THE EFFECT OF PUBLIC POLICIES ON SMOKING BEHAVIORS: PERSPECTIVES FROM A DEVELOPING COUNTRY.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mabel Alejandra Andalón López

May 2010

THE EFFECT OF PUBLIC POLICIES ON SMOKING BEHAVIORS: PERSPECTIVES FROM A DEVELOPING COUNTRY.

Mabel Alejandra Andalón López, Ph.D.

Cornell University 2010

Hinging on economic theory predictions, this dissertation empirically examines the extent to which public policies influence youths and young adults smoking behaviors in Mexico. In the following, the decision to conduct the analysis separate for men and for women, in acknowledgement of separate gender paradigms, has yielded illuminating insights. This study contains three separate but related essays.

The first essay develops a new index of clean indoor air policies, analyzing the role of these policies --and other variables-- as determinants of cigarette consumption. The new index and previously developed indices are positively correlated, but imperfectly so. More stringent laws are negatively associated with the number of cigarettes smoked by men, but have no influence on smoking participation for either men or women. The statistical and economically meaningful predictors of smoking vary with the dependent variable (i.e. smoking participation and conditional intensity), and by gender.

The second essay investigates the causal effect of *Oportunidades*, a federal social program, on the smoking behaviors of its participants. I find that the unbundled components of the program, inclusive of sizable cash transfers, health information sessions and schooling, have no effect on adult smoking; they have a small positive effect on adolescent's cigarette consumption. Differential program treatments by sex position the research to isolate the effect of income, which shall be demonstrated to be not-statistically different from zero. Null effects for women and adolescents are

consistent with either: a) effects of each benefit in different directions that offset each other, or b) a zero effect of all components.

The third essay measures, and disentangles, the correlation between schooling and smoking. The positive association among women is shown to be due to unobserved heterogeneity. Smoking is more prevalent among more sophisticated women coming from better-off backgrounds. Conversely, 'Third Factors' are not found to be behind the negative association between schooling and smoking among men. Using variation in junior high school classroom openings during the nineties to deal with the endogeneity of schooling, I find that higher educational achievement may cause less smoking among men.

BIOGRAPHICAL SKETCH

Mabel Alejandra Andalón López was born on September 27th, 1978 in Oaxaca, Mexico. Her parents, Ramón and Teresa, inculcated in her good morals, perseverance, independence, positivism and a sense of curiosity. Mabel grew up in Puebla, Mexico with her sister, Viviana, and her brother, Gerardo. She majored in Economics at the *Universidad de las Américas-Puebla*. Soon after obtaining her Bachelor's Degree Mabel moved to Mexico City, where she worked for the Mexican Government for almost three years. Mabel arrived to Ithaca, New York in 2003. She obtained a Master's Degree in Economics from Cornell University in 2006, and continued as a graduate student in a research program in the same university. Mabel liked very much her life as a Cornell student. She particularly enjoyed spending time with friends, swimming, and dancing. Mabel married Giuseppe Infusini in 2008. This dissertation completes the requirements for her to earn a Ph.D in Policy Analysis and Management. Mabel will continue doing research in health and development health economics at the University of Melbourne in Australia.

To the memory of my mother, Teresa, and my brother, Gerardo

ACKNOWLEDGMENTS

This dissertation has come to an end thanks to the encouragement, support and constant guidance of the members of my dissertation committee Don Kenkel, Gary Fields and John Cawley. I thank Don for giving me the freedom to bring forth my own ideas and for providing me with uncountable opportunities to discuss them with him. He guided this research by pointing me to the relevant literature and by asking plenty of questions that forced me to “think as an economist” and challenged me to come up with answers myself. Despite his busy schedule, Don was always there to talk, particularly when answers did not come easy. And after each conversation with him I felt that my feet were on the ground again.

Gary has been a friend and a mentor. I was able to discuss with him my ideas and dissertation progress as needed, and I needed it quite often. Gary taught me how to write academic papers, from start to finish. Gary once suggested answering the question: What is the best use of my time now? Many times the answer involved doing some research. Other times the answer was to relax, go dancing, or enjoy the (scarce) good weather in Ithaca. Thanks to Gary’s suggestion I (almost) always felt that I was keeping some life balance while making good progress on my dissertation, which helped to keep anxiety levels relatively low. I thank Gary for all this and for his generous emotional support during difficult times at Cornell.

I thank John for sharing with me numerous tips on how to pursue research and present that research to others. He always encouraged me to think about the relevance of my research to the economics field. He was always willing to meet with me and review my work on a very short notice. John gave me very insightful comments on my dissertation work and many relevant suggestions for my academic career.

I am thankful to Jordan Matsudaira, Rob Masson, Tapan Mitra, and Kaushik Basu, for their support at different stages of my student life at Cornell. I owe my deepest gratitude to Luis Felipe López Calva for inciting my desire to pursue PhD studies, and for always being a strong advocate for me. I owe a special debt to Maru Ibarrarán, Gonzalo Castañeda, Carlos Ibarra, Enrique Cárdenas, Kristiano Raccanello, and Miguel Székely for aiding the PhD application process.

My sincere thanks go to Luis Rubalcava, Evangelina Morales, Olivia Guzmán, Patricia Cravioto, Vicente García, and staff at INEGI for their help in gathering the data used in this dissertation. I am grateful to Alejandra Macías for answering various questions regarding the data and operation of the Oportunidades program and for her assistance with the data used in Chapter 4 of the dissertation. I would also like to express my gratitude to Geysa Smiljanic and all the professional staff in the Department of Policy Analysis and Management. Their help with forms, signatures, letters of recommendation, room bookings etc. was essential to finish this project.

I am indebted to John, Tolu, Marcus, José, Doug, Yearim, Laura, Juan Carlos, Robert, and Manolo for their friendship, for the countless things they taught me about economics, computer science, animals, nutrition, life, religion, music, different cultures etc., and for helping me with a number of things when I got stock. I lived with them and with Jean, Luca, Ana, Rena, Kiran, Asia, Eva, Rafa, Maya, Mayte, Eunkyeong and Kai-Wen the most amazing moments at Cornell. I thank my friends Pepe, Nayeli, Liz, Jorge, Migue, Beto, Alvaro, Ramón, Carlos, Heidi, Yatziri, Claudia, Nancy, Bertha and Cary for always receiving me with open arms each time I was back in Mexico.

I thank my family for believing in me, for their emotional support and for encouraging me to pursue my interests. My sister, Viviana, was a tremendous source of psychological, emotional and even financial support from start to finish of this degree. She accompanied me on my first trip to Ithaca: checked out the city, set up my house, bought me books, and left a note wishing me "the best of lucks". Our daily conversations made me feel that I was closer to her than what the geographic distance suggested. Fortunately, we were often able to shorten that distance meeting in Ithaca, San Francisco, Los Angeles, Cancún, Mexico City, Guadalajara, Italy and the Caribbean! I thank Viviana for her unconditional love and words of encouragement. If sisters could be chosen, I would have chosen her.

My husband, Giuseppe, was on my side during the most rewarding (and many times painful) experiences of the PhD. He let me use his computer, helped me out with graphs and programs, gave me suggestions on posters and presentations, and kept me company during late nights of typing. Every time I felt that this was an impossible mission, he was there to hug me and say “you can do it”. I thank Giuseppe for all this and for sparkling my long days of work with a bit of fun time: we laughed and chatted for hours over dinner (nearly always prepared by him) and wine, watched all sorts of movies, played video-games, or went dancing. Most of all, I thank Giuseppe for bringing richness to my life.

I thank my father, Ramón, for instilling in me moral values and a sense of responsibility. I thank him an my *Tía* Anita and *Tío* Ballo*,* Lupis and her family, my cousins in Los Angeles, and Silvia for checking on me, for the telephone calls, postcards; in sum, for all the affection. Finally, I thank my mother, Teresa, and my brother, Gerardo, because my memories of their perseverance and joy in life enabled me to enjoy and conclude the PhD.

I acknowledge financial support from CONACyT and from the Department of Policy Analysis and Management and the Graduate School at Cornell University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BMI	Body Mass Index
CIAPI	Clean Indoor Air Policy Index
CONACyT	*Consejo Nacional de Ciencia y Tecnología*, National Council of Science and Technology
ENCASEH	*Encuesta de Características de los Hogares,* Household Characteristics Survey
ENSANUT	*Encuesta Nacional de Salud y Nutrición,* National Health and Nutrition Survey
ENUT	*Encuesta Nacional de Uso del Tiempo,* National Time Use Survey
FCTC	Framework Convention on Tobacco Control
IV	Instrumental Variables
INEGI	*Instituto Nacional de Estadística Geografía e Informática*, National Bureau of Statistics in Mexico
IEPS	Tax on Production and Services
LATE	Local Average Treatment Effect
LPM	Linear Probability Model
MxFLS	Mexican Family Life Survey

OLS	Ordinary Least Squares
PSM	Propensity Score Matching
PRI	*Partido Revolucionario Institucional,* Institutional Revolutionary Party
PAN	*Partido Acción Nacional,* National Action Party
PRD	*Partido de la Revolución Democrática*, Democratic Revolutionary Party
RD	Regression Discontinuity
VIF	Variance Inflation Factor
2SLS	Two-Stage Least Squares

CHAPTER 1

INTRODUCTION

1. Questions

Since the study of smoking behaviors in developing economies is still a fresh area of research, much remains to be learned. Analysts are typically drawn to the study of the causes of smoking in Mexico based on the strong consensus that smoking harms health. It is also true, however, that smokers receive benefits from smoking; otherwise they would not pay to do it. The economic models of addiction used as the basis of the empirical research in this dissertation contribute a better understanding to the individual's complex decisions, where this particular harmful, health act is concerned, elucidating the compelling tradeoff between immediate gratification, on the one hand, and long-term health and well-being, on the other.

Data from different sources allow, for the first time, the analysis of smoking behaviors in Mexico at the individual level. Emphasis is put on the effect of three sets of public policies that, based on theoretical and empirical economic research, are predicted to have influenced smoking behaviors substantially, through various channels.

Separately assessing the impact upon men and upon women, the rest of the dissertation poses the following original questions:

- Are clean indoor air policies associated with smoking participation and/or conditional intensity? What are the determinants of the demand for cigarettes among young adults?
- What is the causal treatment effect of participation in the *Oportunidades* program on adolescent and adult smoking? Is the effect of this program

heterogeneous across population subgroups defined in terms of their poverty intensity? What is the isolated income effect of the program on adult smoking?

- Is the positive correlation between schooling and smoking among women, as contrasted with the negative association among men, driven by usually hard-to-observe factors: such as utility function parameters, cognitive ability, and family background? If not, is the casual explanation, which correlates schooling with a diminishment in smoking useful?

Knowledge about the determinants of smoking and the effect of various policies on the smoking behaviors of different population subgroups and across gender can help in targeting strategies that aim to minimize the harm associated with cigarette smoking.

The rest of this chapter reviews up-to-date economic models of smoking, and describes the model that is used as the framework of analysis throughout this research. It also provides background on the demand and supply of cigarettes, against the backdrop of anti-smoking policies in Mexico.

2. Overview of Economic Models of Addiction

Unlike other goods, the nicotine contained in cigarettes makes their consumption addictive: smoking initiation leads to a compulsive need for, and use of, cigarettes. Standard economic theory predicts that the demand for a given good depends on its contemporaneous price, the current price of other products, individual's income, tastes and other factors. Standard theory, however, does not reflect the dependence of current consumption decisions on past behavior, figuring highly in the case of addictive substances. This explains why addictive goods were once considered resistant to standard economic analysis (Chaloupka et al., 2000).

Over the past few decades, a number of economic models, explicitly designed to incorporate the habit-forming nature of cigarette consumption, have emerged. Economic models of addiction can be divided into two major groups: myopic and rational. The main difference between these two groups of models hinges on the assumptions they make about individuals' rationality, where rationality reflects the degree to which the addict considers the future implications of current consumption decisions.

In myopic models of addiction, the individual recognizes the dependence of current addictive consumption decisions on past consumption. Nonetheless, when making current choices, individuals ignore the impact of current and past choices on future consumption. Tastes in myopic models have been treated both as endogenous (Gorman, 1967; Pollak, 1970) and exogenous (Mullahy's, 1985).

In rational addiction models individuals are assumed to be both backward and forward-looking. In the seminal rational-addictive work developed by Becker and Murphy (1988) individuals have stable preferences and seek to maximize life-time utility, taking into account the full price of addictive goods. Full price includes the current and future monetary price of cigarettes and the money value of the consequences of current consumption, such as the negative effects on earnings and the health problems associated with addiction.

One of the most important criticisms to Becker and Murphy's rational addiction model is its failure to capture the fact that just a few of the individuals who express a desire to quit the habit of smoking actually do it. To overcome this "self-control" problem, the decision to smoke was modeled on the basis of a behavioral economics approach that allows individuals to be time-inconsistent (Gruber and Köszegi, 2001). Individuals who are time inconsistent behave as if harboring two personalities --one is a myopic doer who adores tobacco, and the other is a farsighted planner who wants clean lungs (Thaler and Shefrin, 1981). In time-inconsistent models today's self is impatient. Faced with the tradeoff between the short- term pleasures of smoking and the long term health damages of doing so, the individual will prefer immediate gratification, so he will decide to smoke today. However, because tomorrow's self is much more patient, today's choice of tomorrow's consumption path will be consistent with the desire to quit smoking tomorrow. The problem is that when tomorrow comes, the future self who was patient is now the current impatient self. So the smoking continues to the long-term regret of the smoker; a smoker who does not always behave in his own interest.

Testing sophisticated economic models of rational addiction (Becker and Murphy, 1988) or time inconsistent preferences (Gruber and Köszegi, 2001) requires longitudinal data and exogenous changes in full prices, which are not available in Mexico. Therefore, the demand for cigarettes in this dissertation relies on the predictions of the myopic model of cigarette demand developed by Mullahy (1985), which I discuss in turn.

Mullahy's (1985) myopic model is a one period model that captures the three defining characteristics of cigarette addiction, namely reinforcement, tolerance, and withdrawal. Reinforcement, as the key determinant in the myopic model, implies that the more one partakes of an activity, the more one wishes to partake. Due to the body's adaptation to the drug, short-run tolerance implies that the benefit of smoking one more cigarette decreases with the amount of cigarettes already consumed. Because of a congestion phenomenon, long-run tolerance suggests that, holding constant the consumption of cigarettes, the production of nicotine is reduced when habit stocks are larger. Withdrawal reflects the negative physical reactions to the cessation, interruption, or reduction of consumption. Withdrawal effects are larger, the larger the stocks of habits are.

In Mullahy's model, the individual looks backward, but not forward, when facing the following instantaneous utility maximization problem:

$$\begin{aligned} &Max_{A,Y}\ U = \phi(N,H,Y) \qquad (1)\\ &s.t\ \ N = \mu(A,S,\theta)\\ &\quad H = \Psi(A,S,\theta)\\ &\quad AP_A + YP_Y = I\\ &\quad S = S_0,\ \ \theta = \theta_0\\ &\quad A \geq 0, \qquad Y \geq 0 \end{aligned}$$

N is nicotine service flows, H represents the perceptions of the individual in terms of health promotion and Y is a composite of non-addictive goods. The utility function is increasing in H and Y. $\phi_N > 0$ if the individual likes smoking. Both N and H are non-market commodities produced by transforming inputs. The individual's technological possibilities for the production of N and H are represented by the function N=$\mu(.)$ and $H = \Psi(.)$, respectively. A, which stands for "addictive good," is cigarette consumption; and S is a quasi-fixed factor of production which is held at the level $S_0 = S(T)$ in the instantaneous optimization problem. θ is a vector of variables that reflects tastes such as age, gender and education. In the budget constraint restriction, I is income, P_A is the price of the addictive good and P_Y is the price of the composite non-addictive good.

Due to short-run tolerance, the returns of the consumption of the addictive good are positive but decreasing. Therefore $\mu_A > 0$ and $\mu_{AA} < 0$. Long-term tolerance states that, holding the consumption of the addictive good constant, larger habit stocks reduce the production of nicotine services due to congestion; implying that $\mu_S < 0$. Long-term tolerance also suggests that $\mu_{AS} < 0$ because the marginal product of the consumption of the addictive good is diminished by increases in the stock of habits.[1] On the other hand, reinforcement suggests that $\mu_{AS} > 0$. Therefore, the sign of μ_{AS} depends on whether long-term tolerance or reinforcement are dominant.

$\Psi_A \leq 0$ for, when consumption increases, the perceptions of the individual in terms of health promotion do not increase. Ψ_{AA} can either be positive or negative. Since the stock is seen as a commitment to smoking, the greater the stock, the higher the health promotion perceptions implying that $\Psi_S > 0$. $\Psi_{AS} > 0$ (i.e. the perception of incremental health promotion due to additional cigarettes smoked increases with increasing levels of the stock).

The habit stock serves formally as a constraint on the instantaneous optimization problem (1), and affects demands through its role in the production of nicotine services. The stock equals the depreciated sum of an individual's past consumption of the addictive good as expressed in equation (2):

$$S = S(T) = \int_{\varphi} e^{-r(T-t)} A(t) dt \qquad (2)$$

where the process *{A(t)}t*$\in \varphi$ is the lifetime profile of an individuals' use of A, *r>0* is a decay rate, $\varphi = [\breve{T}, \mathrm{T}]$, T is the instant of decision and $\breve{T}$ is the initial time of positive purchase or utilization.

$S(t)$ is the solution to the differential equation relating changes in the state variable S to instantaneous consumption-investment, $A(t)$, and depreciation of existing stocks at the rate δ:

[1] Under these circumstances, an individual would rather prefer to reduce the stocks, but the quasi-fixed nature of the stock precludes any instantaneous diminution of the habit stock. In addition, the myopic behavior implies that individuals do not recognize at the instant of choice that the present use of the habituating good is serving as an investment in the stock.

$$\dot{S} = A(t) - \delta S(t) \quad (3)$$

Therefore, the only means for augmenting (decreasing) S is through investments in cigarette consumption (depreciation).

The demand function of the addictive good resulting as solution to problem (1) above is:

$$A(t) = [\boldsymbol{\theta}(t), S(t), P_A(t), P_Y(t), I(t),] \quad (4)$$

The demand function of addictive goods is thus a function of a vector of measured exogenous covariates; these in turn hold sway over numerous associated factors: the production of nicotine services and the perceptions of health promotion; the stock of the addictive good, own and cross prices, and individual income. Under certain parameter assumptions, not smoking is in the best interest of an individual, but smoking might also be.

3. Overview of the tobacco market in Mexico

Table 1.1, Panel A reports smoking indicators for individuals aged 20 to 65 by sex, calculated from the National Health and Nutrition Survey, 2006. As can readily be seen from this table, a gap exists between men's and women's current and former smoking rates. 30.1 percent of men and 9.5 of women report to be current smokers. Former smokers represent 16.50 percent of men and 5.80 percent of women. Compared to more developed countries, the intensity of smoking, measured as the number of cigarettes smoked a day, is relatively low in Mexico for both men and women. Even so, cigarette smoking is responsible for 9 percent of the annual deaths

Table 1.1 Profile of Smoking in Mexico.

	Men		Women	
	Mean	**SD**	**Mean**	**SD**
Panel A. Smoking Status				
Current	0.301	0.459	0.095	0.294
Former	0.165	0.371	0.058	0.234
No. Cigarettes[&]	5.886	7.725	4.125	5.473
Panel B. Individual Characteristics				
Cohort:				
20-29	0.344	0.344	0.107	0.107
30-39	0.327	0.327	0.098	0.098
40-49	0.304	0.304	0.111	0.111
50-65	0.286	0.286	0.088	0.088
Indigenous Condition:				
No	0.310	0.462	0.102	0.303
Yes	0.181	0.385	0.009	0.094
Marital Status:				
Married	0.298	0.458	0.079	0.270
Divorced	0.279	0.449	0.119	0.324
Single	0.317	0.465	0.130	0.336
Income (pesos):				
No income	0.274	0.446	0.084	0.277
0-1 thousand	0.222	0.416	0.050	0.218
1-2 thousand	0.244	0.430	0.076	0.265
2-5 thousand	0.326	0.469	0.139	0.346
5 thousand or more	0.335	0.472	0.191	0.393
Employment status:				
Employed	0.315	0.464	0.130	0.336
Unemployed	0.414	0.494	0.198	0.402
Inactive	0.229	0.421	0.075	0.263
Schooling:				
Elementary or less	0.282	0.450	0.060	0.237
Junior high school	0.351	0.477	0.099	0.298
High school	0.337	0.473	0.175	0.380
College	0.261	0.439	0.158	0.365
Observations	16991		22121	

Notes: Sample aged 20 to 65 years. [&]Smokers=5098 men and 6503 women.

Source: National Health and Nutrition Survey, 2006

Table 1.1 (Continued)

	Men		Women	
	Mean	**SD**	**Mean**	**SD**
Panel C. Location				
Area:				
Urban	0.318	0.466	0.114	0.318
Rural	0.239	0.427	0.022	0.147
Region:				
Northwest	0.295	0.456	0.095	0.293
Northeast	0.336	0.473	0.118	0.323
Southeast	0.223	0.417	0.037	0.188
Center	0.353	0.478	0.143	0.351
West	0.303	0.460	0.086	0.280
Observations	16991		22121	

Notes: Sample aged 20 to 65 years.

Source: National Health and Nutrition Survey, 2006

among Mexican men aged 35 and over, and 6 percent among women in the same age range (Valdés Salgado, 2005).

