.20	-2.206
29	23,723.80	-1.900
30	43,202.30	-1.900

Appendices to Chapter 2

The calculation of living standards in the Mexican Family Life Survey

A measurement of individual welfare is fundamental in empirical work given that it defines the individual's living standards in comparison with others in the population. For example, measurements of poverty and/or equality are constructed based on specific welfare indicators, and their accuracy depends largely on the kind of information available. Measurements of well-being can take many forms, such as health status, life expectancy, infant mortality rates, housing conditions, schooling and the like. Usually however, a monetary measure to value household welfare is used instead, such as income or consumption expenditure {Haughton, 2009 #36}. In line with international experiences, this analysis uses the level of total consumption expenditure as a leading indicator of welfare. {Deaton, 1997 #8} argues that for the case of developing countries (as is Mexico) living standards are better measured if consumption expenditure is used because it is less volatile than its counterpart, income.

The ultimate purpose is to obtain a measurement of welfare at individual level. It was found that even though information of income is collected at individual level, there are some individuals likely to end up with zero income given their natural circumstances in life at the time of the survey. Such is the case of students or members of the family (usually adult female) whose main activity is to take care of the house and children and from whom no monetary income as such is reported. Having recorded zero income is not entirely consistent for this study given that some individuals are current smokers who spend money on cigarettes, and their consumption is recorded as either the number of cigarettes smoked per week and/or the amount of weekly money spent on cigarettes. Therefore, total consumption expenditure could better explain the level of tobacco consumption by individuals and place them in a point on the income distribution. The methodology, which follows, is an explanation of how total consumption expenditure per capita is calculated.

Methodology

The strategy used to construct the total value of consumption expenditure per capita was first to calculate it at household level. This involved totalling-up the monetary value of all expenditures recorded by the household. One member was asked to report the monetary value of each item listed on a pre-established questionnaire provided. This questionnaire is a compact version of the one used at national level by INEGI which constrains the consumption of only those

items which are more frequently consumed by Mexican households. Table X below shows the complete list of items considered by the MxFLS. However, in the upper part of the table labelled "weekly", the item listed as "food" appears as an aggregate category covering a range of different products from grains to processed foods. The remaining items in the table correspond exactly to the complete list in the questionnaire. Ten items, from corn tortillas to soft drinks, which could have been included in the food category, are listed as in the questionnaire to show the products from which more detailed information have been collected such as quantity purchased. Moreover, Table X shows that items recorded on an annual basis correspond mostly to durable goods. A problem arises, however, in including such goods in the calculation of total consumption expenditure. One potential difficulty derives from the nature of the retrospective question. This is to say that households may not have been able to report items due to recollection difficulties or because they did not purchase any item during the specified time, yet the household may own such an item. Including values of durable goods in the computation of total consumption expenditure may wrongly place the household on a point in the income distribution that does not accurately reflect its true living standards. {Haughton, 2009 #36} provide some preliminary elements to consider if it is decided to include durable goods in the calculations. However, a straightforward solution is simply to omit values of these goods in computing household total expenditure. This is the approach taken here given its convenience and data limitations. Hence, consumption expenditure is based on items recorded on weekly, monthly and quarterly basis only. Weekly and quarterly values have been transformed into monthly values.

As previously mentioned, all the information collected corresponds to retrospective information regarding the purchasing of each item. Households had to answer the amount in pesos spent for the last seven days, past month and past three months (depending on the item in question) of current consumption. In addition, the respondent was asked to report the total value of each item that the household was given as a gift, payment, or obtained from its crops, animals, or business (auto-consumption). Likewise, the respondent was asked to report the total value of each item that the household gave as a present or payment to other people (transfers in kind).

Another step was to convert nominal values into real terms. This involved deflating the values using a price index to a common base. All values were deflated using the overall national price index provided by Banco de Mexico, the central bank[3].

Two measurements of total consumption expenditure have been calculated: total expenditure and net total expenditure. Total expenditure is the sum of monetary and non-

[3] Price Index from the Central Bank's website: http://www.banxico.org.mx/portal-inflacion/index.html
At the time of writing this paper, the Central Bank was still in charge of measuring inflation. It is now the responsibility of INEGI to measure and report this.

monetary (auto-consumption) expenditure which can be thought of “gross expenditure”. In turn, for net expenditure, “transfers” are deducted from the resulting values of total expenditure.