Smoking rates by socio-demographic characteristics are presented in Table 1.1, panel B. Men and women who are not indigenous, single, wealthy, and unemployed are more likely to smoke. There is no clear pattern between smoking rates and age in women; but men are less likely to smoke as they age. For both men and women, there is a U-shaped relationship between income and smoking rates. Finally, the percentage of smokers decreases substantially after junior high school among men. Surprisingly, the schooling-smoking gradient among women is positive and steep up to high school, and it decreases only slightly for those with college diplomas.

The prevalence of smoking varies significantly across geographic units (Table 1.1, panel C). Cigarette consumption is more prominent in urban than in rural areas. The highest prevalence of tobacco consumption is in Central Mexico and the lowest in the Southeast region. This evidence suggests that smoking is more prevalent in developed areas.

Recent epidemiological trends can be summarized as follows. First, the percentage of male smokers has remained relatively stable since 1988. Second, a higher percentage of women and adolescents smoke. And third, compared with older cohorts, the average age of addictive onset is lower among younger cohorts. It is 20.6 years for the cohort born in 1930 and 16.6 years for the cohort born between 1975 and 1978 (Campuzano-Rincón et al., 2005).

The tobacco market in Mexico is oligopolistic. Two companies hold about half of the national market share: British American Tobacco (BAT) and Cigatam-Philip Morris. These companies offer a wide spectrum of cigarette brands, with and without filter. Cigatam's brands include Marlboro--Mexico's bestselling cigarette since 1990---and Benson & Hedges. BAT owns three of the top five brands in Mexico: Boots, Raleigh and Montana.

Cigarettes are sold at accessible prices. Data from barcode scanning in large food stores, reported monthly by The Central Bank of Mexico, suggests that the average price of a pack of twenty cigarettes in 2006 was almost 15 pesos (about 1 US dollar). The minimum price was 12.9 pesos and the maximum 17.5 pesos.

4. Rationale for Government Intervention and Anti-smoking policies in Mexico

Despite the health and other consequences associated with smoking, standard economic theory provides no justification for government intervention in the market when rational consumers know all the risks and bear all the costs of their choices. Two market failures justify government intervention: 1) inadequate information about the addictive nature of tobacco consumption and the health risks of engaging in this behavior, and 2) physical externalities derived from second hand smoking, alongside the financial externalities arising in countries with an element of publicly financed healthcare, where non-smokers pay a share of the medical costs associated with the treatment of tobacco-related illnesses (Jha et al., 2000). In addition, economic models that depart from rational choice suggest that "optimal government policy should depend not only on the externalities that smokers impose on others but also on the "internalities" imposed by smokers on themselves" (Gruber and Köszegi, 2001, p. 1261).

A wide array of anti-smoking policies have been implemented in Mexico and elsewhere. Some of these are 'first-best' policies aiming to address a specific, identified market inefficiency. Other interventions are not justified by economic theory, but follow public health goals (e.g. reducing tobacco use in response to the world's tobacco epidemic). Nearly all tobacco control policies in Mexico emanate from the Framework Convention on Tobacco Control (FCTC). This is an international treaty that Mexico ratified in May of 2004, by which it agreed to adopt or maintain certain tobacco control measures.

The first national anti-smoking program in Mexico was launched in 1990, when the hazards of smoking where recognized in the Federal Health Act for the first time. Anti-smoking policies remained lax until 2000. After that year, taxes were raised, health-warning labels and anti-smoking mass media campaigns launched, and youth access restrictions to cigarettes and clean indoor air laws strengthened. Except for clean indoor air laws, all tobacco control policies are set at the national level. A brief description of these policies follows.

There are two types of taxes levied on manufactured cigarettes in Mexico: a 15 percent value-added tax, applied to the retail price; and a special tax on production and services (*IEPS*, in Spanish). By 2007, the average *IEPS* levied on cigarettes represented 140 percent of their total price. In the same year, the tax incidence on tobacco (i.e. the sum of all taxes on tobacco as a percentage of the retail price) represented 63.8 percent of the retail price of cigarettes (Sáenz de Miera et al., 2007). Unlike in other countries, taxes in Mexico do not vary across geographic units.

Two policy options have been used to address the information market-failure: antismoking campaigns and health-warning labels on cigarette packs. Anti-smoking campaigns consist of print and electronic mass media, disseminating information on the health risks of smoking, and the benefits of quitting smoking. Health warnings on cigarette packages were displayed for the first time in July of 2000, when the Regulation on the Control of Tobacco Consumption was enacted. Since the advent of stronger regulations, both the size and content of the health-warning labels have changed. Since 2004, health-warning labels take up 50 percent of the main display surface of the package and they clearly state the hazards of smoking.

In terms of advertising, the Federal Regulation on Cigarette Advertising enacted in 2000 was mainly focused on regulating sponsorships. Tobacco advertising was not banned in radio and television until 2004 (Sáenz de Miera et al, 2007). In that same year, cigarette publicity was prohibited within 200 meters of schools, hospitals and other public entrances. In addition, the distribution of free tobacco samples to minors was outlawed. To date, low-tar "light" cigarette advertisement is still permitted.

Youth access laws in Mexico include minimum age-at-sale (18 years), banning of self-service displays, limits to vending machines, and bans on loose cigarette sale. Despite the cited bans regarding the sale, or distribution of tobacco to under-age adolescents, Kuri et al. (2006) have found that 73 percent of retailers in Mexico City sell cigarettes to minors; only 7 percent of vendors ask customers to show a valid identification.

Clean indoor air laws are regulations that aim to protect the 41 percent of Mexicans who are affected by second-hand smoking (*INEGI*, 2004). Clean indoor air laws were implemented for the first time in 1990, when the Ministry of Health banned smoking in all health care centers. This regulation was extended to all federal buildings in 2000.

Clean indoor air policies are the only anti-smoking policies that vary across states. By April of 2008, 21 states had enacted clean-air statues, of which only Baja California, Tlaxcala, and Aguascalientes, passed laws before the Tobacco Control Act was enacted in 2000.

5. *Economic Analysis of Smoking in Mexico*

The last decade has seen a growing number of economists devote substantial effort to more fully understanding the tobacco market. Empirical application of recent theoretical perspectives on addictive goods has aided the understanding of the role of prices, policies and other determinants of smoking. Most economic research has focused on developed countries. The lack of research in developing countries is particularly striking, given that more than half of smokers live in these parts of the globe (Warner, 2008).

All the studies of cigarette consumption in Mexico have attempted to identify the effect of prices on smoking participation and conditional intensity (Sesma-Vázquez et al. 2002; Sesma-Vázquez et al., 2005; and Jiménez-Ruíz et al. 2008). The unit of analysis in these studies is households, which might mask features of the smoking decisions that take place individually. Also, their measure of price, computed as a household’s expenditures on cigarettes divided by the number of packs purchased by a household or by a group of households sharing a number of socioeconomic characteristics, has as its denominator an endogenously determined quantity.

To the best of my knowledge, this dissertation constitutes the first thorough analysis of cigarette consumption in Mexico from an economic perspective. This study is novel in various ways. First, I concentrate attention on a series of policies that, based on previous theoretical and empirical health economics research, are predicted to have possible non-intended effects on smoking. I am aware of no previous study that has investigated the effectiveness of clean indoor air laws or any other of the recent tobacco regulations in Mexico. I therefore ask whether differences in the enactment and stringency of clean indoor laws are associated with smoking consumption. I also measure the effect on smoking decisions of the *Oportunidades* program and of an exogenous change in classroom availability that affected schooling. Second, an in contrast with previous studies, I use individual rather than household-level data. Third, the analysis is conducted separately for men and women, which gives new insights in terms of the motivations for health investments by gender.

In the course of this research, I use a number of different data sets. Household-based surveys include the National Health Survey 2000, the National Health and Nutrition Survey 2006, the *ENCASEH* 1997, the *ENCEL* 2003, the Mexican Family Life Survey 2002 and 2005 and the National Survey of Addictions 2002. I also use numerous state-level characteristics. The sources of these data and the definition of the variables are discussed in each chapter. I use so many data sets because no single source of data has information on all the hypotheses I seek to test. Most of the analysis in this dissertation uses quasi-experimental research designs to achieve casual inference.

6. *Layout of the Dissertation*

The remainder of this dissertation is comprised of four chapters. Chapter Two is the first substantive essay. It develops a new index of clean indoor air policies and analyzes the role of these policies and other variables as determinants of cigarette consumption. Chapter Three investigates the causal effect of a federal anti-poverty program in Mexico, *Oportunidades*, on the smoking behaviors of its participants. Chapter Four measures and disentangles the correlation between schooling and smoking. Chapter Five elucidates my research conclusions.

REFERENCES

Becker, Gary S. and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96:4, pp. 675-700.

Campuzano Rincón, Julio C., Mauricio Hernández Avila, Jonathan M. Samet, Ignacio Méndez Ramírez, Roberto Tapia Conyer, and Jaime Sepúlveda Amor. 2005. "Comportamiento de los fumadores en México según las Encuestas Nacionales de Adicciones 1988 a 1998," in *Primer Informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco*. Eduardo Lazcano-Ponce, Mauricio Hernández Ávila and Raydel Valdés Salgado eds. Cuernavaca: Instituto Nacional de Salud Pública.

Chaloupka, Frank J., John A. Tauras, and Michael Grossman. 2000. "The economics of addiction," in *Tobacco control in developing countries*. Prabhat Jha and Frank Chaloupka eds. New York: Oxford University Press.

Gorman, William M. 1967. "Tastes, Habits and Choices." *International Economic Review*, 8:2, pp. 218-22.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *Quarterly Journal of Economics*, 116:4, pp. 1261-303.

Instituto Nacional de Estadística, Geografía e Informática (INEGI). 2004. "Encuesta Nacional de Adicciones, ENA-2002." INEGI: Aguascalientes.

Jha, Prabhat, Philip Musgrove, Frank J. Chaloupka, and Ayda Yurekli. 2000. "The economic rationale for intervention in the tobacco market," in *Tobacco control in developing countries*. Prabhat Jha and Frank Chaloupka eds. New York: Oxford University Press.

Jiménez-Ruiz, J. A., Belén Sáenz de Miera, Luz M. Reynales-Shigematsu, Hugh. R. Waters, and Mauricio Hernández-Avila. 2008. "The impact of taxation on tobacco consumption in Mexico." *Tobacco Control*, 17:2, pp. 105–10.

Kuri-Morales, Pablo A., Mario Cortés-Ramírez, Patricia Cravioto-Quintana, Jesús Hoy María, and Jesús F. González-Roldán. 2006. "Políticas, vigilancia y acceso de los menores de edad a los cigarros en la Ciudad de México." *Salud Pública Mex*, 48:supl 1, pp. S99-S106.

Mullahy, John. 1985. "Cigarette Smoking: Habits, Health Concerns and Heterogeneous Unobservables in a Micro-econometric Analysis of Consumer Demand." Ph.D. Dissertation. University of Virginia: Charlottesville.

O’Donoghue, Ted and Matthew Rabin. 2000. "Risky Behavior Among Youths: Some Issues from Behavioral Economics." *Institute of Business and Economic Research Working Paper No. 285* University of California: Berkeley.

Orphanides, Athanasios and David Zervos. 1995. "Rational addiction with learning and regret." *The Journal of Political Economy*, 103:4, pp. 739-58.

Pollak, Robert A. 1970. "Habit Formation and Long-Run Utility Functions." *Journal of Economic Theory*, 13:2, pp. 272-97.

Sáenz de Miera, Belén, Jorge A. Jiménez, and Luz M. Reynales. 2007. "La economía del tabaco en México." Instituto Nacional de Salud Pública: Cuernavaca.

Sesma-Vázquez, Sergio, Julio C. Campuzano-Rincón, Victor G. Carreón-Rodríguez, Felicia Knaul, Francisco J. López-Antuñano, and Mauricio Hernández-Avila. 2002. "El comportamiento de la demanda de tabaco en México: 1992-1998." *Salud Pública Mex*, 44:supl 1, pp. s82-s92.

Sesma-Vázquez, Sergio, Raymundo Pérez Rico, Esteban Puentes Rosas, and Raydel Valdés Salgado. 2005. "El precio como determinante del consumo de Tabaco en México, 1994 – 2002," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco*. Eduardo C. Lazcano-Ponce Raydel Valdés-Salgado, and Mauricio Hernández-Avila ed. Cuernavaca: Instituto Nacional de Salud Pública.

Strotz, Robert H. 1955-1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *The Review of Economic Studies*, 23:3, pp. 165-80.

Thaler, Richard H. and Hersh M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy*, 89:2, pp. 392-406.

Valdés Salgado, Raydel. 2005. "Las cifras de la epidemia. Daños a la salud y mortalidad atribuible," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco*. Raydel Valdés-Salgado, EC Lazcano-Ponce and Mauricio Hernández-Avila eds. Cuernavaca: Instituto Nacional de Salud Pública.

Warner, Kenneth E. 2008. "The Framework Convention on Tobacco Control: Opportunities and Issues." *Salud Pública Méx,* 50:3, pp. s283-s91.

CHAPTER 2

CLEAN INDOOR AIR POLICIES AND SMOKING

1. Introduction

As of 2008, 21 of the 32 states in Mexico had enacted regulations aiming to restrict smoking indoors. Despite the growing number of public health advocates for ordinances restricting smoking in public places (the *Economist*, 2008), the effectiveness of these laws as tobacco control strategies in Mexico has not been assessed yet. Nor has individual-level data been used to explore the determinants of smoking in Mexico. This paper attempts to fill these gaps by examining whether state-level clean indoor air laws and other socio-economic characteristics affect individual smoking behaviors.

The aim of clean indoor air policies is to protect non-smokers from the physical discomfort associated with passive smoking, such as irritation, annoying effects and health effects. These policies also aim to resolve the financial externalities that come about when, in countries where health care is publicly financed, non-smokers pay a share of the medical costs associated with the treatment of smoking related diseases medical costs associated with tobacco related illness treatments. These costs accounted for about 5 percent of the total expenditure of the Mexican Social Security Institute,--the principal health care provider in Mexico--(Reynales et al., 2006). Government interventions like this are justified by economic theory on efficiency grounds (Jha et al., 2000).

Despite that clean indoor air policies do not intend to address the choice of whether to smoke, economic models of addiction suggest that they might decrease cigarette consumption. These models predict that the optimal level of tobacco consumption occurs when the marginal costs of smoking equal the benefits. Restrictions to smoke in enclosed areas increase the costs associated with smoking. By imposing fines or other penalties to people smoking indoors, stringent and enforced laws increase the monetary costs of cigarette consumption. To adhere with the laws smokers are forced to smoke outdoors. This raises the time, and possibly the discomfort, associated with smoking. Bans on public and workplace smoking augment the individuals' direct and indirect costs of smoking relative to the benefits, which might prompt some smokers to cut back their consumption or quit.

Clean indoor air laws are not necessarily binding for all individuals. As in many other countries, smoking restrictions in Mexico are limited to certain locations (e.g. workplaces and hospitals). Where imposed, the constringency of the provisions is not homogenous. In some places regulations are very restrictive, but in others they are merely symbolic legislative statements. Tobacco control laws impose greater costs on individuals who spend most of their time in locations where the provisions are either more restrictive or better enforced. Put it differently, people who spend no time in places where smoking is limited or prohibited face no extra costs of smoking, so law enactment is unlikely to influence these individual's smoking decisions. Additionally, laws that are not enforced, or perceived not to be, have much the same effect as having no law.

Previous empirical research has examined the relationship between clean indoor air regulation and cigarette consumption. In the U.S., solid support has been found for the hypothesis that restrictions on smoking in enclosed places have a significant effect on both adult and teenage smoking participation or conditional demand (see, for example, Wasserman et al., 1991; Ross and Chaloupka, 2004; DeCicca et al., 2008a). The effect of these policies has been generally measured with a dummy indicating whether a law has been enacted in a given state/county. Only a few papers have made steps to incorporate specific features of the law in the analysis of the effectiveness of clean indoor air policies on smoking. For example, Tauras (2006) and Chirqui et al. (2002) rated each location on the extensiveness of legislative provisions. Both studies then created an index of clean-air laws as a weighted average of specific-location ratings. A caveat of these studies is that, assuming that the effect of bans on smoking is independent of the time people spend in each location, all locations are weighted equally.

Some studies have considered that more comprehensive laws are those that cover places where a large fraction of the population is intended to be affected or where individuals spend a large amount of time. Chaloupka and Pacula (1999) used as weights the fraction of the population who are subject to smoking restrictions in the person's county of residence. An index developed in a Surgeon General's report assigned full weight to private worksites because "individuals spend more time at work than in any other place outside the home" (U.S. DHS, 1986; p. 327), but it neglected that people spend time in other locations besides workplaces. None of these indices measure the degree of stringency or take into account the enforcement aspect of the law.

Building on economic theory predictions, one contribution of this study is to develop a new method to create an index of clean indoor air tobacco control policies. Employing this index, which is more comprehensive than previous indices, helps to deal with the problem of multicollinearity among numerous policies regulating smoking behaviors. As in previous indices, this index takes into account whether smoking is partially or fully banned. It improves upon previous indices in two main ways. First, the contribution of each location to the overall index is a function of the costs imposed on potential smokers. Holding the restriction level constant, the larger the fraction of the population spending most of their time in a given location, the higher the contribution of that location to the overall index is. Second, the effectiveness of the clean indoor air restrictions in this index varies with the enforcement level of the law. The index developed in this chapter serves to generate a "summary score" of the legislation for population subgroups in each state, thus facilitating comparison of clean indoor air restrictions across populations, geographic areas and over time. The index is applied here to Mexican data.

Another contribution of this chapter is the analysis of the determinants of cigarette consumption, paying particular attention to the role of clean indoor air policies. Sesma-Vázquez et al. (2002), Sesma-Vázquez et al. (2005), and Jiménez-Ruíz et al. (2008) have provided some initial steps towards an evidence-based diagnostic of the determinants of cigarette use in Mexico. Instead of using household-level data as in previous studies, I use individual-level data. Indeed, the empirical analysis is conducted separately for men and women. This distinction is important as previous research has suggested that there are sex differences in nicotine metabolism (Ben-Zaken Cohen, et al., 2007). Also, the fact that the male population in Mexico smokes three times as much as the female population and that the gender gap has narrowed over time suggests that men, women, youth and old adults might have different motivations for smoking, and may also respond to the same incentives distinctly.

The empirical analysis is based on individual information from the National Health and Nutrition Survey, 2006, matched to clean indoor air and other policies at the state and at the state and economic activity levels. The sample is restricted to individuals aged 20 to 50 years. The estimation equation is derived from Mullahy's (1985) model of addiction presented in the Introduction of this dissertation. I use a linear probability model for the estimation of smoking participation and ordinary least squares to estimate the conditional intensity of smoking. In alternative models I use a probit model to estimate smoking participation. The preferred measure of clean indoor air policies is the index developed in this chapter, but I also explore other measures of state regulations restricting smoking in public places.

The findings in this chapter suggest that more stringent laws have no influence on smoking participation for either men or women, but are negatively associated with the number of cigarettes smoked by males. More specifically, a one unit increase in the index is associated with 6.2 less cigarettes smoked a day. The statistical and economically meaningful predictors of smoking vary with the dependent variable (i.e. smoking participation and conditional intensity), and by gender.

The rest of this chapter proceeds as follows. In section 2 I develop an index of clean indoor air policies. Section 3 applies this index to Mexican data and compares it to previously developed indices. Section 4 discusses the data and methods to be used in the analysis of the determinants of cigarette consumption among young Mexican adults. Section 5 presents the empirical results on the determinants of smoking, paying particular attention to the role of policies that limit smoking indoors. Section 6 explores the causes of the effects found in Section 5 and discusses the results in the context of the current economic literature. Section 7 concludes and provides guidelines for future research.

2. Clean Indoor Air Policy Index

This section develops a Clean Indoor Air Policy Index (*CIAPI*), which provides a composite snapshot of clean indoor air policies. The *CIAPI* has two dimensions. The first dimension measures the *strictness* of tobacco control restrictions in public places. The second dimension assesses the overall level of *enforcement* of the policies in those locations.

The calculation of the *strictness* dimension involves four steps. First, following Chriqui et al. (2002), the severity of state clean indoor air laws in each location is rated based on general guidelines reflecting health policy goals. For example, states with full bans on smoking in workplaces might receive the maximum rating value when that is the target to aspire to. Compared to those states, states where smoking is permitted in designated areas will get lower scores.

Second, a performance indicator L_{is} is calculated for each location *i* in state *s.* This can be done applying the following formula:

$$L_{is} = \frac{actual\ value - minimum\ value}{target\ value - minimum\ value} \quad (1)$$

L_{is} in (1) ranges between 0 and 1. Values of performances lower than one indicate that the restrictions in a specific location fall far short of what is considered to be optimum legislation.