The final step is to derive a measurement of individual welfare. This study converts total expenditure into per capita terms. This simply involved calculating the ratio between total or net expenditure and the household size. A per capita term implicitly assumes that consumption is equally divided among all members of the household. This might be a strong assumption to make, given that as it is recognized, ‘children do not have the same needs as adults’ (Deaton, 1997 p. 150) therefore inequality and poverty measures may be understated. An alternative solution is usually proposed which involves using “equivalence scales” (or weights as they are also known) to deflate the values of real total expenditure. These equivalence scales would adjust the “per capita” value recognising the presence of economies of scale in household consumption and having families with different compositions. Unfortunately, there is not a general consensus on the precise value for these scales and therefore this study does not take this approach. In fact, Deaton (1997, p. 150) also points out that ‘assigning per capita expenditure equally to individuals is still a best practice’.

Lastly, it is worth mentioning that a small percentage of individuals from the reference sample belonged to households which did not participate in the section “Household expenditure”. This situation resulted in having 415 households (4.92% of the sample) with missing values in the first sweep. To avoid dropping these observations from the sample, total consumption per capita has been imputed. Firstly, total household consumption is imputed based on stratum, total expenditure and household size. Then, the imputed values are divided by household size to obtain a final value of consumption expenditure per capita. The imputation of values have been carried out with the `impute` command in STATA. Results of calculations are provided in Table Y & Z.

Table 1 Catalogue of items listed in the MxFLS' questionnaire.

Weekly
• Food (Vegetables and fruits, cereals and grains, animal origin food and other industrially processed food)
• Cigarettes and tobacco
• Transportation like: bus, subway, taxi, and/or gasoline
• Food and drinks consumed outside the household (breakfast, lunch, dinner)
• Corn tortillas
• Bakery or Store unpacked *bolillo/telera*
• Chicken
• *Milanesa* steak
• Pasteurized milk
• Hen eggs
• Red tomatoes
• Beans Packed or in bulk
• White sugar
• Soft drinks
Monthly
• Personal goods in general such as: toothpaste, shampoo, tissues, toilet soap, and toilet paper
• Women's personal goods as: perfume, deodorant, cosmetics, feminine towels, cold creams, haircut, dyes, manicure, depilation, etc.
• Men's personal goods as: lotion, deodorant, razors, and shaving foam, haircuts, etc.
• Cleaning products and for the house as: detergents, cleaners, light bulbs, brooms, candles, bar of soap, bleaches, glass lampshades, etc.
• Services in general as: domestic service, laundry, dry cleaner, etc.
• Culture and recreation as: books, magazines, newspapers, records, excursions, fairs, etc.
• Lottery and other chance games.
• Telephone, telegraph, money orders, postage stamps, internet, etc.
• Other home services as: water, electricity, gas, garbage collection, firewood, coal, petroleum, etc.
Quarterly
• Clothes and shoes for male adults as: pants, shirts, sweaters, underwear, etc.
• Clothes and shoes for female adults as: blouses, sweaters, skirts, underwear, etc.
• Clothes and shoes for boys as: pants, shirts, sweaters, and underwear, etc.
• Clothes and shoes for girls as: blouses, sweaters, skirts, underwear, etc.
• Toys in general, baby clothes and baby articles as: disposable or cloth diapers, clothes, day-cares, etc.
• Tableware, casseroles, pots, bedspreads, bed sheets, pillows, yarn, needles, and any other domestic utensils.
• Healthcare and services as: medicine, medical and dental visits, hospitalization, etc.
• Maintenance services for vehicles as: fuel, oil, lubricants, pension, parking, cleaning, mechanical shops, appliance, auto parts, etc.
Yearly
• Electronic appliances as: TV sets, radios, cameras, etc. gardening articles and sport articles or sport services.
• Blenders, irons, washing machines, dryers, refrigerators, and other electro domestic appliances.
• Chairs, sofas, and other furniture and /or dwelling repairs or extensions.
• Funerals, vacations, parties, insurances, moving and other transportation services, and other expenditures.
• Property or income taxes
• Vehicle acquisitions

NOTE: School items excluded.
Source: MxFLS 1 [Questionnaire] – Book 1 (Household Consumption).

Table 2 Total and Per capita Consumption Expenditure Per Month by Households (MxFLS 1 & 2).

Wave 1

Quintile	Mean	S.D	Min	p25	p50	p75	Max	*N*
				Total Expenditure				
I	2,096.00	1,093.91	15.06	1,318.40	1,995.56	2,720.44	6,957.64	1689
II	3,594.59	1,574.27	605.66	2,562.08	3,422.91	4,398.88	13,520.27	1688
III	5,016.64	2,152.56	970.21	3,537.19	4,917.01	6,229.00	16,230.92	1688
IV	6,698.61	2,830.37	1,439.46	4,584.42	6,499.09	8,704.27	28,890.05	1688
V	17,436.42	62,768.60	2,323.31	7,025.41	9,912.97	14,564.07	2,136,079.00	1688
Total	6,967.88	28,644.27	15.06	2,683.16	4,489.82	7,500.51	2,136,079.00	8441
				Total Expenditure Per Capita				
I	387.86	133.58	15.06	292.39	397.97	502.16	595.62	1689
II	768.98	104.08	595.70	678.50	765.33	857.16	958.71	1688
III	1,181.82	139.65	959.89	1,059.65	1,174.51	1,305.25	1,435.31	1688
IV	1,809.33	243.98	1,435.35	1,579.30	1,795.76	2,008.44	2,304.19	1688
V	5,964.18	14,003.18	2,306.12	2,736.34	3,376.95	5,040.90	305,154.20	1688
Total	2,022.24	6,582.01	15.06	678.45	1,174.51	2,008.34	305,154.20	8441

Wave 2

Quintile	Mean	S.D	Min	p25	p50	p75	Max	*N*
				Total Expenditure				
I	1,988.18	1,081.22	5.10	1,195.19	1,847.93	2,584.03	7,589.37	1688
II	3,499.43	1,560.59	526.09	2,412.26	3,292.31	4,269.08	12,294.14	1687
III	4,776.56	2,177.86	864.20	3,368.79	4,575.76	5,884.10	17,386.14	1688
IV	6,339.14	2,797.07	1,320.54	4,133.63	6,124.24	7,983.35	20,017.15	1687
V	13,325.03	54,048.61	2,119.30	6,497.70	9,276.98	13,885.97	2,105,780.00	1687
Total	5,985.05	24,547.72	5.10	2,544.05	4,263.09	6,974.18	2,105,780.00	8437
				Total Expenditure Per Capita				
I	337.91	123.16	2.55	246.36	351.03	441.00	525.44	1688
II	687.96	95.67	525.60	606.71	682.31	768.36	859.65	1687
III	1,065.95	126.15	859.73	955.19	1,056.50	1,170.47	1,294.46	1688
IV	1,653.08	226.46	1,294.46	1,453.72	1,631.13	1,843.52	2,094.44	1687
V	4,231.64	10,884.71	2,094.62	2,486.86	2,967.88	4,179.35	421,155.90	1687
Total	1,595.10	5,062.05	2.55	606.71	1,056.05	1,842.51	421,155.90	8437

NOTES: Weighted figures in Mexican pesos and in real terms.
Source: Own elaboration with data from the MxFLS 1 & 2.

Definition of dependent variables

Demographic Variables

- **Age**: Continuous variable which correspond to the self-reported age of the respondent in years.

- **Male**. Dichotomous indicator equal to one for males and zero for females.

- **Indigenous:** Dichotomous indicator equal to one if the respondent belongs to an ethnic group, zero otherwise.

- **Marital Status:**

  Single: Dichotomous indicator equal to one if the respondent is single and zero otherwise-omitted.

  Married: Dichotomous indicator equal to one if the respondent is married or living in cohabitation and zero otherwise.

  Divorced: Dichotomous indicator equal to one if the respondent is divorced or separated and zero otherwise.

  Widowed: Dichotomous indicator equal to one if the respondent is widowed and zero otherwise.

- **Urban:** Dichotomous indicator equal to one if the respondent lives in stratum 1 (for households living in areas of more than 100,000 inhabitants), stratum 2 (for household living in areas with a population between 15,000 and 100,000) or in stratum 3 (for areas with a population between 2,500 inhabitants and 15,000) and equal to zero if the respondent lives in stratum 4 (for households living in areas of less than 2,500 inhabitants).

- **Region:**

  Centre – North East: Dichotomous indicator equal to one if the respondent lives in the Mexican state of Coahuila, Durango or Nuevo Leon, zero otherwise.