Third, the total population is divided into groups based on broad categories of main economic activity (e.g. public workers and students) and the average time each of the groups spends at each location is estimated. The total number of groups might be larger than the groups affected by the law. This is possible when the legislation does not consider bans in places where relevant groups spend most of the time, and the principles for banning smoking in those places are similar to those in other place (e.g. regulations in homes as workplaces). The fourth step involves the calculation of the *strictness* of the law dimension. For each group g considered, the strictness index S_{sg} is calculated as a weighted average of location-specific performance levels:

$$S_{gs} = \sum_{i=1}^{n} \alpha_{igs} L_{is} \quad (2)$$

The weights α_{igs} are the average proportion of time that people in group g spend at location i and state *s*. The fifth step is optional. State-level strictness indices might be constructed based on group-specific indices if the proportion of the population in each group (P_g) is known. This can be done as follows:

$$S_s = \sum_{g=1}^{4} P_g S_{gs} \tag{3}$$

The calculation of the enforcement dimension involves three similar steps. In the first step two other features of the legislation are tallied based on targets. One aspect measures the degree to which the law defines sanctions to smokers and proprietors (i.e. sanctions S_s). The other evaluates whether an appropriate authority is designated to enforce the law (i.e. enforcement agency A_s). The second step involves calculating a performance indicator for each enforcement aspect using a formula similar to (1) above. Finally, the enforcement dimension E_s in state *s* is calculated as a weighted average of its two components. The weights are equal because higher penalties are not effective if no proper authority is designated to enforce the law and vice versa:

$$E_s = \frac{1}{2} A_s + \frac{1}{2} S_s \tag{4}$$

In any given year, the Clean Indoor Air Policy Index $CIAPI$ for a given state and population subgroup can be expressed as:

$$CIAPI_{gs} = B_{gs} * E_s \tag{5}$$

The two components enter equation (5) multiplicatively to reflect that the efficacy of smoking restrictions should depend on both the stringency of the law and its enforcement level. The overall index thus decreases as differences in the actual and desired level of enforcement increase. The CIAPI is a standardized index ranging from 0 to 1. As previously mentioned, the CIAPI index might also be expressed at the state level. To do that B_{gs} in equation (5) may be replaced by B_s.

The clean indoor air policy index developed in this section improves upon previous indices in two main ways. First, it assesses the extent to which different people should be affected by clean indoor air policies varying the location-specific weights according to the proportion of people attending the place in question and the average duration of their visits. Second, the effectiveness of the restrictions in this index depends on the enforcement of the law. Additionally, defining the index at the economic activity and state level allows comparisons across states, population subgroups and over time. The first application of this index to Mexican data is discussed in turn.

3. Clean Indoor Air Policy Index in Mexico

As of May of 2006, 13 of the 32 states in Mexico had laws restricting smoking in Mexico. The state-level legislation for the calculation of the Clean Indoor Air Policy Index developed in the previous section comes from *Orden Jurídico Naciona*l, a repository of laws available online and personal communications with state Congress officers. The index scores the strictness of regulations in eleven locations: public workplaces, private workplaces, schools, houses (as workplaces), public transportation, restaurants, retail stores, recreational facilities, cultural facilities, health care facilities and private homes. The Appendix of this chapter provides details on how the laws were coded and how the weights were calculated.

In principle, location-specific weights of the CIAPI should be constructed calculating the time individuals spend in each place they report to be. I am not aware of any dataset asking detailed information on where people were, either in Mexico or elsewhere. The Mexican Time Use Survey, 2002 used in this chapter is not ideal, but it is the most suitable available data to compute weights. In many circumstances this survey asked what people did in the reference week and where they did what they did. For example, people reported the time spent studying at school and at home. In other circumstances, however, individuals only reported what they did. Hence, based on people's responses on what they had done I inferred where people were. For example, I assumed that all the people who reported to have cleaned their houses did it indoors, but they could have done it outdoors.

The CIAPI index was calculated separately for men and women. To be consistent with the sample restrictions used in the analysis of cigarette consumption, the index weights were based on the time spent by people aged 20 to 50 years. The index discussed in this section is composed of four groups: public workers, private workers, students and house workers.

Table 2.1 presents evidence suggesting that there is multicollinearity between smoking restrictions in different locations. The pairwise correlations of smoking restriction scores reach up to level 0.95 (top of the table). Except for two cases, the variance inflation factors exceed 10; the cutoff that has been suggested to denote collinearity (see Belsley, 1991). R-squares from auxiliary regressions that use each location as the dependent variable and other locations as independent variable range from 0.84 to 0.99, indicating that restrictions in other locations explain most of the variation of the restrictions in a given location. The CIAPI index developed in the previous section addresses the correlation problem among policy variables.

Table 2.1 Multicollinearity tests of Clean Indoor Air Policies in Locations Considered in the Mexican Legislation.

	Pairwise Correlation										
	Work	Schools	Transp.	Restaur.	Retail Stores	Recreat. Place	Cultural Places	Health Care Centers	Sanctions	**VIF**	**R-squared**
Work	1									6.07	0.84
Schools	0.781	1								37.43	0.97
Transport.	0.649	0.797	1							25.29	0.96
Restaur.	0.651	0.699	0.588	1						20.87	0.95
Retail Stores	0.673	0.795	0.810	0.592	1					98.02	0.99
Recreat. Places	0.258	0.642	0.337	0.287	0.496	1				9.23	0.89
Cultural Places	0.814	0.858	0.808	0.843	0.906	0.369	1			187.33	0.99
Health Care Centers	0.795	0.772	0.863	0.740	0.691	0.119	0.848	1		19.35	0.95
Sanctions	0.696	0.769	0.951	0.616	0.753	0.306	0.786	0.884	1	17.22	0.94
Enforcement Agency	0.777	0.920	0.800	0.814	0.653	0.441	0.841	0.841	0.800	44.43	0.98

Notes: Tranport=Transportation, Restaur.=Restaurants, Recreat.=Recreational,VIF=Variance Inflation Factors and R-squared comes from regressions where regulations in a given location are regressed on restrictions in all other locations.

Source: Own calculations based on Mexican statutory state legislation

Table 2.2 illustrates the calculation of the $CIAPI$ at the state-level in the states with the lowest CIAPI for women and the highest for men (henceforth, state 1 and 2). The *strictness* dimension in state 1 is very low because the population share with the highest index (students) is small, and that of the lowest index (home workers) is large. In contrast, state 2 has a relatively high index because the restrictions in private and public workplaces in this state rank the highest, and the percentage of people affected by these provisions represents 86 percent of the population. Both state indices drop when the enforcement component is taken into account. This is because, based on health targets, the enforcement level is 50 percent of what it should be in state 1 and 75 percent of what is ought to be in state 2.

Table 2.2 Example of the Calculation of the Clean Indoor Air Policy Index (CIAPI) at the State-Level.

	Group-specific strictness index, S_{gs} (1)	**Population Share, P_s (2)**	**(1)*(2)**
State 1 - Lowest CIAPI:			
Public Workers	0.106	0.08	0.008
Private Workers	0.067	0.41	0.027
Students	0.248	0.01	0.002
Home Workers	0.021	0.51	0.010
	State-level strictness index, S_s (3)		0.048
Sate 2 - Highest CIAPI:			
Public Workers	0.544	0.22	0.122
Private Workers	0.446	0.64	0.286
Student Workers	0.150	0.06	0.009
Home Workers	0.225	0.08	0.017
	State-level strictness index, S_s (3)		0.433
	State-level strictness index S_s (3)	**Enforcement E_s (4)**	**$CIAPI_s$ (3)*(4)**
State 1 - Lowest CIAPI	0.048	0.50	0.024
Sate 2 - Highest CIAPI	0.433	0.75	0.325

Source: Own calculations based on ENUT (2002) and Mexican statutory state legislation.

Table 2.3 presents basic descriptive statistics of the state-level *CIAPI* index and its components by gender. Only the 13 states that had clean indoor air laws in effect as of May of 2006 were included in these calculations. Considering that the overall index and both of its dimensions range from 0 to 1, the strictness dimension for both males and females is considerably low. In fact, it is much lower for women than for men because a larger proportion of females spend most of their time in places where smoking is not regulated at all. Moreover, since the enforcement component in all the states is less than what is desirable, the CIAPI index is even lower. There variability of the the index is higher for men than for women.

Table 2.3 Descriptive Statistics of the State-Level CIAPI Index and its Components.

	Mean	Std. Dev.	Min	Max
Men:				
Strictness index	0.271	0.106	0.098	0.433
Enforcement index	0.596	0.146	0.375	0.750
CIAPI	0.164	0.089	0.061	0.325
Women:				
Strictness index	0.119	0.042	0.048	0.169
Enforcement index	0.596	0.146	0.375	0.750
CIAPI	0.070	0.029	0.024	0.115

Source: Own calculations based on ENUT (2002) and Mexican statutory state legislation.

Table 2.4 reports correlation coefficients of the *CIAPI* and the other indices that have been previously used in this literature for the states that had enacted laws by May of 2006. Following Tauras (2006), the unrated index is the sum of indicator variables measuring whether tobacco restrictions and enforcement provisions existed in a given location. The rated index corresponds to the sum of the rated scores in each location. The rating was based on the codebook presented in table A2.1 of the

Appendix. This index is the Mexican version of the one used by Chriqui et al (2002). As expected, the correlation among indices is positive, but imperfectly so. [2]

Table 2.4 Pairwise correlation of Clean Indoor Air Policy Indices.

	Men			Women		
	Unrated	Rated	CIAPI	Unrated	Rated	CIAPI
Unrated	1			1		
Rated	0.436	1		0.436	1	
CIAPI	0.101	0.567	1	0.038	0.752	1

Notes: Unrated refers to an index constructed as the sum of dichotomous variables for each location. A given location is coded as 1 if smoking is restricted in that specific location.

Rated refers to an index constructed as the sum of the rated scores for each location. The rating is based on the codebook presented in table A 2.1.

CIAPI is a weighted index of rated tobacco restrictions in specific locations. The weights are based on the proportion of people affected in each of the locations and the average time of their visits.

Source: Own calculations based on ENUT (2002) and Mexican statutory state legislation.

4. Data and Methods for the Estimation of the Determinants of Smoking

4.1 Data

This chapter uses the Mexican Health and Nutrition Survey (*ENSANUT*, in Spanish), 2006. This is a cross-sectional data set that contains a wide array of individual-level data and self-reported cigarette consumption. The *ENSANUT* is representative of Mexican states, rural-urban areas and metropolitan cities. A detailed description of this survey can be found in Olaiz et al. (2006).

The *ENSANUT* provides information of one individual aged 20 or older per household. Throughout the analysis I restrict attention to people aged 20 to 50 years. The potential sample in this age range consisted of 32,669 individuals. Elimination of observations with missing information in state of residence and other relevant variables yields a final sample of 13,816 men and 18,314 women.

[2] The correlations are considerably higher when all states are included because the 18 states with no laws are coded as zero in all the indices.

Smoking participants are those who report to be current smokers given that they have smoked at least 5 packs of cigarettes in their life. The intensity of smoking corresponds to the number of cigarettes smoked per day. This measure is constructed with information on the number of cigarettes smoked and the frequency of smoking (occasionally, daily, weekly and monthly). Occasional smokers are treated as non-smokers. Given the difficulties to econometrically overcome the tendency of individuals to report amounts consumed as packs or half packs, I proceed on the assumption of error- free data. The stock of cigarettes is proxied by the number of years the person has smoked regularly. This measure takes into consideration both the time the person has smoked and the intensity of smoking. Its only caveat is that "smoking regularly" is somewhat subjective.

Table 2.5 reports descriptive statistics of the estimation sample. The top panel presents individual-level variables and the lower panel state-level variables that were matched to the survey data based on each respondent's state of residence. There were two exceptions to this rule. The CIAPI was merged based on both the individual's main activity and state of residence. Since the *ENSANUT* does not distinguish between private and public workers, indices in each state were constructed for three groups of individuals: workers (public and private combined), students and individuals working at home. Cigarette prices were merged to individual-level data based on the state of residence and month of interview.

State-level data comes from different sources. The sources of legislative records were discussed in the previous section. The percentage of people who think that smoking is either "dangerous" or "very dangerous" comes from the National Addiction Survey (*INEGI*, 2004). Information on the percentage of urban people, unemployment rates and tobacco producer states comes from the National Institute of Statistics. Cigarette prices, which include taxes, are derived from barcode scanning in

Table 2.5 Descriptive Statistics of the Estimation Sample.

	Men		Women	
	Mean	**S.D**	**Mean**	**S.D**
Individual-Level:				
Smoker	0.325	0.468	0.106	0.307
Former	0.126	0.332	0.050	0.217
Avg. no. cigarettes	5.243	7.051	3.545	4.769
Stock	4.145	7.020	1.232	4.275
Age	33.711	8.867	33.945	8.702
Some middle school	0.649	0.477	0.577	0.494
Income (thousands)	0.357	0.451	0.142	0.280
Indigenous	0.059	0.236	0.069	0.253
Single	0.290	0.454	0.232	0.422
Married	0.686	0.464	0.686	0.464
Divorced	0.023	0.151	0.082	0.274
Worker, public or private	0.872	0.334	0.360	0.480
Student	0.042	0.201	0.033	0.178
Worker at home	0.086	0.280	0.607	0.489
Pregnant	-	-	0.023	0.149
Oportunidades	0.169	0.374	0.185	0.388
Rural	0.200	0.400	0.205	0.404
Northwest	0.087	0.283	0.081	0.273
Northeast	0.151	0.358	0.138	0.345
Southwest	0.265	0.441	0.270	0.444
Central	0.297	0.457	0.304	0.460
West	0.201	0.400	0.206	0.405
State-Level:				
CIAPI	0.045	0.064	0.019	0.034
Law (=1)	0.411	0.492	0.409	0.492
Rated law	8.072	9.854	7.963	9.765
CIAPI-state only	0.047	0.063	0.023	0.029
% Think smoking dangerous	0.765	0.070	0.856	0.059
Tobacco producer state	0.113	0.317	0.115	0.319
% Urban	0.701	0.204	0.694	0.205
Unemployment rate	2.601	0.751	2.582	0.756
Cigarette price (pesos), 2000	10.332	1.059	10.325	1.063
Cigarette price (pesos), 2006	14.909	0.958	14.928	0.950
Observations	13816		18314	

Notes: There are 4,310 men and 1,510 women smokers in the sample.

Source: Own calculations based on the National Survey of Health and Nutrition, 2006, and state-level characteristics from various sources. Please see text for details.

large food stores reported monthly by the Central Bank of Mexico for 46 urban cities. State cigarette prices are averages of city prices. Real prices result by dividing nominal prices in each city by the city-price index.

4.2 Empirical Strategy

Based on the predictions of the cigarette demand function discussed in Chapter I, the econometric model for the estimation of cigarette consumption $\mathrm{C_{is}}$of individual *i* residing in state *s,* working in activity *a* can be written as:

$$C_{isa} = \alpha_0 + S_i\alpha_1 + \boldsymbol{\theta}_i\boldsymbol{\alpha}_2 + CIAPI_{sa}\alpha_3 + P_s\alpha_4 + \boldsymbol{D_s\alpha}_5 + \mu_{is} \qquad (6)$$

where the smoking variable is expressed as a function of the individual stock of cigarettes and a vector of individual attributes $\boldsymbol{\theta_i}$**,** which includes tastes and income, the Clean Indoor Policy Index $CIAPI_{sa}$ defined at the state and economic activity levels, the price of cigarettes in the state of residence P_s, and a vector of state-level characteristics $\boldsymbol{D_s}$. Denoting with * the omitted variables, the vector of personal characteristics includes: age, dummies for marital status (single*, married, divorced), a dummy for junior high school, a dummy indicating whether the individual is indigenous, a dummy variable for rural area, a set of dummy variables for economic activity (workers*, students, home-workers), a set of dummy variables for region of residence (southeast, northwest*, northeast, central and west) and individual's income I_i (expressed in thousands).

Equation (6) relates current policies and prices to contemporaneous smoking consumption holding constant the lifetime stock of cigarettes and other important individual characteristics. While the coefficient on the stock cannot be interpreted as a causal effect on smoking, including this variable does minimize the specification problem that is present in models of smoking participation and conditional demand that ignore the addictive nature of cigarette smoking. [3] Mis-specification issues in the context of smoking have been discussed by DeCicca et al. (2005, 2008b).

Identification of clean indoor air policies derives from variation across states, and within states, across activities. Individual's smoking decisions are small to affect tobacco control laws defined at the state-occupation levels, thus reducing potential endogeneity biases. Yet, if states that passed tobacco control laws are different from states that did not passed them in observable or unobservable ways, the $CIAPI_{sa}$coefficient in equation (6) would suffer from omitted variable biases. To address this issue, equation (6) includes a set of controls that should capture differences in state-level tastes for smoking. They are: a dummy indicating whether the state is a tobacco producer, a variable reflecting state-awareness of the dangers of smoking, measured as the percentage of people in the state who reported to know that smoking is either 'very dangerous' or 'dangerous', the unemployment rate, the percentage of the population in urban areas, and a set of dummies to indicate the political party with majority vote in the local. Still, the coefficient on the clean indoor air policy index should not be interpreted as a causal effect because unobserved state-level heterogeneity might cause the correlation between smoking and tobacco control policies.

Cigarette prices are averages of city-prices in each state. In a developing country like Mexico, state price variation might be the result of transportation costs and a less than fully market integration where the presence of potential arbitrage is not

[3] Alternatively, one could estimate a reduced form equation where smoking is modeled as a function of previous determinants of smoking status. This implies that current participation would be modeled as a function of the relevant history of prices, tobacco control policies and other time varying determinants of smoking. This strategy requires knowledge on the year in which the person started smoking and her migration history coming from either retrospective information or panel data. This information is not available in the survey used in this study. Indeed, it is not even possible to follow DeCicca, Kenkel and Mathios (2005) and estimate a smoking participation model that includes both current and past prices and policies because no information is readily available on the place where the person lived when he/she started smoking.

enough to equalize prices between geographical locations. As Deaton (1997) recognizes, this type of "variation in developing countries is not usually available in developed economies" (p. 283). Tax variation across states has been used to identify price effects in other countries (see, for example DeCicca et al, 2000; Gruber and Zinman, 2000; Cawley et al., 2006), but this strategy cannot be used in Mexico as taxes are set nationally (see Chapter I), Nonetheless, the use of state-level prices improves upon previous studies in Mexico where prices, calculated dividing household cigarette expenditure by consumption (the RHS variable), suffer from endogeneity biases.

I estimate different versions of equation (6) corresponding to two cigarette consumption measures. One is smoking participation, where C_{is} is a dummy variable taking the value of one if the individual smokes. The other measures the number of cigarettes a person smokes conditional upon participation. Smoking participation is the usual limited-dependent variable set-up, where one observes someone smoking only when the underlying net utility gain from doing so is positive.

I use a linear probability model (LPM) for the estimation of smoking participation and ordinary least squares (OLS) for the estimation of conditional intensity. I use LPM for ease of computation and interpretation. However, to account for the limited nature of the smoking participation variable, I also conduct the analysis of the propensity to smoke using a probit model, which completes Cragg's (1971) two-part model. This model is more flexible than its closest alternative, a Tobit model. This is because the parameters that govern the probability of observing an above-threshold realization of the dependent variable (i.e. smoking participation) are allowed to differ from the coefficients determining the quantity of cigarettes smoked, conditional on smoking.

Individuals residing in the same state are assigned identical values of the state-level variables, which generates within-group dependence of the errors. In fact, in the benchmark models that uses a dummy variable for law-enactment the number of clusters is 2. The maximum number of clusters that variation in my data generates is 38, corresponding to clean indoor air policies merged at the state and individual activity levels. To account for this issue, the standard errors reported in all specifications are cluster-robust. As Wooldridge (2003), Bertrand, Mullainathan and Duflo (2004) and Cameron et al. (2008) have noted, standard asymptotic tests in applications with few clusters tend to over-reject the null hypothesis considerably. To correct for this problem, the preferred specifications use the bootstrap procedure for determining the corrected standard errors proposed by Cameron et al. (2008). This procedure is referred to as wild cluster bootstrap-t standard error correction.

5. Results

Each table of results contains estimations for men and women. Unless otherwise indicated, all regressions include individual and state-level controls. Table 2.6 presents benchmark estimates of models of cigarette participation and conditional intensity that explore the alternative measures of smoking restrictions indoors that have been used in this literature. Each cell of this table contains, from top to bottom, the coefficient, the standard error in parentheses, and the wild cluster bootstrap p-value italicized in brackets. The stars denote significance levels based on adjusted p-values *à la* Cameron et al. (2008).

Since smoking restrictions in different locations are highly collinear, it is not possible to include all these policies in a single model. Doing so would make the coefficient estimates sensitive to model specification and increase their standard errors. Therefore, the regressions in Table 2.6 include one policy variable at a time. As

Table 2.6 Relationship between Clean Indoor Air Policy Measures Previously Used in the Literature and Smoking.

	Men		Women	
	Participation	Intensity	Participation	Intensity
Work	0.006	-0.625 *	0.004	-0.036
	(0.002)	(0.060)	(0.001)	(0.127)
	[0.19]	*[0.07]*	*[0.21]*	*[0.84]*
Schools	0.002	-0.310	0.004 *	0.213
	(0.001)	(0.036)	(0.001)	(0.117)
	[0.56]	*[0.18]*	*[0.07]*	*[0.46]*
Transportation	0.001	-0.211	0.003 *	0.251
	(0.000)	(0.126)	(0.001)	(0.083)
	[0.12]	*[0.67]*	*[0.07]*	*[0.59]*
Restaurants	0.001	-0.399 *	0.002	0.058
	(0.001)	(0.050)	(0.002)	(0.112)
	[0.22]	*[0.08]*	*[0.33]*	*[0.79]*
Retail stores	0.006	-0.516	0.005	0.330
	(0.003)	(0.113)	(0.000)	(0.088)
	[0.40]	*[0.14]*	*[0.14]*	*[0.14]*
Recreational	0.002	-0.329	0.007	0.706
	(0.004)	(0.169)	(0.002)	(0.250)
	[0.84]	*[0.62]*	*[0.20]*	*[0.57]*
Cultural	0.002 **	-0.292	0.002 *	0.116
	(0.001)	(0.072)	(0.000)	(0.068)
	[0.04]	*[0.19]*	*[0.10]*	*[0.49]*
Health care	0.000	-0.193	0.002 *	0.082
	(0.001)	(0.095)	(0.001)	(0.080)
	[0.70]	*[0.10]*	*[0.07]*	*[0.38]*
Sanctions	0.000	-0.231	0.003	0.214
	(0.001)	(0.104)	(0.001)	(0.090)
	[0.69]	*[0.38]*	*[0.20]*	*[0.19]*
Enforcement	0.000	-0.353	0.003	0.149
	(0.001)	(0.066)	(0.001)	(0.164)
	[0.62]	*[0.39]*	*[0.38]*	*[0.87]*

*Notes: Each entry comes from a regression where the dependent variable was regressed on the policy variable denoted in the first column a set of individual level characteristics (as in Table 2.8) and a set of state-level characteristics including: a dummy for tobacco producer state, a variable reflecting state-level awareness of the dangers of smoking, the unemployment rate, the percentage of the population in urban areas, state-level cigarette prices and a set of dummies to indicate the political party with majority vote in the local Congress. The coefficient is presented in the first row of each followed by the standard error in parentheses, and the adjusted wild p-value italicized in brackets. The stars reflect significance levels based on clustered wild adjusted p-values (see, Cameron et al., 2008). * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations from the National Health and Nutrition Survey, 2006.