  Centre – West: Dichotomous indicator equal to one if the respondent lives in the Mexican state of Guanajuato, Michoacán or Jalisco, zero otherwise.

  Centre: Dichotomous indicator equal to one if the respondent lives in the Mexican state of Mexico, Morelos, Puebla or Mexico City, zero otherwise-omitted.

  North West: Dichotomous indicator equal to one if the respondent lives in the Mexican state of Baja California Sur, Sinaloa or Sonora, zero otherwise.

  South – South East: Dichotomous indicator equal to one if the respondent lives in the Mexican state of Oaxaca, Veracruz or Yucatan, zero otherwise.

Socio-economic variables

- **Income (proxy):** A continuous variable that represents the monthly consumption expenditure per capita in Mexican pesos.

- **Household owner:** Dichotomous indicator equal to one if the respondent is living in a household which is of his/her own or currently paying it, zero otherwise.

- **Education:**

  No education: Dichotomous indicator equal to one if the respondent has no attended school or his/her last level of education is kinder garden, zero otherwise-omitted.

  Primary: Dichotomous indicator equal to one if the respondent's last level of education is "primary" and zero otherwise.

  Secondary: Dichotomous indicator equal to one if the respondent's last level of education is "secondary" or "open secondary", zero otherwise.

  High School: Dichotomous indicator equal to one if the respondent's last level of education is "high school" or "open high school", zero otherwise.

  University: Dichotomous indicator equal to one if the respondent's last level of education is "*normal superior*" (teaching training), "college" or "postgraduate", zero otherwise.

- **Activity last week:**

  House maker: Dichotomous indicator equal to one if the respondent answered being the housemaster or housewife the previous week to the survey, zero otherwise.

- **Health insurance:** Dichotomous indicator equal to one if the respondent benefits from health services provided either by a private insurer or public insurer such as IMSS, ISSSTE, PEMEX, SEDENA or Marina, zero otherwise.

Psycho-social factors

- **Exercise:** Dichotomous indicator equal to one if the respondent engages as a routine in physical exercise, zero otherwise.

- **Other smoker:** Dichotomous indicator equal to one if the respondent lives with an active smoker in the household or someone who had the habit of smoking, zero otherwise.

- **Hot drinks:** Dichotomous indicator equal to one if the respondent consumes hot drink with food or drinks it in parties, gatherings or *fiestas*, zero otherwise.

- **Alcohol:** Dichotomous indicator equal to one if the respondent consumes alcoholic drinks such as beer, Tequila, *pulque* or any fermented juice of the maguey with food, as well as in parties, gathering or fiestas, zero otherwise.

Table 3 Sample Statistics, All Adults, MxFLS-1.

	All		Non-Smokers		Smokers	
Variable	MEAN	SD	MEAN	SD	MEAN	SD
Continuous						
Cigarettes	4.78	21.507	0.00	0.000	40.58	49.715
Age	37.64	17.133	37.59	17.326	38.04	15.610
Ln(Tot. Exp. PC)	6.48	1.179	6.45	1.161	6.65	1.296
Categorical						
Male	0.445	0.497	0.405	0.491	0.742	0.438
Indigenous	0.160	0.367	0.167	0.373	0.108	0.310
Urban	0.576	0.494	0.565	0.496	0.657	0.475
Single	0.294	0.456	0.298	0.457	0.263	0.440
Married/Cohabitation	0.612	0.487	0.609	0.488	0.639	0.480
Divorced/Separated	0.047	0.211	0.044	0.206	0.064	0.244
Widowed	0.047	0.212	0.049	0.216	0.034	0.182
No education	0.089	0.285	0.091	0.287	0.078	0.267
Primary	0.412	0.492	0.416	0.493	0.377	0.485
Secondary	0.262	0.440	0.258	0.438	0.293	0.455
High-School	0.141	0.348	0.140	0.347	0.146	0.353
University	0.096	0.295	0.095	0.293	0.107	0.309
South - East	0.204	0.403	0.214	0.410	0.126	0.332
Centre	0.193	0.395	0.191	0.393	0.205	0.404
West	0.198	0.398	0.198	0.398	0.197	0.397
North - West	0.212	0.408	0.207	0.405	0.248	0.432
North - East	0.194	0.395	0.190	0.392	0.224	0.417
House maker	0.322	0.467	0.348	0.476	0.129	0.336
Homeowner	0.729	0.444	0.732	0.443	0.705	0.456
Health insurance	0.442	0.497	0.437	0.496	0.478	0.500
Alcohol	0.381	0.486	0.343	0.475	0.667	0.471
Hot drinks	0.003	0.054	0.003	0.057	0.001	0.029
Exercise	0.196	0.397	0.192	0.394	0.227	0.419
Other smoker	0.291	0.454	0.237	0.425	0.692	0.462
N	**19804**		**17469**		**2335**	

Table 4 Sample Statistics, All Adults, MxFLS-2.