Table 2.6 (Continued)

	Men		Women	
	Participation	**Intensity**	**Participation**	**Intensity**
Law (=1)	0.002	-0.770	0.007	0.437
	(0.000)	(0.163)	(0.002)	(0.297)
	[0.25]	*[0.25]*	*[0.25]*	*[0.74]*
Rated law	0.000	-0.039	0.000 *	0.021
	(0.000)	(0.009)	(0.000)	(0.011)
	[0.89]	*[0.14]*	*[0.07]*	*[0.22]*
Observations	13816	4301	18314	1510

*Notes: Each entry comes from a regression where the dependent variable was regressed on the policy variable denoted in the first column a set of individual level characteristics (as in Table 2.8) and a set of state-level characteristics including: a dummy for tobacco producer state, a variable reflecting state-level awareness of the dangers of smoking, the unemployment rate, the percentage of the population in urban areas, state-level cigarette prices and a set of dummies to indicate the political party with majority vote in the local Congress. The coefficient is presented in the first row of each followed by the standard error in parentheses, and the adjusted wild p-value italicized in brackets. The stars reflect significance levels based on clustered wild adjusted p-values (see, Cameron et al., 2008). * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations from the National Health and Nutrition Survey, 2006.

Ross and Chaloupka (2004) point out, these models demonstrate how individual policies affect cigarette consumption when multicollinearity is eliminated at the expense of introducing omitted variable biases. The prediction that compulsory restrictions in enclosed areas would increase the costs of smoking, and hereby decrease smoking is confirmed for workplaces and cultural restaurants among men. Surprisingly, cultural places are positively associated with smoking participation among women and men. Restrictions on smoking in schools, transportation services and health care facilities are also positively correlated with female smoking participation. Reverse causality might be behind these unexpected signs (i.e. high smoking participation among women might have driven more stringent policies).

Aggregate measures of clean indoor air policies are presented at the bottom of Table 2.6. *Law* is a dummy variable that reflects whether or not a state has enacted a clean indoor air law. *Rated* is the sum of the rated restrictions in all locations, where the maximum score in each location is 4. The evidence in this table indicates that the stringency of the law, rather than the enactment of the law, is positively associated with female smoking participation.

Based on the evidence from the regressions where policies in different locations where added one at a time, the significance of the rated law in models of female smoking participation and the lack of significance on this coefficient in models of male conditional intensity is not surprising, but is seems inappropriate. This is because, despite the fact that women spend a little time in locations that are statistically significant in the models of smoking participation, these restrictions contribute significantly to the overall rated index where all smoking limits are equally weighted. In contrast, even though the costs of smoking when smoking restrictions are imposed in workplaces—where men spend most of their time--, the significance of workplaces and restaurants in the model of male conditional intensity is offset by the lack of significance of the rest of the coefficients. The CIAPI index should correct for this problem.

Table 2.7 presents the results using the most preferred measure of compulsory restrictions; the Clean Indoor-Policy Index developed in this chapter. This index is defined at the economic activity (public/private worker, student, home worker) and state levels. Including state fixed-effects in these specifications would wipe out interesting variation across states. This is a drawback if one thinks that state and the laws are correlated. In order to alleviate this problem, I control instead for region of residence.

Panel 1 reports the coefficients on the Clean Indoor Air Policy Index from benchmark specifications that do not include state-level controls. The other panels report estimations from specifications that do include state controls. In the benchmark models, stringent smoking restrictions are negatively and statistically significantly associated with conditional cigarette demand among men, but not among women.

Table 2.7 Relationship between the CIAPI and Smoking.

	Men		Women	
	Participation	**Intensity**	**Participation**	**Intensity**
Panel 1: LPM without state controls				
CIAPI	0.032	-3.685 ***	0.028	4.152
	(0.046)	(0.658)	(0.069)	(2.532)
Panel 2: LPM				
CIAPI	0.072 **	-6.227 ***	0.016	5.780 *
	(0.035)	(1.413)	(0.051)	(3.013)
Panel 3: Probit model				
CIAPI	0.310 *		0.328	
	(0.162)		(0.342)	
	{0.105}		{ .031}	
Panel 4: LPM with SE-adjusted				
CIAPI	0.072	-6.227 **	0.016	5.780
	[0.17]	*[0.05]*	*[0.82]*	*[0.22]*
Observations	13816	4301	18314	1510

*Notes: Each entry comes from a regression where the dependent variable was regressed on the Clean Indoor Air Policy Index (CIAPI), individual level characteristics (as in Table 2.8) and a set of state-level characteristics, which includes: a dummy for tobacco producer state, a variable reflecting state-level awareness of the dangers of smoking, the unemployment rate, the percentage of the population in urban areas, state-level cigarette prices and a set of dummies to indicate the political party with majority vote in the local Congress. Results in Panel 1 are estimated via a Linear Probability Model (LPM) that does not include state-level controls. The rest of the estimates control for state-level characteristics. Panels 2 and 4presents results estimated through a LPM and Panel 3 via a Probit Model. Panel 4 reports estimates where the standard errors are adjusted for the 'small number of clusters'. Standard errors are in parentheses, marginal effects in curly brackets, and the adjusted wild p-value italicized in brackets. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations from the National Health and Nutrition Survey, 2006.

In Panel 2, when state controls are included in linear probability models of smoking participation and OLS models of conditional intensity, the magnitude of all, but the coefficient on smoking participation among women, increases. Moreover, the estimated coefficient of the CIAPI index in the model of smoking participation among men, and conditional intensity among women become statistically significant. This suggests that the omission of important state controls induces a downward bias on the CIAPI coefficient.

Panel 3 presents comparable results for models of smoking participation that are estimated using a probit model. Each cell contains, from top to bottom, the coefficient, the standard errors in parentheses, and the marginal effects evaluated at the mean of the independent variable in curly brackets. As can be readily seen from this table, using a probit model for the estimation of smoking participation leads to similar qualitative conclusions: the positive marginal effect of the stringency of the laws among men and the non-statistical effect among women indicate that smoking participation increases with the stringency of the laws among men, but not among women.

In Panel 4 of Table 2.7 I present the results of models of smoking participation and conditional intensity estimated using the wild cluster bootstrap-t standard error procedure. This is the most preferred estimation because it corrects for the small number of clusters. Since this method is not yet available for non-linear models, smoking participation is estimated using a LPM. When standard errors are corrected, only the negative and significant association between stringent laws and the intensity of smoking among men remains. In particular, a one unit increase in the CIAPI is associated with 6.2 fewer cigarettes smoked per day. Therefore, living in a state with no clean indoor air law compared to living in Zacatecas, currently the state with the most stringent law (0.3032), is associated with 1.9 more cigarettes smoked per day.

The estimates for the demographic and socioeconomic determinants of cigarette demand of the preferred model (Panel 4, Table 2.7) are presented in Table 2.8. As in Table 2.6, each cell of this table contains, from top to bottom, the coefficient, the standard error in parentheses, and the wild cluster bootstrap p-values italicized in brackets. The stars denoting significance levels are based on wild p

Table 2.8 Determinants of Cigarette Demand.

	Men		Women	
	Participation	Intensity	Participation	Intensity
Cig. stock	0.043 ***	0.217 ***	0.046 ***	0.092
	(0.001)	(0.015)	(0.001)	(0.025)
	[0.01]	*[0.00]*	*[0.01]*	*[0.11]*
Age	-0.011 ***	0.004	-0.003 ***	0.017
	(0.001)	(0.020)	(0.000)	(0.016)
	[0.00]	*[0.87]*	*[0.00]*	*[0.35]*
Indigenous	-0.012	2.611	-0.002	-1.126
	(0.014)	(1.219)	(0.005)	(0.496)
	[0.48]	*[0.52]*	*[0.78]*	*[0.11]*
Married	-0.005	-0.006	-0.012	-0.089
	(0.010)	(0.370)	(0.006)	(0.307)
	[0.67]	*[0.96]*	*[0.13]*	*[0.85]*
Divorced	0.041	1.127	0.035	0.342
	(0.015)	(0.776)	(0.008)	(0.291)
	[0.13]	*[0.17]*	*[0.13]*	*[0.51]*
Middle school	-0.043 ***	-0.391	0.022 **	-0.206
	(0.007)	(0.404)	(0.007)	(0.189)
	[0.01]	*[0.92]*	*[0.03]*	*[0.37]*
Income	-0.002	0.537	0.014	-0.251
	(0.009)	(0.219)	(0.014)	(0.287)
	[0.96]	*[0.29]*	*[0.74]*	*[0.52]*
Student	-0.006	-0.157	0.042	0.481
	(0.021)	(0.279)	(0.031)	(0.303)
	[0.77]	*[0.70]*	*[0.91]*	*[0.51]*
Home worker	-0.015	0.519	-0.012	0.896
	(0.013)	(0.369)	(0.004)	(0.239)
	[0.32]	*[0.52]*	*[0.24]*	*[0.25]*
Oportunidades	-0.020 *	-1.214	-0.004	0.075
	(0.008)	(0.290)	(0.002)	(0.514)
	[0.09]	*[0.12]*	*[0.38]*	*[0.98]*
Pregnant			-0.055 ***	-0.049
			(0.010)	(1.378)
			[0.00]	*[0.98]*

Notes: Details on omitted and other independent variables included are on the next page at the end of this table.

*Each cell contains, from top to bottom, the coefficient, the standard error in parentheses, and wild p-value italicized in brackets. The stars reflect significance levels based on clustered wild adjusted p-values (see, Cameron et al., 2008). * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Table 2.8 (Continued)

	Men				Women			
	Participation		Intensity		Participation		Intensity	
Rural	-0.026		-1.072		-0.021	**	-0.275	
	(0.010)		(0.224)		(0.003)		(0.371)	
	[0.58]		*[0.19]*		*[0.02]*		*[0.53]*	
Northeast	0.073	**	-1.917		0.017		0.366	
	(0.010)		(0.349)		(0.011)		(1.375)	
	[0.02]		*[0.14]*		*[0.27]*		*[0.77]*	
Southeast	0.051		-4.919	*	0.007		-0.620	
	(0.018)		(0.286)		(0.010)		(0.956)	
	[0.14]		*[0.06]*		*[0.59]*		*[0.74]*	
Center	0.040	**	-3.709	**	0.008		-1.371	
	(0.008)		(0.288)		(0.010)		(0.516)	
	[0.03]		*[0.02]*		*[0.82]*		*[0.16]*	
West	0.015		-2.357		-0.002		0.026	
	(0.014)		(0.269)		(0.010)		(0.731)	
	[0.21]		*[0.12]*		*[0.76]*		*[0.96]*	
R-squared	0.40		0.12		0.45		0.06	
Observations	13816		4301		18314		1510	

Notes: The omitted variables are single, public/private workers and northwest.

Independent variables in these regressions include the CIAPI, and state-level characteristics: a dummy for tobacco producer state, a variable reflecting state-level awareness of the dangers of smoking, the unemployment rate, the percentage of the population in urban areas, cigarette prices and a set of dummies to indicate the political party with majority vote in the local Congress. Smoking participation models are estimated through a LPM and conditional intensity models via OLS.

*Each cell contains, from top to bottom, the coefficient, the standard error in parentheses, and wild p-value italicized in brackets. The stars reflect significance levels based on clustered wild adjusted p-values (see, Cameron et al., 2008). * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

values. The first part of the table reports the estimates on individual-characteristics. The stock of cigarettes is associated with a higher probability of being a smoker for both men and women. This result is consistent with the addictive nature of cigarettes, but may not be interpreted as a causal effect. This is because the stock of habits might be correlated with previously determined market-level characteristics as well as unobservable individual characteristics.

Controlling for the stock of habits, people are less likely to smoke as they age. One of the most striking results is that, conditioning on other relevant factors, having a middle school diploma decreases the probability of male smoking, but decreases the probability of female smoking. These findings, however, conform to the profile of smoking presented in the introduction of this dissertation. Participation in *Oportunidades*, a federal social program providing income and health information sessions to adults, decreases the probability of smoking for men, but the coefficient is barely significant. Finally, pregnant women are less likely to smoke, but being pregnant does not alter the number of cigarettes smoked for women who smoke.

Compared to men who live in the Northwest, men residing in the Northeast and Central regions are more likely to smoke, whereas men living in the Southeast and Central regions smoke with lower intensity once they decide to pursue the habit. The probability of smoking is higher for women living in urban versus rural areas.

Individual and socio-demographic variables do not predict the number of cigarettes people smoke. With the exception of the CIAPI and the stock of cigarettes none of the coefficients in the regressions of conditional intensity of smoking among male are significant. Furthermore, the R-squares of these models are particularly low.

The findings in this chapter can be summarized as follows. First, the exclusion of observable characteristics bias downwards the coefficient on the stringency of the laws. Second, ignoring that the regressions include only a small number of clusters leads to an over-rejection of the null hypothesis that clean indoor air policies affect smoking decisions. Third, once state-level variables are controlled for and the standard errors are adjusted I find a negative and significant association between more stringent laws and the number of cigarettes smoked by men. Fourth, the predictors of smoking participation by men and women differ significantly. And fifth, other than the clean indoor air laws and the stock of cigarettes, none of the other individual or state-level variables predict the number of cigarettes smoked by men and women who smoke. The next section explores the possible causes of these findings and contrasts the results with those found in the previous literature.

6. Discussion

Using variation of clean indoor air policy indices across activities and states, the conclusion that emerged in the previous section is that clean indoor air policies and the number of cigarettes smoked by men are negatively correlated. This finding is consistent with the prediction that compulsory restrictions in enclosed areas increase the costs of smoking and thereby decrease cigarette consumption. This is not, however, the only possible explanation for this correlation.

An important question that arises in interpreting the results of the previous section is, what is the political economy behind the enactment of clean indoor air policy laws? There are many potential explanations. States that enacted stringent or lax laws compared to states that did not pass a law at all might vary in various ways. Variation might occur in terms of the public awareness of the health and economic consequences of smoking, public sentiment towards smoking, state-level cigarette price determination, economic progress and political forces. The later issue could have been particularly important in Mexico were the last decade brought extraordinary changes to the Mexican political system. Once an authoritarian regime controlled by the hegemonic Institutional Revolutionary Party (PRI), party-switching at the state-level started in 1989, which ultimately paved the way for Fox's election in 2000 running on the ticket of the center-right PAN.

Stringent laws may have resulted from low levels and/or intensities of cigarette smoking rather than caused them. Lower levels of smoking would facilitate the passage of the laws, since they would encounter less opposition. Conversely, more stringent laws could have attempted to decrease high levels of smoking.

I predict the enactment of clean indoor air laws and the stringency of these laws, measured by the CIAPI, as a function of various state-level characteristics and gender-specific prevalence and intensity of smoking, all of them measured prior to the passage of the laws. State controls include economic variables, such as the unemployment rate and a set of dummies for region reflecting differences in purchasing power; a set of dummies indicating the political party holding the majority vote in the local Congress (*PRI, PAN, PRD*); and multiple tobacco related variables: a measure of state-level awareness regarding the health-related consequences of smoking, a dummy for state tobacco producer, and cigarette prices.

The factors that predict the enactment (Law) and stringency (CIAPI) of the laws are reported in table 2.9. The unemployment rate is the only state-level variable that is negatively associated with both the enactment and the stringency of the laws. Additionally, the urban population share and the *PRD* in Congress are negatively correlated with law enactment in the regression for women. The signs of the tobacco-related variables are unexpected, but none of these variables are significant.[4]

The last two rows of table 2.9 show evidence that the prevalence of smoking affected positively and significantly the enactment of the laws, but not the stringency of the laws. In contrast, the average number of cigarettes is not found to significantly affect the passage of the laws. This suggests that, while policy endogeneity is a potential problem in the estimation of smoking participation, it is less of a problem in the estimation of the number of cigarettes smoked.

---

4 The results are similar if the states that passed laws before 2000 (Aguascalientes, Baja California and Tlaxcala) are excluded from the regressions. Results are available from the author upon request.

Table 2.9 Preset Factors that Predict the Enactment and Stringency of Clean Indoor Air Laws.

	Men		Women	
	Law (=1)	**CIAPI**	**Law (=1)**	**CIAPI**
Economic Variables:				
% Urban	-0.655	-0.022	-1.458 **	-0.061
	(0.745)	(0.039)	(0.612)	(0.036)
Unemployment rate	-0.292 *	-0.015 *	-0.283 **	-0.014 **
	(0.150)	(0.008)	(0.125)	(0.007)
Northeast	-0.347	-0.020	-0.262	-0.017
	(0.462)	(0.021)	(0.360)	(0.019)
Southeast	-0.579	-0.032	-0.502	-0.031
	(0.628)	(0.031)	(0.736)	(0.036)
Central	-0.168	-0.019	0.050	-0.011
	(0.613)	(0.030)	(0.714)	(0.039)
West	-0.655	-0.030	-0.463	-0.022
	(0.419)	(0.021)	(0.509)	(0.024)
Political variables:				
PAN in Congress	-0.025	0.008	-0.005	0.009
	(0.338)	(0.018)	(0.355)	(0.019)
PRD in Congress	-0.617	-0.025	-0.699 *	-0.030
	(0.394)	(0.020)	(0.372)	(0.019)
Tobacco-related variables:				
% Smoking danger.	-0.890	-0.038	-1.219	-0.057
	(2.039)	(0.097)	(2.403)	(0.112)
Tobacco producer	0.408	0.015	0.200	0.006
	(0.411)	(0.018)	(0.305)	(0.012)
Cigarette price	0.016	0.003	0.031	0.004
	(0.111)	(0.005)	(0.114)	(0.006)
% Smokers[&]	4.817 *	0.225	10.397 **	0.509
	(2.483)	(0.130)	(4.226)	(0.295)
Avg. no. cigarettes[&]	-0.084	-0.004	-0.542	-0.028
	(0.143)	(0.006)	(0.770)	(0.041)
R-squared	0.41	0.39	0.50	0.49
Observations	32	32	32	32

Notes: The omitted variables Northwest and the PRI party.

All independent variables are measured prior to the passage of the law.

Standard errors are in parentheses.

All in all the evidence in this section indicates that neither observable characteristics nor initial intensity of smoking were behind the enactment of more stringent laws. This suggests that the laws might in fact have cause a decrease in the number of cigarettes smoked by men. It is important to note, however, that the inclusion of state controls increased the estimated coefficients of the CIAPI and changed its significance. If states self-selected to law passage based on characteristics that I cannot observe, such as the anti-smoking sentiment discussed by DeCicca et al. (various years), the conclusion that the estimates in the previous section are bias-free would be inaccurate. For that reason, I neglect to interpret the results above as casual.

The evidence of cross-sectional studies relating tobacco measures with smoking participation and conditional intensity in the U.S. is mixed. For example, Ross and Chaloupka (2004), who use an index that adds up dummy variables, each representing full restrictions in a given location, found that clean indoor air restrictions reduce smoking participation but have not effect on smoking intensity. In contrast, Wasserman et al. (1991) found that more stringent regulation has a consistently strong negative influence on the number of packs smoked. This conclusion was based on an index that considered, albeit roughly, that more comprehensive laws should discount restrictions in relatively unimportant places and give more credit to regulations that affect a large share of the population. The choice of index might explain why the conclusions reached by Wasserman and his colleagues in the U.S. are similar to the ones in this chapter.

7. Conclusions and Further Research

The indices of smoking restrictions indoors that have been used in previous research are admittedly imperfect measures of tobacco control policies for at least two reasons. First, creating the index variables has required somewhat arbitrary assumptions. Second, the indices do not reflect the level of enforcement of the laws (DeCicca et al., 2002).

The first main focus of this study was to develop a new index of smoking restrictions, the Clean Indoor Air Policy Index (*CIAPI*), which advances previous indices in two main aspects. First, regulations in locations where most of the people spend little time have less of an impact to the overall index. And second, the *CIAPI* index is penalized as differences in the actual and desired level of enforcement increase. Additionally, the strictness of the law is calculated for four population subgroups, allowing comparisons across types of individuals and geographic units. In the case of Mexico, the data enable distinguishing between private workers, public workers, students and individuals who work at home. This index is not without limitations, but it solves the problem of multicollinearity of the policies and summarizes the overall impact of clean indoor air policies relying on more credible assumptions than previous indices.