	All		Non-Smokers		Smokers	
	MEAN	SD	MEAN	SD	MEAN	SD
Continuous						
Cigarettes	4.00	20.446	0.00	0.000	41.93	52.825
Age	38.45	17.801	38.48	17.986	38.16	15.943
Ln(Tot. Exp. PC)	6.33	1.102	6.31	1.098	6.54	1.113
Categorical						
Male	0.439	0.496	0.406	0.491	0.757	0.429
Indigenous	0.160	0.366	0.166	0.372	0.103	0.304
Urban	0.593	0.491	0.581	0.493	0.710	0.454
Single	0.301	0.459	0.303	0.459	0.282	0.450
Married/Cohabitation	0.602	0.489	0.599	0.490	0.631	0.483
Divorced/Separated	0.044	0.206	0.043	0.203	0.053	0.225
Widowed	0.053	0.224	0.055	0.228	0.034	0.180
No education	0.071	0.257	0.073	0.260	0.053	0.225
Primary	0.384	0.486	0.388	0.487	0.347	0.476
Secondary	0.271	0.444	0.269	0.443	0.289	0.453
High-School	0.163	0.369	0.161	0.367	0.180	0.385
University	0.111	0.314	0.109	0.312	0.130	0.337
South - East	0.205	0.403	0.215	0.411	0.105	0.307
Centre	0.185	0.389	0.182	0.386	0.216	0.411
West	0.212	0.409	0.208	0.406	0.246	0.431
North - West	0.214	0.410	0.214	0.410	0.220	0.414
North - East	0.184	0.387	0.180	0.385	0.214	0.410
House maker	0.349	0.477	0.373	0.484	0.119	0.324
Homeowner	0.801	0.399	0.802	0.398	0.787	0.409
Health insurance	0.407	0.491	0.402	0.490	0.451	0.498
Alcohol	0.335	0.472	0.299	0.458	0.672	0.470
Hot drinks	0.002	0.043	0.002	0.044	0.001	0.032
Exercise	0.135	0.342	0.130	0.336	0.184	0.388
Other smoker	0.333	0.471	0.291	0.454	0.734	0.442
N	**20606**		**18639**		**1967**	

Table 5 Sample Statistics, Adults who were not followed-up - Attritors, MxFLS-1.

	All		Non-Smokers		Smokers	
	MEAN	SD	MEAN	SD	MEAN	SD
Continuous						
Cigarettes	5.49	21.919	0.00	0.000	36.38	45.437
Age	35.31	18.108	35.48	18.614	34.35	14.917
Ln(Tot. Exp. PC)	6.52	1.322	6.49	1.316	6.69	1.341
Categorical						
Male	0.543	0.498	0.504	0.500	0.761	0.427
Indigenous	0.099	0.299	0.105	0.306	0.070	0.255
Urban	0.661	0.474	0.650	0.477	0.719	0.450
Single	0.382	0.486	0.393	0.489	0.317	0.466
Married/Cohabitation	0.522	0.500	0.509	0.500	0.597	0.491
Divorced/Separated	0.045	0.208	0.042	0.201	0.063	0.242
Widowed	0.051	0.221	0.056	0.231	0.024	0.152
No education	0.074	0.263	0.078	0.268	0.056	0.229
Primary	0.328	0.470	0.329	0.470	0.321	0.467
Secondary	0.289	0.454	0.281	0.450	0.335	0.472
High-School	0.176	0.381	0.178	0.382	0.167	0.373
University	0.132	0.339	0.134	0.341	0.121	0.326
South - East	0.176	0.380	0.184	0.388	0.128	0.334
Centre	0.216	0.412	0.212	0.409	0.242	0.429
West	0.177	0.382	0.179	0.383	0.168	0.374
North - West	0.199	0.400	0.198	0.399	0.206	0.405
North - East	0.231	0.422	0.227	0.419	0.256	0.437
House maker	0.216	0.411	0.236	0.425	0.102	0.302
Homeowner	0.671	0.470	0.677	0.468	0.638	0.481
Health insurance	0.458	0.498	0.452	0.498	0.491	0.500
Alcohol	0.422	0.494	0.379	0.485	0.665	0.472
Hot drinks	0.001	0.035	0.001	0.038	0.000	0.000
Exercise	0.235	0.424	0.228	0.420	0.273	0.446
Other smoker	0.307	0.461	0.238	0.426	0.694	0.461
N	**4768**		**4049**		**964**	