Clean Indoor Policy Indices in Mexico may be improved in various ways. Future efforts might conjoin statutory and regulatory rankings to quantify the restrictive nature of statewide clear indoor air policies. Ordinances at the municipal level should be codified and ranked in order to evaluate local restrictions. When doing that, it would be important to account for state preemption, which occurs when local areas are not allowed to enact smoke-free ordinances that are more rigorous than those of the state. Finally, *de Jure* enforcement might be replaced by a measure of *de facto* enforcement. This has been done in the U.S. by Chaloupka and Ross (2004).

A second goal of this study was to analyze the role of clean indoor air policies --and other variables-- as determinants of cigarette consumption. Compared to previous studies, the estimates in this chapter were refined by using individual level data, rather than household-level data. Using cross-sectional data for 2006, I found that more stringent policies are associated with less smoking intensity among men. The fact that more stringent laws were not found to be determined by observable characteristics or initial levels of smoking intensity provided suggestive, but not definite, evidence in favor of the hypothesis that the laws changed the individuals' smoking decisions.

The cross-sectional data used in this study did not allow me to explore whether unobserved state-level characteristics might be causing both the passage of stringent laws and the decrease in the quantity of cigarettes consumed. The ideal data to address unobserved heterogeneity issues would be to have panel data of individuals residing in different states for whom smoking decisions are observed before and after the law enactment in their state of residence. The only Mexican dataset that has followed individuals over time is the Mexican Family Life Survey (MxFLS). This survey was not used because the second round became available when this study was already finished. Even though this survey is representative at a national level, not all states where interviewed in this survey. None of the states that had enacted laws before 2002 were included in the MxFLS sample. 12 of the states included in the sample did not have changes between 2002 and 2005, and four states did. Only 8 Mexican states enacted laws between the baseline survey in 2002 and the follow-up in 2006. Exploiting the longitudinal nature of this data in a regression framework that includes individual and state fixed-effects would shed some light on the role of unobserved heterogeneity and pave the way for the estimation of causal effects of tobacco control laws on smoking in Mexico.

Besides the stock of cigarettes and the clean indoor air policies, none of the individual characteristics explain differences in the number of cigarettes smoked by men or women. However, I found significant gender differences in the models of smoking participation. Two findings are noteworthy. First, participation in *Oportunidades*, the largest anti-poverty program in Mexico, decreases the probability of cigarette participation among men. Even though this program did not intend to affect smoking, the benefits of this program, inclusive of cash transfers and health information sessions, might have altered individual's decisions to smoke. The next chapter explores this possibility. Second, even though schooling is negatively associated with smoking among men, women with middle school diplomas are more likely to smoke. Differences in the schooling-smoking gradient by gender are intriguing and worth exploring. This will be done in the third essay of this dissertation.

APPENDIX

APPLICATION OF THE CIAPI TO MEXICAN DATA

1. Rating of legislation

The first step in the calculation of the *CIAPI* involved coding and rating state smoke-free laws. The coding was done using Atlas Ti V. 5.2.; a commonly used software for qualitative data analysis. Only statutory laws were considered in the present analysis (i.e. the norms with the highest rank in the Mexican legal system). Nonetheless, clean indoor air policies are also in the form of regulations (Ibañez, 2005).

I coded state-level statutory laws in effect by May of 2006, which was the last interview month of the ENSANUT, 2006. By that date, 13 states had laws restricting smoking. Three of these states had already promulgated laws when the National Health Survey 2000 was carried out and the rest passed them between 2001 and 2006.[5] In 12 of the 13 states the legislation analyzed were 'Laws for the Protection of Non-Smokers. In Chihuahua, tobacco control restrictions are considered in the Environmental Protection Law. In most cases, the stated goal of the law is to protect people from second hand smoking. Only Mexico's City law states explicitly the objective of reducing cigarette consumption.

[5] As of April 2008 eight other states had promulgated laws for a total of 21 states with laws restricting or prohibiting smoking in certain places. Tamaulipas enacted the law in December of 2005 (one month after the ENSANUT of 2006 was carried out in that state), Durango, and San Luis Potosi promulgated laws in late 2006. Coahuila, Veracruz, Quintana Roo, Yucatán and Sinaloa enacted laws in 2007. Additionally, Aguascalientes and Mexico City reformed their laws in October of 2006 and the beginning of 2008, respectively.

Tobacco partial and full bans were only coded when the laws were unambiguous in their application. The following standard coding scheme was employed for most of the locations:

0 No provision/no meet a restriction

1 Restrict smoking to designated areas or require separate ventilation with exemptions

2 Restricts smoking to separately ventilated areas

3 No smoking permitted unless restricted to enclosed, separately ventilated designated smoking areas

4 Ban at all times (i.e. health policy goal)

The codebook is presented in table A2.1. A summary of the specific decision rules in each location follows.

Government workplaces refers to buildings of the three branches of the Government (legislative, executive and judicial), or all the buildings of the Public Administration. No points were deducted if the law did not explicitly mention that the law applied to autonomous agencies like the Central Bank of Mexico. Private workplaces include banks, financial, commercial and industrial offices and offices of firms producing goods and services

In the case of schools, laws that did not explicitly indicate whether private schools were or were not included, were coded as having restrictions in both public and private schools. Laws that did not explicitly mentioned prohibitions in college and universities were severely penalized. The same was true when restrictions only applied to classrooms. Some legislation considered classrooms, laboratories, faculty and administrative offices etc.

Public transportation includes public transit and inter-city buses. Public transit includes bus, light rail and subway.

Table A2.1 Codebook of Law Rating.

Variable	Description	Value	Coding Explanation
govXX	**Gov't worksites by year (XX)**	4	Government worksites are 100% smoke-free, no exemptions
		3	No smoking permitted in government worksites unless restricted to enclosed, separately ventilated designated smoking areas
		2	i) Smoking in government worksites restricted to designated smoking areas which are separate and enclosed or ii) Smoking in separately ventilated designated smoking areas at government worksites, with a minimal exemption
		1	i) Smoking in government worksites restricted only to designated smoking areas or ii) Any stricter requirement that applies to some but not all types of worksites (e.g., worksites with access to the public)
		0	No restrictions, or requirement(s) that smoking be permitted
pvtXX	**Private worksites by year (XX)**	4	All rooms in all private worksites are 100% smoke-free, no exemptions
		3	Offices in all private worksites are 100% smoke-free, with a minimal exemption (e.g., other rooms in the building are not considered)
		2	i) Smoking in private worksites restricted to designated smoking areas which separately ventilated in at least two types of offices or ii) any stricter requirement applying with a minimal exemption (e.g. worksites with access to the public) to at least one type of worksites
		1	i) Smoking in private worksites restricted to designated smoking areas in at least two types of offices or ii) Any stricter requirement that excludes completely at least one type of office (e.g., no provisions for warehouses)
		0	No restrictions
schXX	**Schools by year (XX)**	4	Private and public schools including colleges and universities are 100% smoke-free.
		3	No smoking in schools is permitted with minimal exemption(s) (e.g., colleges and universities have separately ventilated designated smoking areas)
		2	i) No smoking in classrooms is permitted (college and university classrooms included) or ii) No smoking in school is permitted, but this requirement excludes colleges and universities
		1	i) Smoking in classrooms restricted only to designated smoking areas, or ii) Any stricter requirement with more than minimal exception(s) (e.g., all private schools explicitly exempted)
		0	No restrictions

Table A2.1 (Continued)

Variable	Description	Value	Coding Explanation
hosXX	**Health Care Facilities by year (XX)**	4	Health care facilities are 100% smoke-free
		3	Smoking in health care facilities restricted to enclosed separately ventilated smoking areas
		2	Smoking in health care facilities restricted to separately ventilated smoking areas
		1	i) Smoking in health care facilities restricted to designated smoking areas or ii) Any stricter requirement that applies to only one type of health care facility (e.g. excludes clinics or private hospitals)
		0	No restrictions
Note: *Health care facilities include nursing homes, hospitals and health centers, public and private.*			
tranXX	**Transport by year (XX)**	4	Public transportation is 100% smoke-free including rail, bus services, cabs, ferries and inter-city buses, and requires sign posting.
		3	Public transit is 100% smoke-free including rail, bus services, taxi cabs and ferries, and requires sign posting
		2	Smoking is restricted to enclosed, separately ventilated public transit units including rail, bus services, taxi cabs and ferries, and requires sign posting
		1	i) Smoking is restricted to separately ventilated public transit units including rail, bus services, taxi cabs and ferries, and requires sign posting or ii) any stricter requirement that only applies to some public transit (e.g. taxi drives decide whether they allow smoking in the car)
Note: *Public transportation includes public transit. That is, public transit excludes inter-city buses*			
restXX	**Restaurant by year (XX)**	4	Restaurants (explicitly including bar areas of restaurants) are 100% smoke-free
		3	No smoking permitted in restaurants (including bar areas of restaurants) unless restricted to enclosed separately ventilated designated smoking areas
		2	Smoking in restaurants restricted to separately ventilated designated smoking areas
		1	i) Smoking in restaurants restricted only to designated smoking areas or ii) Restrictions on smoking that apply to some but not all restaurants (e.g., size exemptions)
		0	No restrictions

Table A2.1 (Continued)

Variable	Description	Value	Coding Explanation
retXX	**Retail stores by year (XX)**	4	Retail stores or businesses open to the public are 100% smoke-free
		3	No smoking permitted in retail stores or businesses open to the public unless restricted to enclosed, separately ventilated designated smoking areas
		2	i) Smoking in retail stores or businesses open to the public restricted to designated smoking areas which are separately ventilated or, ii) Bans smoking in either all retail stores or all businesses open to the public and restricts smoking to enclosed, separately ventilated areas in some of the others (e.g. smoking is permitted in separately ventilated areas of commercial businesses open to the public and banned in all other businesses).
		1	i) Smoking in retail stores or businesses open to the public restricted to designated smoking areas, or ii) Any stricter law that applies to either retail stores or businesses open to the public
		0	No restrictions.
recXX	**Recreational/sport facilities by year (XX)**	4	Recreational and sport facilities are 100% smoke-free
		3	i) Restricts smoking to enclosed, separately ventilated areas in all recreational and sports facilities or ii) Bans smoking in either all sports or recreational facilities and restricts smoking to separately ventilated designated smoking areas in all other facilities, or iii) Bans smoking in either all sports or recreational facilities and restricts smoking to enclosed separately ventilated designated smoking areas in some of the other facilities.
		2	i) Smoking in all recreational and cultural facilities restricted to separately ventilated smoking areas or ii) Bans smoking in either all recreational or sport facilities and restricts smoking to designated smoking areas in all the other facilities or iii) Bans smoking in either all recreational or sport facilities and restricts smoking to separately ventilated designated smoking areas in some of the other facilities
		1	i) Smoking in either of all recreational or cultural facilities restricted only to designated smoking areas or ii) Bans smoking in either all sports or recreational facilities or iii) Any stricter requirement that applies to only some of the recreational and/or cultural facilities
		0	No restrictions

Table A2.1 (Continued)

Variable	Description	Value	Coding Explanation
culXX	**Cultural facilities by year (XX)**	4	i) Restricts smoking to enclosed, separately ventilated designed smoking areas in all cultural facilities or ii) Bans smoking in at least four cultural facilities and restricts smoking to separately ventilated designated smoking areas in the others
		3	i) Restricts smoking to separately ventilated designed smoking areas in at least four cultural facilities and restricts smoking to separately ventilated smoking areas in the other or ii) bans smoking in at least three cultural facilities and do has some type of less strict requirements on the others
		2	i) Restricts smoking to separately ventilated designated areas in three cultural facilities and restricts smoking to designated areas in at least other cultural facility or ii) Bans smoking in at least three cultural facilities imposing no restrictions on the others
		1	i) Restricts smoking to designated areas in at least four cultural facilities or ii) Bans smoking in at least two cultural facilities imposing no restrictions on the others or iii) Restricts smoking to separately ventilated designated areas in at least three cultural facilities imposing no restrictions on the others
		0	i) Restricts smoking to designated areas in three or less cultural facilities or ii) No restrictions, or requirement(s) that smoking be permitted

Note: *All cultural facilities include museums, auditoriums, indoor movie theaters, theaters and libraries.*

Variable	Description	Value	Coding Explanation
sanXX	**Sanctions by year (XX)**	4	Graduated penalties or fines, applicable to smokers, proprietors and employers, for repeated violation of clean indoor air legislation. In addition, the law has some provisions for infractions not considered in the law.
		3	i) Graduated penalties or fines, applicable to all types of infractors (smokers, proprietors and vehicle owners) on some clean indoor air provisions with penalties or fines applicable to the other clean indoor air provisions or ii) Graduated penalties or fines, applicable to some of the infractors (either smokers, proprietors and vehicle owners) on all clean indoor air provisions with penalties or fines applicable to the other type of infractors. In addition, the law explicitly considers provisions for infractions that are "not considered in the law" in either i) or ii)

Table A2.1 (Continued)

Variable	Description	Value	Coding Explanation
		2	i) Penalties or fines, applicable to all types of infractors (smokers/proprietors/vehicle owners) and all specified clean indoor provisions orii) Penalties or fines applicable to some of the infractors or some of the clean indoor provisions if and only if the law considers provisions for infractions that are "not considered in the law".
		1	Penalties or fines, applicable to either smokers, proprietors or vehicle owners to only a limited number of specified clean indoor air provisions
		0	None of the above
ageXX	**Agency by year (XX)**	4	Designate an enforcement authority for clean indoor air legislation, and require the agency or other officials to conduct ordinary and random compliance inspections in businesses and vehicles and require sign posting.
		3	i) Designate an enforcement authority for clean indoor air legislation, and require the agency or other officials to conduct ordinary compliance inspections in businesses and vehicles and require sign posting or ii) Designate an enforcement authority for clean indoor air legislation, and require the agency or other officials to conduct ordinary and random compliance inspections in businesses and vehicles but no requirement for sign posting.
		2	i) Designate an enforcement authority for clean indoor air legislation and require sign posting or ii) Designate an enforcement authority for clean indoor air legislation and require the agency or other officials to conduct ordinary compliance inspections in either businesses or vehicles but no requirement for sign posting
		1	Designate an enforcement authority for clean indoor air legislation, but no requirement for sign posting or inspections.
		0	No provision
totXX	**Total score by year (XX)**		

Some states have smoking restrictions in restaurants or bars. States were given no credit for restrictions in bars-- ‘places where people dance’--, if they did not specify whether smoking was banned in the whole building or only in specific areas where people actually dance. Retail stores include self-service stores, malls and other businesses that are open to the public.

Most of the legislation explicitly referred to prohibitions or restrictions on all recreational facilities and/or all sport facilities. When they did not, an exhaustive list of examples was needed in order to consider prohibitions in all recreational or all sports facilities. More specifically, laws that mentioned prohibitions in gyms and/or parks were given the lowest possible score.

Cultural facilities include museums, auditoriums, indoor movie theaters, theaters and libraries. Health care facilities include both private and public hospitals and health centers, nursing homes and clinics.

Legislation that did not consider how to sanction “infractions not considered in the law” received at most two points. If penalties or fines where applicable to either of smokers/proprietors/public vehicle owner and/or only applicable to certain clean indoor air provisions, two points were given at most. Graduated fines or penalties were given credit only when the law defined: i) the terms of repeated violations (e.g. repeated offense within 6 months of the first offense), and ii) the term “serious infraction”.

Table A2.2 reports the scores by item and state. The number of public places with restrictions varies from 7 to 9. Most of the states with laws in effect have some sort of regulations on all public places with the exception of recreational activities and smoking at home.

Table A2.2 Scores of Statutory Laws in Mexico by State and Item.

code	state	gov06	pvt06	sch06	hwk06	hos06	tran06	res06	ret06	rec06	cul06	hom06	san06	age06	tot06
01	Aguascalientes	1	3	2	0	4	3	1	1	0	2	0	1	2	20
02	B. C. N	4	1	2	0	4	1	4	0	0	2	0	2	3	23
03	B. C.S	0	0	0	0	0	0	0	0	0	0	0	0	0	0
04	Campeche	0	0	0	0	0	0	0	0	0	0	0	0	0	0
05	Coahuila	0	0	0	0	0	0	0	0	0	0	0	0	0	0
06	Colima	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07	Chiapas	1	3	2	0	4	3	1	1	0	2	0	1	2	20
08	Chihuahua	1	0	3	0	1	3	1	1	3	1	0	3	2	19
09	Distrito Federal	3	1	2	0	3	3	3	2	0	4	0	2	2	25
10	Durango	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Guanajuato	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	Guerrero	1	1	2	0	4	3	1	1	0	2	0	1	2	18
13	Hidalgo	1	3	2	0	4	3	1	1	0	2	0	2	2	21
14	Jalisco	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	México	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	Michoacán	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	Morelos	3	2	2	0	3	1	3	1	1	3	0	2	2	23
18	Nayarit	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	Nuevo León	4	1	2	0	2	3	2	2	1	3	0	4	2	26
20	Oaxaca	1	1	2	0	1	1	1	0	1	1	0	1	3	13
21	Puebla	2	2	4	0	2	1	1	2	2	3	0	1	2	22
22	Querétaro	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	Quintana Roo	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	San Luis Potosí	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	Sinaloa	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	Sonora	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	Tabasco	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	Tamaulipas	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	Tlaxcala	1	2	2	0	1	1	1	0	0	1	0	1	2	12
30	Veracruz - Llave	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	Yucatán	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	Zacatecas	4	3	2	0	4	3	1	1	0	2	0	1	2	23

Table A2.3 presents the frequency of individual ratings (0-4 points) for 11 *strictness* items and 2 *enforcement* items. Smoking is more restrictive in hospitals, schools and transportation services. The item-specific means are low because many of the states allow smoking in separately ventilated areas. The regulations in government are stricter than in private workplaces. In fact, 3 states ban smoking fully in government workplaces. Sanctions vary significantly across states; only in five states the legislation considered very strict sanctions. In general, enforcement is delegated to a state agency, normally the state Ministry of Health. Most of the states have mild provisions regarding monitoring and law enforcement. Enforcement is handled passively rather than actively; the responsible agency responds to complains, but does not actively monitor policy compliance by surveying worksites, restaurants or public places.

Table A2.3 Frequency of Individual Clean Indoor Air Regulations.

Item	Rating				
	1	2	3	4	Mean
Restrictions in locations					
Gov't workplaces	7	1	2	3	0.84
Private workplaces	5	3	4	0	0.72
Schools	0	11	1	1	0.91
House - Work	0	0	0	0	0.00
Transportation	5	0	8	0	0.91
Restaurants	9	1	2	1	0.66
Retail Stores	7	3	0	0	0.41
Recreational Facilities	3	1	1	0	0.25
Cultural Facilities	3	6	3	1	0.88
Health Care Facilities	3	2	2	6	1.16
Home - Relax	0	0	0	0	0.00
Enforcement items					
Sanctions	3	4	1	5	1.06
Agency	0	11	2	0	0.88

Notes: The horizontal summation of ratings is13 at the most, which equals the number of states that passed laws by May 2006. None of the states consider restrictions at home.

Source: Own calculations based on Mexican statutory state legislation

1.2 Construction of weights

The weights were estimated using the Time Use Survey, 2002 (*ENUT*, in Spanish). This is a supplement to the National Survey of Income and Expenditure providing estimates of how Mexicans spend their time (e.g. eating, relaxing, and working). It also provides information on the full range of nonmarket activities, from childcare to volunteering. To be consistent with the sample used in the study of cigarette consumption, the weights were calculated using the sample of people aged 20 to 50 years.

The proportion of people in the four occupational groups (public or private worker, student or house-worker) and the average time they spend at different locations by gender is presented in Table A2.4. At a national level, the percentage of females working at home (50%) is considerably higher than that of males (5.3%).[6] Consequently, the share of time that people spend at different locations varies by gender (see Table A2.4, column All). Women spend most of their weekly time at home (76%), either working or relaxing. Only 16% of their time is spent at work or at school. In contrast, men spend about the same proportion of their weekly time at work and at home (45% and 41%, respectively). Therefore banning smoking at work increases disproportionately the non-monetary costs of smoking for men compared to women. On the other hand, allowing smoking at home has a greater impact on women as, compared to men, they spend 30% more of their time at home. While at home, the type of activities performed also varies by gender. Women work at home 5 times as much as men do. In contrast, men use about 4% more of their time relaxing compared to women. This pattern prevails across the four groups defined according to their economic activity.

---

[6] Pedrero (2004) discusses in detail the fact that the main economic activity of Mexican women is to perform domestic work at home.

Table A2.4 Percentage of Weekly
Time Spent at Different Locations by Gender and Economic Activity.