Table 6 Sample Statistics, Adults who were not followed-up - New Comers, MxFLS-2.

	All		Non-Smokers		Smokers	
	MEAN	SD	MEAN	SD	MEAN	SD
Continuous						
Cigarettes	3.32	17.208	0.00	0.000	36.70	45.283
Age	29.45	17.431	29.24	17.666	31.60	14.700
Ln(Tot. Exp. PC)	6.27	1.192	6.26	1.167	6.34	1.419
Categorical						
Male	0.510	0.500	0.482	0.500	0.790	0.408
Indigenous	0.105	0.307	0.109	0.312	0.065	0.248
Urban	0.626	0.484	0.616	0.486	0.724	0.447
Single	0.505	0.500	0.516	0.500	0.399	0.490
Married/Cohabitation	0.429	0.495	0.417	0.493	0.546	0.498
Divorced/Separated	0.034	0.182	0.034	0.181	0.040	0.195
Widowed	0.032	0.175	0.033	0.179	0.016	0.125
No education	0.056	0.229	0.057	0.232	0.042	0.200
Primary	0.271	0.445	0.267	0.442	0.315	0.465
Secondary	0.354	0.478	0.356	0.479	0.333	0.472
High-School	0.239	0.426	0.241	0.428	0.214	0.411
University	0.080	0.271	0.079	0.269	0.095	0.294
South - East	0.183	0.387	0.191	0.393	0.105	0.307
Centre	0.184	0.387	0.181	0.385	0.214	0.411
West	0.234	0.423	0.228	0.419	0.294	0.456
North - West	0.211	0.408	0.215	0.411	0.173	0.378
North - East	0.188	0.391	0.185	0.389	0.214	0.411
House maker	0.261	0.439	0.277	0.448	0.099	0.299
Homeowner	0.786	0.410	0.788	0.409	0.766	0.424
Health insurance	0.387	0.487	0.384	0.486	0.413	0.493
Alcohol	0.319	0.466	0.280	0.449	0.716	0.451
Hot drinks	0.001	0.033	0.001	0.034	0.000	0.000
Exercise	0.170	0.376	0.167	0.373	0.200	0.401
Other smoker	0.318	0.466	0.280	0.449	0.702	0.458
N	**5570**		**5066**		**504**	

Table 7 Sample Statistics, Adults who were followed-up, MxFLS-1 only.

	All		Non-Smokers		Smokers	
	MEAN	SD	MEAN	SD	MEAN	SD
Continuous						
Cigarettes	4.56	21.371	0.00	0.000	42.45	51.407
Age	38.38	16.745	38.22	16.868	39.68	15.635
Ln(Tot. Exp. PC)	6.46	1.130	6.44	1.109	6.63	1.275
Categorical						
Male	0.413	0.492	0.375	0.484	0.733	0.442
Indigenous	0.180	0.384	0.186	0.389	0.125	0.331
Urban	0.549	0.498	0.539	0.498	0.629	0.483
Single	0.266	0.442	0.269	0.444	0.238	0.426
Married/Cohabitation	0.641	0.480	0.639	0.480	0.658	0.474
Divorced/Separated	0.047	0.212	0.045	0.207	0.064	0.245
Widowed	0.046	0.209	0.047	0.211	0.039	0.194
No education	0.094	0.292	0.095	0.293	0.087	0.282
Primary	0.438	0.496	0.442	0.497	0.402	0.490
Secondary	0.254	0.435	0.251	0.434	0.274	0.446
High-School	0.130	0.336	0.129	0.335	0.137	0.344
University	0.085	0.278	0.083	0.275	0.100	0.300
South - East	0.213	0.409	0.223	0.417	0.126	0.332
Centre	0.186	0.389	0.185	0.389	0.189	0.391
West	0.204	0.403	0.203	0.403	0.209	0.407
North - West	0.215	0.411	0.209	0.407	0.266	0.442
North - East	0.182	0.386	0.178	0.383	0.210	0.408
House maker	0.356	0.479	0.382	0.486	0.142	0.349
Homeowner	0.748	0.434	0.749	0.434	0.735	0.442
Health insurance	0.437	0.496	0.433	0.495	0.473	0.499
Alcohol	0.369	0.482	0.333	0.471	0.668	0.471
Hot drinks	0.003	0.059	0.004	0.061	0.001	0.035
Exercise	0.184	0.387	0.181	0.385	0.207	0.405
Other smoker	0.286	0.452	0.237	0.425	0.691	0.462
N	**15036**		**13420**		**1616**	