Percentage of men	All		Public worker		Private worker		Student		House worker	
	100		10.4		81.3		2.90		5.4	
	% time	R	% time	R	% time	R	% time	R	% time	R
Gov't work	4.85	5	45.1	1	0.00	11	0.76	10	2.90	5
Private work	40.2	1	0.00	11	47.7	1	5.94	4	22.7	2
School	1.53	8	2.33	6	0.58	9	27.6	2	0.36	11
Housework	6.89	3	6.83	3	6.81	3	6.48	3	8.39	3
Transportation	5.70	4	5.63	4	5.90	4	5.09	5	3.14	4
Restaurant	0.85	10	0.98	9	0.77	8	1.03	9	1.68	8
Retail stores	1.63	7	2.07	7	1.58	6	1.14	8	1.75	7
Recreational	3.03	6	3.73	5	2.99	5	2.97	6	2.28	6
Cultural	1.05	9	1.03	8	0.96	7	2.60	7	1.56	9
Health center	0.13	11	0.09	10	0.12	10	0.05	11	0.43	10
Home-relax	34.1	2	32.2	2	32.6	2	46.3	1	54.8	1
	100		6.58		40.4		2.98		50.0	
Percentage of Women	All		Public worker		Private worker		Student		House worker	
	100		6.58		40.4		2.98		50.0	
	% time	R	% time	R	% time	R	% time	R	% time	R
Gov't work	2.24	7	29.2	1	0.00	11	0.96	10	0.58	8
Private work	13.0	3	0.00	11	27.4	2	5.92	4	3.57	5
School	1.37	8	1.78	7	0.98	7	22.3	2	0.39	10
Housework	42.0	1	28.4	2	34.7	1	21.4	3	51.0	1
Transportation	2.48	6	4.49	4	3.29	4	3.63	5	1.50	6
Restaurant	0.66	10	0.93	9	0.73	9	0.88	9	0.56	9
Retail Stores	3.23	4	2.91	6	2.85	5	2.00	7	3.66	4
Recreational	3.07	5	3.01	5	2.45	6	1.50	6	3.67	3
Cultural	0.75	9	1.18	8	0.85	8	1.10	8	0.58	7
Health center	0.27	11	0.28	10	0.19	10	0.01	11	0.35	11
Home-relax	30.9	2	27.8	3	26.5	3	40.3	1	34.2	2

Notes: Workplace Public/Private for Students and Home Workers is the time spent searching for jobs. I calculated the number of hours people spend searching for jobs in public versus private places using total time on the job search and the percentage of current workers in public and private workplaces as weights. 86% (89%) of female (male) workers work in private places and 14% (11%) in private workplaces. The highest rank (rank columns) corresponds to the place in which the least time is spent.

Source: Own calculations based on ENUT, 2002.

Table A2.4 also informs about the time use of people in different occupations. For example, females working in private workplaces spend more time working at home than at private workplaces. Females working in public workplaces spend about the same time at work than at home. Therefore, females who have private or public jobs spend at least the same proportion of time working at home than working at their respective private or government workplaces. One possible explanation for this phenomenon is part-time jobs. In contrast, men who work in public or private workplaces spend about 45% of their time working outside home and only 6% of their time working at home.

Students only spend about 25% of their time at school. In addition, women spend 25% of their time getting work done at home and about 40% relaxing at home (listening to music, using the computer etc.). Conversely, men who are students spend relatively more time relaxing (46%) than doing housework activities at home.

The time spent at transportation facilities is about 5% of total time. Private, public workers and students spend more time using transportations services than home workers, regardless of their gender. The time spent at retail stores is about 1% points higher for women than for men. For both men and women, the time spent at restaurants and hospitals is negligible implying that the effect of restrictions on smoking at these locations should be relatively minor. The student population in the age range 20-50 years spends literally no time at hospitals were the stringiest regulations are in effect.

REFERENCES

Belsley, David A. 1991. *Conditioning diagnostics: collinearity and weak data in regression*. New York: John Wiley & Sons, Inc.

Ben-Zaken Cohen, Sigal, Peter D. Paré, S. F. Paul Man, and Don D Sin. 2007. "The growing burden of chronic obstructive pulmonary disease and lung cancer in women: examining sex differences in cigarette smoke metabolism." *Am J Respir Crit Care Med.*, 176:2, pp. 113-20.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119:1, pp. 249-75.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics*, 90:3, pp. 414-27.

Cawley, John, Sara Markowitz, and John Tauras. 2006. "Body Weight, Cigarette Prices, Youth Access Laws and Adolescent Smoking Initiation." *Eastern Economic Journal*, 32:1, pp.149-70.

Chaloupka, Frank J. and Rosalie Liccardo Pacula. 1999. "Sex and race differences in young people's responsiveness to price and tobacco control policies." *Tobacco Control*, 8:4, pp. 373-77.

Chaloupka, Frank J. and Kenneth Warner. 1999. "The Economics of Smoking." *NBER Working Paper No.7047:* Cambridge, MA.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to Demand for Durable Goods." *Econometrica*, 39:5, pp. 829-44.

Chriqui, Jaime F., Marcy Frosh, Ross C. Brownson, Dana M. Shelton, Russell C. Sciandra, Robin Hobart, P. H. Fisher, Regina el Arculli, and Marianne H. Alciati. 2002. "Application of a rating system to state clean indoor air laws (USA)." *Tobacco Control*, 11:1, pp. 26-34.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Washington, DC: The Johns Hopkins University Press.

DeCicca, Philip, Donald S. Kenkel, and Alan Mathios. 2000. "Racial Difference in the Determinants of Smoking Onset." *Journal of risk and uncertainty*, 21:2, pp. 311-40.

DeCicca, Philip, Donald S. Kenkel, and Alan Mathios. 2002. "Putting Out the Fires: Will Higher Taxes Reduce the Onset of Youth Smoking?" *Journal of Political Economy*, 110:1, pp. 144-69.

DeCicca, Phillip, Donald S. Kenkel, and Allan Mathios. 2005. "The fires are not out yet: Higher taxes and young adult smoking," in *Advances in Health Economics and Health Services Research*. Michael Grossman and Björn Lindgren ed, pp. 301-20.

DeCicca, Philip, Donald S. Kenkel, Alan Mathios, Yoon-Jeong Shin, and Jae-Young Lim. 2008a. "Youth smoking, cigarette prices, and anti-smoking sentiment." *Health Economics*, 17:6, pp. 733-49.

DeCicca, Philip, Donald S. Kenkel, and Alan Mathios. 2008b. "Cigarette taxes and the transition from youth to adult smoking: Smoking initiation, cessation, and participation." *Journal of Health Economics*, 27:4, pp. 904-17.

Falba, Tracy, Mireia Jofre-Bonet, Susan Busch, Noelia Duchovny, and Jody Sindelar 2003. "Reduction of Quantity Smoked Predicts Future Cessation among Older Smokers." *Addiction,* 99:1, pp. 93-102

Gruber, Jonathan and Jonathan Zinman. 2001. "Youth smoking in the United States: evidence and implications," in *Risky behavior among youths: an economic analysis*. Jonathan Gruber ed. Chicago University of Chicago Press.

Ibáñez Hernández, Norma A. 2005. "Disposiciones jurídicas federales sobre la prohibición de fumar," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco, México*. Raydel Valdés-Salgado, Eduardo C. Lazcano-Ponce and Mauricio Hernández-Avila eds. Cuernavaca: Instituto Nacional de Salud Pública.

Instituto Nacional de Estadística, Geografía e Informática (INEGI). 2004. "Encuesta Nacional de Adicciones, ENA-2002." INEGI: Aguascalientes.

Jha, Prabhat, Philip Musgrove, Frank J. Chaloupka, and Ayda Yurekli. 2000. "The economic rationale for intervention in the tobacco market," in *Tobacco control in developing countries*. Prabhat Jha and Frank Chaloupka eds. New York: Oxford University Press.

Jiménez-Ruiz, J. A., Belén Sáenz de Miera, Luz M. Reynales-Shigematsu, Hugh. R. Waters, and Mauricio Hernández-Avila. 2008. "The impact of taxation on tobacco consumption in Mexico." *Tobacco Control*, 17:2, pp. 105–10.

Landes, William and Lewis Solomon. 1972. "Compulsory Schooling Legislation: An Economic Analysis of Law and Social Change in the Nineteenth Century." *The Journal of Economic History*, 32:1, pp. 54-91.

Lleras-Muney, Adriana. 2002. "Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939." *Journal of Law & Economics*, 45:2, pp. 401-35.

Mullahy, John. 1985. "Cigarette Smoking: Habits, Health Concerns and Heterogeneous Unobservables in a Micro-econometric Analysis of Consumer Demand." Ph.D. Dissertation. University of Virginia: Charlottesville.

Olaiz-Fernández, Gustavo, Juan Rivera-Dommarco, Teresa Shamah-Levy, Rosalba Rojas, Salvador Villalpando-Hernández, Mauricio Hernández-Avila, and Jaime Sepúlveda-Amor. 2006. "Encuesta Nacional de Salud y Nutrición 2006." Instituto Nacional de Salud Pública: Cuernavaca.

Pedrero Nieto, Mercedes. 2004. "Género, trabajo doméstico y extradoméstico en México. Una estimación del valor económico del trabajo doméstico." *Estudios Demográficos y Urbanos*. El Colegio de México: Mexico City.

Reynales-Shigematsu, Luz M, Rosibel Rodriguez-Bolaños, Jorge A Jiménez Ruíz, Sergio A Juárez-Márquez, Angelica Castro-Rios, and Mauricio Hernández-Avila. 2006. "Costos de la atención médica atribuibles al consumo de tabaco en el Instituto Mexicano del Seguro Social." *Salud Pública Mex*, 48:supl 1, pp. s48-s64.

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

Sáenz de Miera, Belén, Jorge A. Jiménez, and Luz M. Reynales. 2007. "La economía del tabaco en México." Instituto Nacional de Salud Pública: Cuernavaca.

Sesma-Vázquez, Sergio, Julio C. Campuzano-Rincón, Victor G. Carreón-Rodríguez, Felicia Knaul, Francisco J. López-Antuñano, and Mauricio Hernández-Avila. 2002. "El comportamiento de la demanda de tabaco en México: 1992-1998." *Salud Pública Mex*, 44:supl 1, pp. s82-s92.

Sesma-Vázquez, Sergio, Raymundo Pérez Rico, Esteban Puentes Rosas, and Raydel Valdés Salgado. 2005. "El precio como determinante del consumo de Tabaco en México, 1994 – 2002," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco*. Eduardo C. Lazcano-Ponce Raydel Valdés-Salgado, and Mauricio Hernández-Avila ed. Cuernavaca: Instituto Nacional de Salud Pública.

Tauras, John A. 2006. "Smoke-Free Air Laws, Cigarette Prices and Adult Cigarette Demand " *Economic Inquiry*, 44:2, pp. 333-42.

The Economist. February, 7th 2008. "How to save a billion lives."

U.S. Department of Health and Human Services. 1986. "The health consequences of involuntary smoking " *A Report of the Surgeon General* Rockville, MD.

Wasserman, Jeffrey, Willard G Manning, Joseph P Newhouse, and John D Winkler. 1991. "The effects of excise taxes and regulations on cigarette smoking." *Journal of Health Economics* 10:1, pp. 43-64.

Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *The American Economic Review*, 93:2, pp. 133-38.

CHAPTER 3

OPORTUNIDADES TO REDUCE SMOKING IN MEXICO?

1. Introduction

This chapter investigates the causal effect of *Oportunidades*, a federal program in Mexico, on the smoking behaviors of its participants. The aim of this program is not to affect smoking directly, but to reduce poverty through investments in health, nutrition and education. Yet, based on a long line of health economics research there are several reasons to predict that the *Oportunidades* intervention will affect smoking among poor Mexicans. First, it provided households with subsidies that were very large compared to their baseline incomes. Based on estimated income elasticities that show that smoking is a normal good in Mexico (Jiménez-Ruiz et al., 2008), these subsidies are expected to increase smoking participation by about 9 percent. Second, *Oportunidades* provided participant women and adolescents in high school with health information sessions, called *pláticas*. Beneficiaries have reported that these sessions helped them understand the importance of healthy lifestyle behaviors, hygiene and prevention (Adato et al., 2000). Extrapolating from studies of the effects of health information campaigns on smoking (Nuño-Gutiérrez et al., 2008), it is reasonable to predict that the program might reduce smoking participation by about 25 percent. Third, *Oportunidades* increased the schooling of adolescent participants, and a long line of health economics research suggests that increased schooling should improve health and reduce unhealthy behaviors like smoking. Based on Behrman et al., (2005) the schooling attainment of children after four years of participation in *Oportunidades* should have increased by 0.35 years. From previous research, this is expected to reduce the prevalence of smoking among Mexican adolescents by 14 percent.

The contributions of this chapter are the following. First, estimating the causal treatment effect of participation in *Oportunidades* on adolescent and adult smoking, documents the existence of unintended effects of the program. Furthermore, since smoking is an input in the production of health (Grossman, 1972) and health might reduce poverty via its effects on economic growth (Mayer, 2001; Fields, 2001), estimating the effect of the program on smoking elucidates the possibility of eliminating long-lasting poverty. Second, Gutierrez et al. (2005) and Duarte Gómez et al. (2005) have used propensity score matching (PSM) techniques based on retrospective data to estimate average treatment effects on smoking habits of program participants. In contrast, the approach in this essay consists on exploiting an exogenous jump in program participation at the poverty threshold for eligibility by means of a fuzzy Regression Discontinuity (RD) design (Imbens and Lemieux, 2007).[7]

The RD is a quasi-experimental design that allows estimating a local average treatment effect (LATE) on the treated at the poverty threshold. This effect predicts what would happen with the smoking behaviors of non-participants if they were made eligible and decided to participate. As such, the LATE is particularly relevant when program expansions to cover better-off households via a small change in the cut-off for eligibility (van der Klaauw, 2008). Comparing the LATE with the average program effects previously analyzed, I investigate whether heterogeneous impacts of the program exist.[8] Third, this essay is the first to estimate the causal effects of

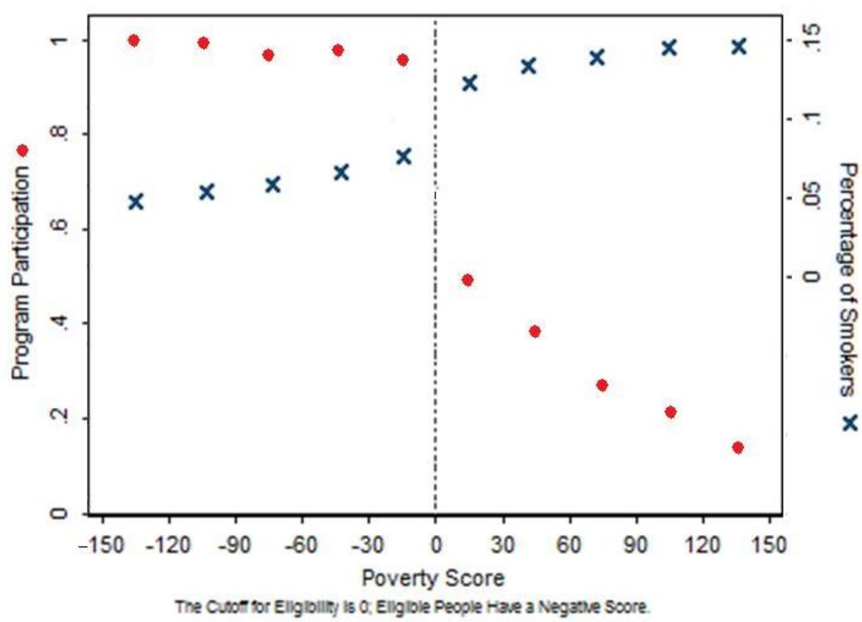
---

7 The distinction between a sharp and fuzzy RD design is due to Trochim (1984). In both designs the probability of participation changes discontinuously at the threshold. In the sharp design this probability changes from zero to one while in the fuzzy design the magnitude of the probability change is smaller.

8 Heterogeneous effects of *Oportunidades* on household consumption have been recently documented by Djebbari and Smith (2008).

Oportunidades by gender. Furthermore, because of the differential treatments between men and women in the *Oportunidades* program, and in the absence of peer effects, estimating the program's impact on adult male smoking enables isolating the income effect.

This paper’s identification strategy relies on the fact that only the households scoring below a poverty threshold (here normalized to zero) were eligible to participate in the program. As it is clear from Figure 3.1, the eligibility rules induced a considerable discontinuity in program participation rates (denoted with open circles) at the eligibility cutoff among nearly identical individuals. Hence, comparing smoking behaviors of those who barely made it to be eligible to those who failed to be eligible potentially eliminates any confounding program selection and omitted variable biases.



Notes: Average program participation and smoking participation rates are plotted as a function of five categories of the poverty score on each side of the eligibility cutoff. Participation rates are based on true data, but smoking rates are hypothetical.

Figure 3.1 Raw Averages of Program Participation and Smoking for Adults as a Function of the Poverty Scores that Determined Eligibility.
(The individuals in this sample live in comunities where *Oportunidades* started operating in 1998)

In consequence, it allows estimating the causal effect of *Oportunidades* on smoking at the threshold for eligibility. In principle, the negative effect of health information sessions and schooling might be offset by the positive income effect. Suppose it does not. Then, the program would have a positive effect on smoking and the discontinuity in program participation rates at the cutoff score would be echoed by a discontinuity in average smoking rates of the type shown in Figure 3.1 with x's. Smoking rates of individuals below the threshold would be much lower than smoking rates of individuals above the threshold. The discussion that follows demonstrates that the hypothetical average smoking rates are not representative of the true ones. This is shown in two ways. I first present graphical plots similar to Figure 3.1. Then I report the results of flexible parametric models that approximate the relationship between both program participation and eligibility and smoking and eligibility with an intercept shift at the threshold for eligibility. The rest of the chapter is organized as follows. Section II provides background on the *Oportunidades* program and discusses the pathways through which participating in this program is expected to change smoking behaviors. Section III scrutinizes the research design and the econometric methods used with the available data. Section IV presents the empirical findings for adults and adolescents. Section V addresses the validity of the design. Section VI isolates the likely effect of income on smoking among adults, and Section VII brings together the principal results.

2. Background on Oportunidades

2.1 Eligibility rules and identification strategy

Oportunidades, formerly known as PROGRESA, is the backbone of the social policy in Mexico. It currently covers approximately 25 million people, which represents 25 percent of the Mexicans in that country and 90 percent of the population living in extreme poverty (GEUM, 2008). *Oportunidades'* benefits include sizeable cash transfers conditional on households conforming to a set of coresponsabilities. The purpose of the later is to improve the human capital of the poor through investments in health, schooling and nutrition.[9]

At its original stage, this program was targeted to households in extreme poverty of rural marginalized localities with access to a school and a health clinic.[10] Due to budgetary restrictions, the program was randomly phased into communities over two years. In some communities it started in 1998 (treatment-98), and in the remaining communities, coverage started one year later (treatment-99). Within all the marginalized communities, the first step in the selection of beneficiary households involved a multi-dimensional approximation to the poverty condition (Skoufias et al., 1999). Using a discriminant analysis, several individual and households characteristics coming from a census survey conducted in 1997 (ENCASEH-97) were combined to generate a poverty score for each household.[11] Eligible households were those with

---

[9] The program provides nutritional supplements to infants, under-nourished children and pregnant and breastfeeding women (Skoufias and McClafferty, 2001). The impact of this benefit in the outcome of interest is only important if the food supplement releases resources that households can use to buy more cigarettes. This is the same effect that I expect to find as a result of the cash transfers, which I fully address in the paper.

[10] Localities were marginalized if their marginality index was very high or high. Other options of this index were very low, low and medium The marginality index was developed using the method of principal components, based on seven variables: 1) Share of illiterate adults (> 14 years) in the locality, 2) Share of dwellings without water, 3) Share of dwellings without drainage systems, 4) Share of dwellings without electricity, 5) Average number of occupants for room 6) Share of dwellings with dirt floor, and 7) Share of population working in the primary sector (Skoufias and McClafferty, 2001).

[11] The value of the poverty index depends on the following variables: a crowding index (number of people in the household/number of rooms), whether the head of household is female, whether the household has access to medical service, the total number of children in the household less than eleven years old, years of education of the household head, age of the household head, whether the household has a bath and whether the bath has running water, whether the floor of the house has a dirt floor, whether the household has a gas heating system, a refrigerator, a washing machine, whether assets include a vehicle, whether the home is in a rural area and region of residence dummies for the 19 census regions (Parker et al., 2006).

poverty scores below a region-specific cutoff level. The rights and responsibilities of the eligible households were informed in a community assembly. In particular, *Oportunidades* required household withdrawal from pre-existing social programs. The assembly was also aimed to gather feedback from the community about families that needed to be eligible but were not and vice versa. Disapproval of eligibility status by other members of the community was an exception more than a rule (González de la Rocha and Escobar, 2001).

In July of 1999, some of the households in rural areas that were initially ineligible became eligible through a densification process. Such a revision was undertaken to increase the number of households with certain characteristics that were felt to be under-represented when the eligibility status was first determined. In principle, making comparisons of individuals below and above the two thresholds would enable testing for nonlinear impacts of the program at different poverty levels. Unfortunately, performing this test was hindered by the impossibility to get (or infer from the data) the region-specific thresholds that resulted from the densification process.[12] The quasi-experimental evidence presented in this essay is based on the threshold for eligibility in 1997.

The densification process shrank the discontinuity in program participation at the original threshold in treatment-98 communities, and eliminated the discontinuity in treatment-99 communities. This can readily be seen in Appendix Figure A3.1 Panel A shows pre-densification average participation rates in treatment-98 communities as a function of five equal-size categories of the poverty score on each side of the eligibility cutoff. Panel B plots post-densification average participation rates in treatment-98 communities. A comparison of these figures suggests that the

[12] In fact, little is known about the criteria actually followed to boost the eligible population. For a discussion see Buddelmeyer and Skoufias, 2003 and Rubalcava and Teruel, 2003.

densification process leaved unchanged the participation rates of the originally eligible. It is interesting to note that some of the individuals in non-eligible households were already participating before the densification took place. This was due to both the community assembly and ad-hoc changes in eligibility status in communities that were affected by natural disasters in the late nineties (Rubalcava and Teruel, 2003). Indeed, the densification process in 1999 only slightly increased the participation rates of the originally ineligible. Panel C in Figure A3.1 of the Appendix post-densification average participation rates in treatment-99 communities. In these communities there is not sharp discontinuity in participation rates for individuals in the neighborhood of the original threshold. This raises concerns regarding the validity of the RD identification in these communities. In consequence, the data used in this essay is limited to treatment-98 rural communities.

2.2 Analysis of Potential Impacts of Oportunidades on Smoking

Oportunidades provides families with two subsidies: a food subsidy and a schooling subsidy. They are both handed directly to the mother of the household. The monthly amount of the food subsidy is of 150 pesos (about $15 dollars in 2002), which represents about 20 percent of the baseline household income. Households are encouraged to spend this money to improve nutrition, but they use it as they like. The food subsidy is disbursed conditional on (a) regular health clinic attendance by all family members, and (b) monthly attendance to health information sessions by the mother and adolescents in grades 10-12. Women have attended these sessions since the program started, and adolescents only during the school year since 2001 (De la Torre, 2005). The sessions cover 35 topics. Many aim to promote self-care and one of them covers addiction prevention (SDS, 2003).

Families with children are eligible to receive the schooling subsidy subject to children school attendance in one of the subsidy-eligible grade levels for at least 85 percent of days in a given month. Table 3.1 shows the schooling subsidy amounts and how they increase with grade level to offset the higher opportunity costs of working for older children. Scholarships are higher for girls as school attendance has traditionally been lower for them. Schooling benefits are permanently discontinued if a child fails a grade more than once.

Table 3.1 Schooling Subsidy Amounts by Grade and Gender Handed to the Mother of Households in *Oportunidades* with Children Attending School (pesos per month in the second semester of 2002*).

Educational Level and Grade	Gender of Child who Attends School and Participates in *Oportunidades:* Boy	Girl
Elementary School:		
Grade 3	100	100
Grade 4	115	115
Grade 5	150	150
Grade 6	200	200
Middle School:		
Grade 7	290	310
Grade 8	310	340
Grade 9	325	375
High School:		
Grade 10	490	565
Grade 11	525	600
Grade 12	555	635

Notes: ***In 2002, 10 pesos were approximately equivalent to 1 US dollar.***

Schooling subsidies are given to poor eligible families with children in subsidy-eligible grade levels. To receive these subsidy children must attend school at least 85 percent of the days in a given month. In addition to schooling subsidies, participant households receive a food subsidy of about $15 dollars (in 2002). The food subsidy is given conditional on: (a) regular health clinic attendance by family members; and (b) monthly attendance to health information sessions by the mother and adolescents in high school.

Source: Oportunidades Program Webpage: http://www.Oportunidades.gob.mx/informacion_general/main_ma.html

The mechanisms through which the incentives of the program are expected to affect smoking can be derived from an economic model of addiction. As discussed in the introduction, sophisticated economic models of rational addiction (Becker and Murphy, 1988) and time inconsistent preferences (Gruber and Köszegi, 2001) have been recently developed and tested. Nonetheless, a simpler model suffices to predict changes in smoking resulting from program participation. The myopic model of addiction by Mullahy (1985) suggests that, in a given time period t, the demand for cigarettes C, is a function of the stock of habits $\mathcal{H}$, the price of cigarettes P_c, the price of all other goods P_{OAG}, the individual's income I, schooling S and information ι:

$$C(t) = [\mathcal{H}(t), P_c(t), P_{OAG}(t), I(t), \mathrm{S}(t), \iota(t)] \qquad (1)$$

Schooling and information could also be collapsed in a vector of covariates affecting both the production of nicotine services and the perceptions of health promotion. In any event, the effects of *Oportunidades* are expected to affect smoking participation through an income effect, an information effect and a schooling effect. The rest of this subsection discusses the sign and magnitude of the expected changes based on available health economics empirical evidence.

The schooling and food subsidies of *Oportunidades* are expected to affect smoking participation through an income effect. An average of three years of participation in *Oportunidades* increased the average per capita income of those living in extreme poverty by nearly 28 percent (Rubalcava and Teruel, 2003). The income of the participants in my sample should have increased by 37 percent after four years of participation in the program. It has been shown that the increase in income due to transfers from *Oportunidades* resulted in higher food expenditure (Hoddinott, Skoufias and Washburn, 2000). Part of the money was spent on processed foods, such as sodas and cookies. Doctors and health promoters feared that the money could also be used to buy and consume cigarettes (Adato et al., 2000). Nonetheless, an increase in income does not necessarily imply that smoking should increase because relaxing the budget constraint enables more access to health information or to buying smoking quitting devices. The sign and magnitude of the effect on smoking depends on whether smoking is normal or inferior. This remains an empirical question. In developed nations, early demand studies concluded that it was normal, but recent ones find it to be inferior (Chaloupka and Warner, 1999). In Mexico, Jiménez-Ruiz et al. (2008) estimate an "income elasticity" of household smoking participation of 0.25. To the extent that this figure is representative of the income elasticity in rural areas, *Oportunidades* subsidies would result in 9 percent increase in smoking participation, which represents about 1.5 percentage points.

Another channel through which *Oportunidades* might induce participants to choose healthier behaviors is health information. Assuming ex-ante incomplete information, economic theory predicts that providing information regarding the consequences of tobacco consumption and the addictive nature of tobacco would solve an information failure. Consequently, people would accurately estimate the costs of smoking and ultimately smoke less (i.e. the sign of $\iota(t)$ in equation (1) above would be negative). Public health experts in Mexico support this view. They agree that "providing information to the population regarding the health damages caused by tobacco smoking [is] an effective tool to reduce this behavior" (López Antuñano, 2005).

Recent evidence of the effects of a school anti-smoking campaign in Mexico involving parents, teachers and peers, points to a reduction in the rate of experimental smokers of about 50 percent and no effect on smoking among regular smokers (Nuño-Gutiérrez et al., 2008). Data from the National Addiction Survey (2002) indicates that about 50 percent of smokers are light smokers (INEGI, 2004). This evidence suggests that the provision of information would modify the health perceptions of *Oportunidades* participants, increase the costs of smoking, and reduce smoking participation by about 25 percent.

The impact of the health information sessions is not expected to be confounded with other mass-media policies. This is because tobacco advertising in radio and television was banned in 2004, the same year in which health-warning labels increased from 25 to 50 percent of the back face of the cigarette packs (Sebrie, 2006), but smoking in this paper is measured in 2003. The small health-warning labels that were introduced nationally in 2000 are not likely to have been effective (Hammond et al. 2007); leaving enough room from the health information sessions to have an impact.

Oportunidades might have also affected cigarette consumption among adolescents through a schooling effect. "Years of formal schooling completed have been identified as the most important correlate of good health [and less unhealthy behaviors]" (Grossman, 2004; 32). Farrell and Fuchs (1982) claim that the observed correlation between health and education is mainly due to unobservable characteristics affecting both, investments in health and schooling (e.g. time preferences). An alternative explanation for the correlation is a causal effect from schooling to health or vice versa. Recent empirical evidence suggests a causal effect going from schooling to less unhealthy behaviors. Using the approach of instrumental variables, Currie and Moretti (2003), Kenkel et al. (2006), and de Walque (2007) find a negative effect of schooling on current smoking in the US. These studies, however, provide less guidance about the specific causal pathways involved.[13]

De Walque (2007) claims that an additional year of schooling beyond college decreases the probability of smoking by 6 percent. Behrman et al. (2005) simulate that

[13] Schooling affects health in three main ways. First, because education is an investment that raises the future level of income and consumption, educated individuals have more incentives to invest in their health (Becker 1993). Second, schooling improves the 'allocative efficiency' of individuals. That is, through more schooling people get access to health knowledge which makes them choose healthier behaviors. Third, schooling increases people's 'productive efficiency'. This is because they produce more health from the same amount of inputs (Grossman, 1972). Kenkel (1991) indicates that the productive efficiency pathway of schooling might induce healthier behaviors.

participating in *Oportunidades* over an 8-year time period would increase the average educational attainment of children by 0.7 years. Because children in the current sample have participated in the program an average of 4 years, their educational attainment is about 0.35 years more than that of non participant children. Extrapolating from this evidence, it is possible to conclude that *Oportunidades* would decrease smoking participation by 2 percentage points, which represents a 14 percent reduction in smoking participation rates in rural Mexico. Although informative, this ad-hoc estimation should be taken with serious caution for the following reasons. First, it assumes that the effect of schooling on smoking is linear, but the combined evidence in Kenkel et al. (2006) and de Walque (2007) suggests that the marginal returns to schooling are positive but decreasing.[14] If so, because *Oportunidades* main impact is on middle school attendance, its effect on smoking would be higher. Second, it generalizes the instrumental variables estimations to the entire population in the US, and then to the Mexican population.

Table 3.2 summarizes the expected effects of the program. To the extent that there are no spillover effects of information, the overall effect of the program on men smoking should be at around 9 percent. Since the negative information and schooling effects can potentially be offset by the income effect, the theoretical effect among women and adolescents would be ambiguous. Based on the back-of-the envelope calculations, and in the absence of peer effects, participation in *Oportunidades* would expect to decrease the smoking prevalence among women by as much as 16 percent and among adolescents by as much as 30 percent.

[14] While Kenkel et al., 2006 find that high school completion decreases the probability of smoking by 25 percent; de Walque (2007) finds that finishing college decreases the probability of smoking by 16 percent.

Table 3.2 Mechanisms Behind, and Sign of, the Potential Effect of Participation in *Oportunidades* on Smoking based on Differential Treatments among Adult Men, Adult Women and Adolescents.

Benefits of the Program	Mechanism behind a Potential Impact of the Program on Smoking	Expected sign of the Effect of Each Mechanism and of the Overall Effect for:		
		Adult Men	**Adult Women**	**Adolescents**
Food and schooling subsidies	*Income effect*	+	+	+
Health sessions providing information on addiction prevention	*Allocative efficiency:* Information makes people choose healthier behaviors.		-	-
Schooling	*Allocative efficiency:* Information at school makes people choose healthier behaviors more efficiency *Productive efficiency:* schooling enables people choose healthier behaviors with the same amount of inputs.			-
Overall expected effect (in the absence of peer effects):		+	?	?

Before I discuss the data and methods of this paper, one issue regarding program effects on smoking participation in relation to the dynamic aspects of smoking deserves further consideration. As DeCicca et al. (2008) highlight, because of addiction, current participation reflects past decisions regarding initiation and cessation. Accumulated evidence from both developed and developing countries suggests that these decisions occur in different points of the life cycle. According to the 2002 Mexican Family Life Survey (MxFLS) data set, on average people start smoking at 18.4 years of age and quit smoking at 35.9 years. This evidence suggests that most of the program impacts on smoking participation would come about changes in smoking initiation among adolescents, and smoking cessation, among adults.

3. Data and Econometric Methods

3.1 Data

The data used in this chapter come from the ENCASEH 1997 and ENCEL 2003 rural surveys. The ENCASEH provides baseline characteristics and household poverty scores. It was composed of two groups of communities: treatment-98 and treatment-99. I restrict attention to the individuals in treatment-98 communities because the densification process invalidated the RD design in treatment-99 localities (Appendix Figure A3.1, panel C). Treatment-98 are communities where the program started operating since 1998. The ENCEL 2003 is a follow-up survey containing self-reported information about smoking and other health behaviors. The empirical analysis is conducted on separate subsamples of adults and adolescents.

The LATE computed in this chapter is informative about smoking behaviors of individuals participating in the program for an average of four years relative to those of non-participants. This is because program participants started receiving the benefits of the program between March of 1998 and December of 2000, and their smoking behaviors are measured in 2003. Smoking participants are those who answered ‘yes’ to the question: Do you currently smoke? Table 3.3 presents summary statistics of the estimation sample.

3.2 Econometric Framework

Let C_i be cigarette smoking in 2003, and T_i be an indicator for participation in *Oportunidades*. By definition, an individual is never simultaneously observed in both

Table 3.3 Descriptive Statistics by Participation Status in *Oportunidades* for the Sample of Adults and Adolescents.

	Adults In Oportunidades?:		Adolescents In Oportunidades?:	
	No	**Yes**	**No**	**Yes**
Panel A:Dependent Variables Measured in 2003				
Smoking rate	0.050	0.055	0.091	0.071
	(0.007)	(0.005)	(0.011)	(0.007)
Participation in *Oportunidades*	0.000	1.000	0.000	1.000
	(0)	(0)	(0)	(0)
Panel B. Demographic Variables Measured in 2003				
Female	0.728	0.733	0.552	0.527
	(0.013)	(0.009)	(0.020)	(0.013)
Age	42.043	40.084	17.239	16.929
	(0.429)	(0.239)	(0.071)	(0.044)
Married	0.849	0.916	n.a.	n.a.
	(0.011)	(0.006)	(-)	(-)
Single	n.a.	n.a.	0.845	0.892
	(-)	(-)	(0.014)	(0.008)
Indigenous	0.344	0.481	0.276	0.383
	(0.014)	(0.011)	(0.018)	(0.013)
Panel C. Covariates Associated with Smoking Measured in 2003				
No. years since started smoking	7.240	5.807	0.903	0.630
	(0.447)	(0.261)	(0.070)	(0.039)
Tobacco control laws	0.259	0.155	0.266	0.139
	(0.013)	(0.008)	(0.018)	(0.009)
Cigarette price	12.281	12.267	12.257	12.213
	(0.025)	(0.018)	(0.034)	(0.024)
Panel D. Other Covariates Measured in 1997				
Despensa[&]	0.135	0.138	n.a.	n.a.
	(0.011)	(0.007)	(-)	(-)
Niños de Solidaridad[&]	n.a.	n.a.	0.166	0.215
	(-)	(-)	(0.016)	(0.011)
Land Property	0.688	0.610	0.713	0.662
	(0.015)	(0.011)	(0.020)	(0.013)
Cigarette price	11.734	11.689	11.725	11.634
	(0.040)	(0.029)	(0.058)	(0.038)
Observations	1108	2262	627	1443

*Notes: *One US dollar was about 10.5 pesos in 2003 and 8 pesos in 1997. Standard errors are reported in parentheses.*

[&]*Niños de Solidaridad gave grants for children in marginalized communities to finish elementary school. Despensas provided a monthly package of basic food products to very poor families*

Source: Own calculations based on ENCASEH 1997 and ENCEL 2003.

the treatment and the absence of treatment states. Therefore, $T_i = 1$ if an individual i belongs to a household that started receiving the benefits of the program in 2000 or before and continues receiving the benefits in 2003, and $T_i = 0$ if an individual belongs to a household that never received the benefits of the program. Let $C_i(1)$ be the outcome given treatment, and $C_i(0)$ the outcome in the absence of treatment. Then the actual outcome we observe is: $C_i = T_i C_i(1) + (1 - T_i) C_i(0)$. A common regression model representation expresses the outcome as a function of program participation and an unobserved error term representing all causes of cigarette consumption other than participation ε_i :

$$C_i = \alpha + \theta T_i + \varepsilon_i \tag{2}$$

For simplicity, the previous expression excludes a vector of observable characteristics determining smoking other than treatment, but its inclusion is straightforward. The foremost parameter of interest in this paper, the causal effect of program participation on smoking, is given by θ. To the extent that program participation is exogenous, least squares estimation of equation (2) would produce an unbiased estimate of θ.

As discussed in section 2.1, *Oportunidades* gives households incentives in the form of cash transfers and coresponsabilities, and assigns them to a certain eligibility status based on their poverty scores. Program participation is voluntary. Based on expected benefits and costs of the program, households make the decisions to join the program or not. As such, some households might find it more beneficial to join the program than others; they may 'self-select into the program' (Heckman, 2008). In the *Oportunidades* context, the self-selection problem might be aggravated by the fact that the agents making the choice to join the program may be different from the agents receiving treatment. Note, for instance, that participant adolescents are younger and are more likely to be indigenous than non-participant youths (Table 3.3). If these differences persist in characteristics not observed by the econometrician, program participation endogeneity concerns would be at stage. Fortunately, the RD design used in this essay overcomes this problem and enables the estimation of an unbiased causal effect of θ: a local average treatment effect.

The identifying assumption in a RD design relies on the fact that program participation is a function of eligibility E (i.e. $T = fn(E)$). We also know that the first stage of the selection of eligible households was based on a known poverty score. In particular, households scoring below a predetermined cutoff poverty score (here normalized to zero) were eligible to participate in the program. That is, letting E_i be program eligibility, and P_i be the poverty index in 1997: $E_i = 1 \ if \ P_i < 0$ and $E_i = 0$. The list of eligible beneficiaries was finalized after getting feedback from the community, and was later changed when the densification process took place. The variables that leaded to these changes are unobserved by the econometrician implying that eligibility depends on the poverty score in a stochastic manner, but in such a way that the propensity of treatment is known to have a discontinuity at the threshold for eligibility. Given this feature, *Oportunidades* can be best characterized by means of a fuzzy Regression Discontinuity design. In contrast, a sharp RD design would require treatment to be a deterministic function of the poverty score (Trochim, 1984).

Estimating θ in a sharp RD design, requires continuity at the cutoff $E[\varepsilon_i | P_i = p]$. Local continuity requires that individuals just above and below the cut-off have similar average potential outcomes when receiving treatment and when not. This identifying condition implies that individuals in the neighborhood of the cutoff should share the same predetermined characteristics (i.e. there is local randomization). Judging from the eligibility rules, this assumption seems plausible, but cannot be taken for granted. As in other research designs, it is impossible to test this assumption for unobserved characteristics. However, in section 5 of this essay, four different tests provide evidence suggesting that observable characteristics of individuals close to the cutoff were nearly identical. Hence, following Hahn et al. (2001), the average treatment effect of *Oportunidades* if the sharp discontinuity in the probability of treatment was one would be given by the difference in smoking behaviors for individuals just below and just above the eligibility cutoff:

$$\theta^{sharp} = \lim_{p\uparrow 0} E[C_i|P_i = p] - \lim_{p\downarrow 0} E[C_i|P_i = p] \qquad (3)$$

Since the probability of participation in *Oportunidades* at the threshold for eligibility does not jump from zero to one (see Figure 3.1), θ^{sharp} in equation (3) would not generally lead to correct inferences regarding an average treatment effect (van der Klaauw, 2008b). Nonetheless, assuming local conditional independence (i.e. individuals do not select into treatment on the basis of anticipated gains from treatment), a fuzzy RD design identifies an average treatment effect of participation precisely when the change in the probability of participation is less than one. This effect is estimated as follows:

$$\theta = \frac{\lim_{p\uparrow 0} E[C_i|P_i=p] - \lim_{p\downarrow 0} E[C_i|P_i=p]}{\lim_{p\uparrow 0} E[T_i|P_i=p] - \lim_{p\downarrow 0} E[T_i|P_i=p]} \qquad (4)$$

In the presence of self-selection on the basis of expected gains from the treatment, Hahn et al. (2001) show that under a weaker local monotonicity assumption, the ratio (4) will instead identify a local average treatment effect (LATE) at the cutoff point. The local monotonicity assumption requires the individual level of treatment to be a monotonically increasing (or decreasing) function of the poverty score. This assumption is fundamentally untestable. However, it is expected that the probability of treatment would decrease as the poverty index increases (people become richer). If so, the denominator of equation (4) captures the (less than one) discontinuity in program participation at the threshold. As such, it magnifies the estimate in the discontinuity of smoking status given by (3) by the inverse of the fraction of compliers (Matsudaira, 2008). Compliers are individuals who were induced to participate in the program because their poverty score happened to be slightly below the cutoff score .The LATE differs from the average treatment effect estimated by Duarte Gómez et *al.* (2005) and Gutiérrez et *al.* (2005). The later effect is an estimation of the average effect of the program on compliers and people who participate in the program no matter where the threshold for eligibility is located.

3.3 Estimation

For the estimation of the LATE (as in equation is (4) above), suppose that cigarette smoking C_i measured in 2003, and program participation T_i (as defined in the previous section), can be expressed as a function of the 1997 poverty score P_i as follows:

$$C_i = \alpha_1 + D_i\pi_1 + m_1(P_i) + \varepsilon_{1i} \quad \text{where } E\left[v_1 \middle| P_, E\right] = 0 \text{ and } E_i = 1 \; if \; P_i < 0 \tag{5}$$

$$T_i = \alpha_0 + D_i\pi_0 + m_0(P_i) + v_{0i} \quad \text{where } E\left[v_0 \middle| P_, E\right] = 0 \text{ and } E_i = 1 \; if \; P_i < 0 \tag{6}$$

Under the assumptions that (a) $m_1(.)$ and $m_0(.)$ are continuous at $P = 0$, and (b) the parametrization of the $m(.)$ function is accurate; π_1 in (5) represents the size of the discontinuity in smoking and π_0 in (6) represents the size of the discontinuity in program participation. Note that π_1 and π_0 are the numerator and denominator of equation (4), respectively. Therefore, the causal effect of program participation on smoking is given by their ratio: $\theta = \pi_1/\pi_0$.