Appendices to Chapter 3

Table 1 Sample Statistics, Adults who were followed-up, MxFLS 1 & 2.

	Participants (Smokers)			
	Sweep 1		Sweep 2	
	MEAN	SD	MEAN	SD
Continuous				
Cigarettes[a]	42.45	51.407	43.72	55.08
Age	37.79	15.96	41.19	15.94
Ln(Tot. Exp. PC)	6.59	1.24	6.56	0.98
Categorical				
Male	0.739	0.439	0.739	0.439
Indigenous	0.123	0.329	0.123	0.329
Urban	0.635	0.482	0.672	0.469
Single	0.283	0.451	0.233	0.423
Married/Cohabitation	0.624	0.484	0.665	0.472
Divorced/Separated	0.057	0.232	0.060	0.237
Widowed	0.036	0.186	0.042	0.201
No education	0.080	0.271	0.064	0.244
Primary	0.388	0.488	0.370	0.483
Secondary	0.284	0.451	0.273	0.446
High-School	0.147	0.354	0.157	0.364
University	0.101	0.301	0.136	0.343
South - East	0.122	0.328	0.122	0.327
Centre	0.198	0.398	0.198	0.399
West	0.222	0.416	0.222	0.416
North - West	0.242	0.429	0.242	0.429
North - East	0.216	0.411	0.216	0.411
House maker	0.129	0.336	0.134	0.341
Homeowner	0.738	0.440	0.805	0.396
Health insurance	0.474	0.499	0.455	0.498
Alcohol	0.625	0.484	0.615	0.487
Hot drinks	0.001	0.037	0.001	0.037
Exercise	0.213	0.409	0.151	0.358
Other smoker	0.601	0.490	0.730	0.444
N	**2232**		**2232**	

NOTE: [a] if positive.

Table 2 Sample Statistics, Adults who were followed-up, MxFLS 1 & 2.

	Non-Participants (Non-smokers)			
	Sweep 1		Sweep 2	
	MEAN	SD	MEAN	SD
Continuous				
Cigarettes	0	0	0	0
Age	38.48	16.88	41.88	16.89
Ln(Tot. Exp. PC)	6.44	1.11	6.32	1.07
Categorical				
Male	0.357	0.479	0.357	0.479
Indigenous	0.190	0.392	0.190	0.392
Urban	0.534	0.499	0.565	0.496
Single	0.263	0.440	0.223	0.416
Married/Cohabitation	0.644	0.479	0.666	0.472
Divorced/Separated	0.045	0.208	0.046	0.209
Widowed	0.048	0.213	0.064	0.246
No education	0.096	0.295	0.079	0.270
Primary	0.447	0.497	0.436	0.496
Secondary	0.249	0.432	0.234	0.423
High-School	0.127	0.333	0.131	0.337
University	0.082	0.274	0.120	0.325
South - East	0.229	0.420	0.228	0.420
Centre	0.184	0.387	0.184	0.388
West	0.201	0.401	0.201	0.401
North - West	0.211	0.408	0.211	0.408
North - East	0.176	0.381	0.176	0.381
House maker	0.396	0.489	0.425	0.494
Homeowner	0.749	0.434	0.806	0.395
Health insurance	0.431	0.495	0.407	0.491
Alcohol	0.324	0.468	0.293	0.455
Hot drinks	0.004	0.062	0.002	0.048
Exercise	0.179	0.383	0.117	0.322
Other smoker	0.231	0.421	0.270	0.444
N	**12804**		**12804**	