There are several ways to estimate θ. One approach is to use non-parametric and semiparametric estimations such as one-sided kernel, local polynomial regression and estimators based on partially linear model estimation with and without covariates (see, for example Haht et al., 2001; Porter, 2003; and Frölich, 2007). The approach in this paper is to use a flexible parametric model following DiNardo and Lee (2004) and Matsudaira (2008). Indeed, it is possible to estimate π_1, π_0, and hence θ, by means of a flexible parametric model. $m_1(.)$ and $m_0(.)$ in equations (5) and (6), represent a *n* degree polynomial in P, fully interacted with the indicator for eligibility $E_i = 1 \; if \; P_i < 0$ allowing the shape of the conditional expectation vary on either side of the cutoff. This parametrization is expressed as:

$$C_i = \alpha_1 + E_i\pi_1 + E_i \sum_{d=1}^{n} \varphi_{1d} \, (P_i)^d + (1 - E_i) \sum_{d=1}^{n} \varphi_{1p}{}' \, (P_i)^d + \varepsilon_{1i} \tag{7}$$

$$T_i = \alpha_0 + E_i\pi_0 + E_i \sum_{d=1}^{n} \varphi_{0d} \, (P_i)^d + (1 - E_i) \sum_{d=1}^{n} \varphi_{0p}{}' \, (P_i)^d + \varepsilon_{0i} \tag{8}$$

The estimation of π_1 and π_0 depends on the particular functional form of the model relating C_i and T_i to the eligibility score. For that reason, I will perform various specification checks varying the order of the polynomial on either or both sides of the cutoff. However, the preferred estimations will come from a specification where the order of the polynomial is chosen using the Schwarz (1978) criterion. This criterion, also known as the Bayesian Information Criterion (BIC), is commonly used to compare competing regression models. It penalizes a larger model for using additional degrees of freedom while rewarding improvements in goodness of fit (Baum, 2006).

Because this model is exactly identified I estimate it via two-stage least squares (TSLS). As Wooldridge recognizes (2002, 636) the LATE is identical to the instrumental variables estimator of θ in equation (2) when eligibility E_i is used as an instrument for participation T_i. Eligibility is a valid instrument for participation to the extent that it does not have an independent causal impact on smoking besides its effect through participation in *Oportunidades*. One can hardly argue against this assumption. Hence, the first stage of the TSLS estimates program participation as a function of eligibility and the interaction terms. This is analogous to estimating equation (8) above, implying that π_0 is the coefficient on the dummy for eligibility in the first stage. The instrument (eligibility) must be powerful in predicting a statistically significant discontinuity in participation at the threshold for eligibility. If not, $\hat{\pi}_0$ would be zero, and so we would have no variation to work with. Knowing that eligibility induces a discontinuity in program participation, the second stage of the TSLS uses predicted program participation along with the interaction terms to predict smoking. As such, θ is the coefficient on the dummy for participation in the second stage. The relationship between $\pi_0,\ \pi_1,$ and θ implies that $\pi_1 = \theta * \pi_0$. The standard errors of π_1 are estimated from a reduced form specification of Y_i on the dummy for eligibility and the polynomial interaction terms (as in equation (7)). Robust standard errors are reported throughout. In that way, I account for the fact that members of the same family share the same poverty score, and allow for heteroscedasticity due to misspecification of the $m(.)$ function.

All the regressions in this chapter exclude a vector of covariates. This is because θ is unchanged to the inclusion of smoking covariates when the identifying assumptions are met. In section 5 I prove that this is indeed the case. In addition, I also contrast the results of the regressions that exclude covariates with those including standard socio-demographic variables, state-level cigarette prices derived from barcode scanning in large food stores reported monthly by the Central Bank of Mexico and a dummy for state-level tobacco control laws.

4. Results

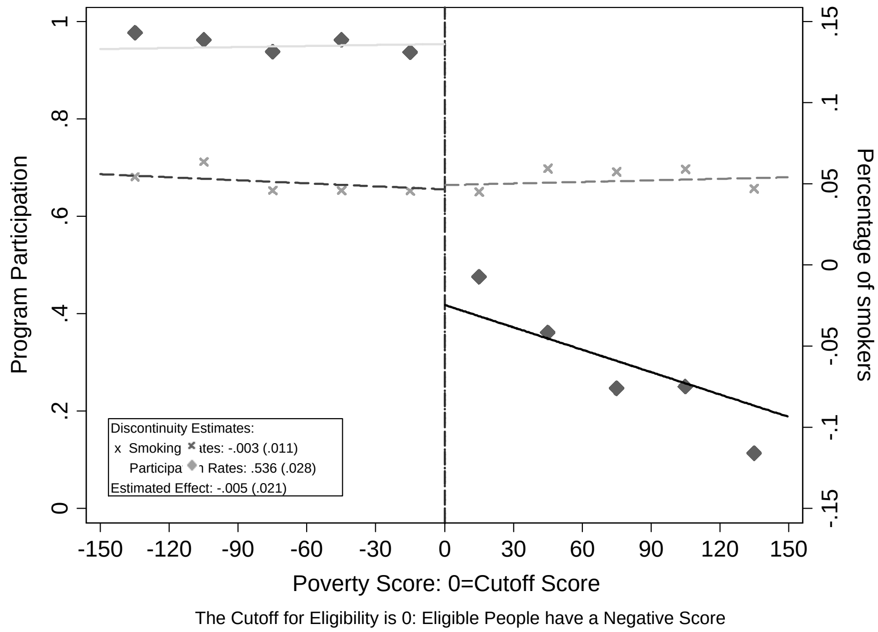
In this section I present estimates of the effect of program eligibility on participation and smoking, and of the impact of participation in *Oportunidades* on the smoking behaviors of adults and adolescents.

4.1 Effect of eligibility on participation in Oportunidades

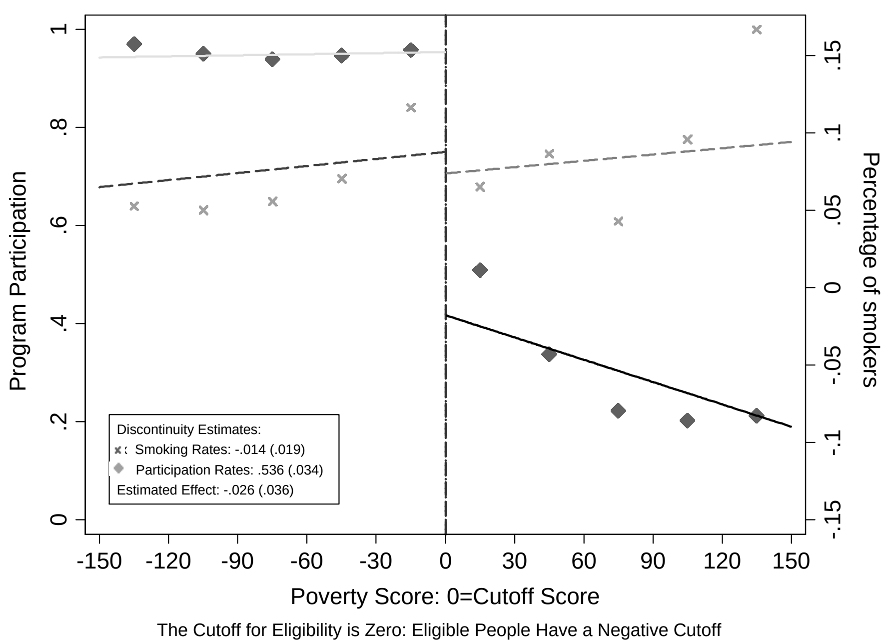
Represented with small diamonds, Figure 3.2 shows average program participation rates of adults (Panel A) and adolescents (Panel B) for five equally spaced poverty score categories on each side of the cutoff. A sizeable jump in program participation of adults at the eligibility cutoff of at least 40 percent is apparent in both figures. Table 3.4 presents estimations of the size of this jump. These are the coefficients on the dummy for eligibility coming from flexible parametric models where participation is modeled as a function of a full set of polynomial terms along with the eligibility indicator (see eq. (8)). The Schwarz preferred specification includes first-order polynomial terms on both sides of the cutoff. The predicted values of this model are superimposed to the average participation rates in Figure 3.2. In spite that the poverty score in the figure is truncated at +/- 150 points, the regressions are estimated on the entire range of data. This is because, under the assumption that the first-order polynomial is the true function of the underlying data, the LATE is "efficiently estimated using data that are both close to and far from [either side of] the discontinuity threshold" (DiNardo and Lee, 2004; 1400).

The Schwarz preferred estimation indicates that being eligible to participate in the program is associated with a 53 percentage point increase in the probability of participation for both adults and adolescents with standard errors of .028 and .034,

Panel A. Discontinuity Estimates for Adults



Panel B. Discontinuity Estimates for Adolescents



Notes: Average program participation and smoking participation rates are plotted as a function of five categories of the poverty score on each side of the eligibility cutoff. The lines are conditional expectations of specifications as in equations (7) and (8) in the text. They are estimated using the whole range of data with poverty scores of +/- 500. The first order polynomial approximation presented here is the Schwarz (1978) preferred specification.

Figure 3.2 Impact of Eligibility on Program Participation and Smoking and Effect of Program Participation on Smoking.

Table 3.4 Identification of Program Effects: Discontinuity Estimates of Participation in *Oportunidades* at the Poverty Score Eligibility Cutoff for Adults and Adolescents.

Polynomial order on both sides of the eligibility cutoff	Discontinuity Estimates of Participation at Cutoff for: Adults	Adolescents
1	0.536[&]	0.536[&]
	(0.028)	(0.034)
2 right, 1 left*	0.458[‡]	0.454[‡]
	(0.037)	(0.044)
2	0.449	0.457
	(0.037)	(0.045)
3	0.408	0.407
	(0.047)	(0.058)
4	0.403	0.386
	(0.056)	(0.069)
5	0.414	0.380
	(0.066)	(0.083)
6	0.399	0.360
	(0.076)	(0.097)
7	0.376	0.351
	(0.085)	(0.113)
Observations	3370	2070

*Notes: *This regression is estimated using a quadratic parametrization to the right of the cutoff for eligibility and a linear to the left.*

& are the preferred estimates based on Schwarz (1978).

‡ are the second best estimates based on Schwarz (1978).

Each cell comes from the first stage regression of the two-stage least squares (TSLS) estimation. In particular, each entry represents π_0 in equation 8 in the text. This is the dummy for eligibility coefficient of a flexible parametric regression of program participation on eligibility that also includes interactions of the dummy for eligibility and polynomial terms of the poverty score that defined eligibility. The order of the polynomial terms is specified in the first column of the table. Except for the estimates in the second row of the table, I use the same polynomial order to the right and to the left of the cutoff.

None of the regressions include covariates. Clustered standard errors are reported in parentheses.

respectively (first row, Table 3.4). The second-best specification, which includes linear polynomial terms to the left of the cutoff and quadratic to the right, suggests a 45 percentage point difference in participation rates around the threshold. Even thought the size of the discontinuity decreases as higher-order polynomials are included, the magnitude remains sizeable and significant. This evidence suggests that scoring below a predetermined cutoff had a sturdy impact on the probability of participation in *Oportunidades*. On the econometric side, the persistent and statistically significant discontinuity demonstrates the existence of a strong first stage relationship between eligibility (the instrument) and program participation. Hence, identifying causal effects using the RD design is not only justified, but powerful.

4.2 Effect of participation on smoking behaviors

Represented with x's, average adult smoking rates are plotted in Figure 3.2, Panel A. The similarity of the smoking rates on either side of the cutoff is visually apparent. Moreover, the predictions of the parametric estimates (also plotted in Figure 3.2 with dashed lines) point to a null effect of eligibility on smoking behaviors. In fact, the -0.003 coefficient together with the small confidence interval around the point estimate implied by the standard errors (s.e.: .011) provide strong evidence of a null effect of eligibility on smoking behaviors (Table 3.5). The point estimate of the effect of adult participation in *Oportunidades* on smoking is zero (-.005) and precise. In fact, the confidence interval suggests an effect going from -4 to 3 percentage points. The implications of this non-result are interesting, and will be discussed further in section 5 below.

Panel B of Figure 3.2 plots average program participation and smoking rates and their parametric predictions for adolescents. The parametric estimations of the effects are presented in Table 3.5. The effect of eligibility on smoking participation is 1.4 percent, which is small and not statistically significant. The effect of program participation on smoking suggests that participating in the program increased adolescent smoking rates by 2.6 percentage points. This effect is not statistically significant but also less precisely estimated. Because the confidence interval is more on the positive side, smoking of current non-participants would increase slightly if the poverty threshold was moved to cover better-off households and the current non-eligibible became eligible and participate.

Table 3.5 Effect of Eligibility on Program Participation and Smoking and Effect of Participation in *Oportunidades* on Smoking for Adults and Adolescents.
(Estimations based on the preferred flexible parametric specification)**

	Effect of:			
	Eligibility on participation Π_0^*	**Eligibility on smoking $\Pi_1^{\&}$**	**Participation on smoking $\theta = \Pi_1/\Pi_0^{\ddagger}$**	**Observations**
Adults	0.536	-0.003	-0.005	3370
	(0.028)	(0.011)	(0.021)	
Adolescents	0.536	0.014	0.026	2070
	(0.034)	(0.019)	(0.036)	

*Notes: ** The preferred estimates are based on the Schwarz (1978) criterion. They are estimated using first order polynomials interacted with a dummy for eligibility to Oportunidades on both sides of the eligibility cutoff.*

** The effect of eligibility on participation corresponds to π_0 in equation (8) in the text. This is the coefficient of the dummy for eligibility coming from the First Stage of the TSLS regression of program participation.*

& The effect of eligibility on smoking corresponds to π_1 in equation (7) in the text. That is the coefficient on the dummy for eligibility coming from a reduced form equation of smoking.

‡ The effect of participation in Oportunidades on smoking is the local average treatment effect (LATE): $\theta = \pi_1/\pi_0$ in equation (4) in the text. This is the participation coefficient of the second stage of the TSLS.

None of the regressions include covariates. Clustered standard errors are reported in parentheses.

4.3 Specification checks

The previous subsection discussed RD results based on first order polynomial regressions. The conclusion that emerged is that, around the threshold for eligibility, adult and adolescent participants smoke at the same rates as non-participants. This is valid so long as the first order polynomial specification of the *m(.)* function is the true function of the underlying data. If the true functions do not belong to the class of first-order polynomials, the discontinuity estimates will in general be biased, and may lead to erroneous inferences of statistical significance (Lee, 2008).

Table 3.6 reports sensitivity estimates based on alternative model specifications.

Table 3.6 Program Participation on Smoking for Adults and Adolescents: Sensitivity Estimates.
(Estimations based on the second and third best parametric specifications)**

	Polynomial order:		
	First order to the left of the cutoff and second to the right	**Second order on both sides of the cutoff**	**Observations**
Adults	0.004	0.076	3370
	(0.028)	(0.050)	
Adolescents	-0.008	0.094	2070
	(0.033)	(0.060)	

*Notes: ** Based on Schwarz (1978), the second best specification includes a first order polynomial term to the left and to the right of the cutoff and a second order polynomial term only to the right of the cutoff (see column 2). The third best specification includes second order polynomial terms on both sides (see column 3).*

All entries are local average treatment effects (LATE) of participation in Oportunidades an average of four years on smoking. These estimations come from the Second Stage of the TSLS regression and represent $\theta = \pi_1/\pi_0$ *in equation (4) in the text.*

None of the regressions include covariates. Clustered standard errors are reported in parentheses.

Based on the Schwarz criterion, the second-best specification includes a second order polynomial only to the right of the threshold. As before, the coefficients are not-significant. However, the sign of the coefficients is reversed. The confidence interval implied by the standard errors is similar to that of the preferred model (+/-5 percent). For completeness, I also report results from a model in which second order polynomial terms are included on both sides of the cutoff. The coefficients of this model remain not significant. Moreover, given that this model is far from being a good fit of the underlying data, it is not surprising that the standard errors increase. All in all, the specification checks presented here do not contradict the main finding of this chapter: participation in *Oportunidades* did not affect current smoking rates of adults or adolescents.

5. Internal Validity of the RD estimates

This section's aim is to provide evidence in favor of a randomization around the *Oportunidades* eligibility cutoff, and hence, of the validity of the estimates. The RD impact estimates reported in the previous section are credible so long as the mean outcomes of individuals marginally above the threshold identify the true counterfactual of those marginally below the threshold. For that to be case, the individuals who barely made it to be eligible to participate in the program should be similar in observable and unobservable characteristics determining smoking behaviors to those who almost made it to be eligible. This is analogous to conducting a randomized experiment at the threshold (Hahn et al., 2001). It is precisely this randomization process what will be tested in this section.

As Lee (2008) emphasizes, randomization around the threshold is linked to how much control individuals have on the assignment variable (i.e. the poverty score). Manipulation can be complete or partial (McCrary, 2008). Complete manipulation occurs when the poverty score is entirely under the control of the agent. Partial manipulation occurs when the agent has some control of the assignment variable. Yet, the probability of receiving treatment lies somewhere between 0 and 1 due to an idiosyncratic element.

Partial manipulation of the poverty score might have occurred. Back in 1997, when the first *Oportunidades* survey was carried out, the interviewed families knew that an anti-poverty program was going to be implemented. It is possible that households (or interviewers) reported (registered) some of the variables with error aiming that this would impact the final eligibility assignment. However, complete manipulation was very unlikely because the variables determining the poverty score were not public information until after the program was evaluated for the first time in 1999.

Meaningful parameters using the RD design can be obtained even in the presence of partial manipulation as long as one can prove that the randomization around the threshold worked (Lee, 2008).The critical assumption is that each 'type' of person has an equal chance of scoring just below or just above the threshold. For that to be the case, conditioning on the 'type' of the individual, the density function of the poverty score should be continuous. Since the conditional density is not observable, testing for a discontinuity at the cutoff in the observable density function of the poverty score is informative about the randomization process. This is because continuity of the conditional density implies continuity of the poverty score density.

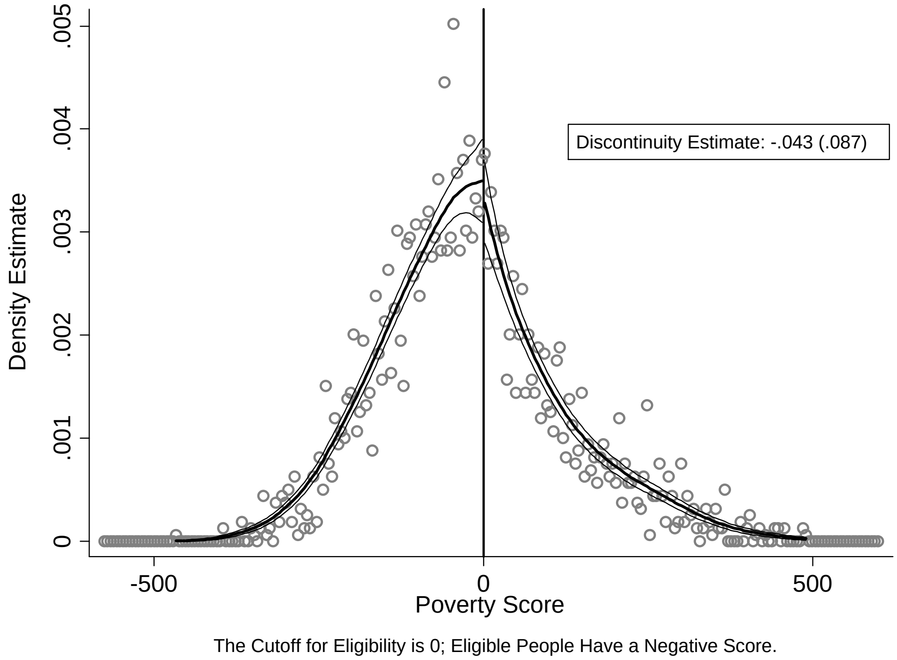
I test the null hypothesis of zero discontinuity in the density function of the poverty score at the threshold using the Wald test proposed by McCrary (2008), which I estimate through local linear density techniques.[15] Figure 3.3, plots the conditional expectation of this test along with confidence intervals for adults (top) and adolescents (bottom).[16] Both the graphical analysis and the point estimates (first row, Table 3.7) establish that the small discontinuity in the log of the baseline poverty score density for adults is not statistically significant. Nonetheless, the discontinuity in the log of the baseline poverty score density of adolescents is statistically significant. The discontinuity of the density of poverty scores for adolescents casts doubt on the identifying assumption, but it does not prove lack of randomization.[17] I

[15] I would like to thank Justin McCrary for providing the Stata code to perform the McCrary (2008) test.

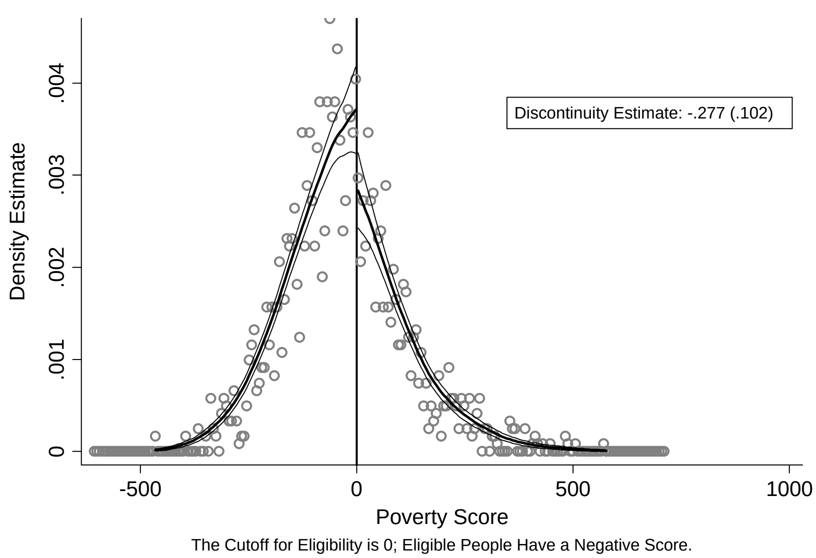
[16] Normal Q–Q plots for the t-test of the (true) null hypothesis of continuity, where t-tests stem from 1000 replications are available from the author upon request. They suggest that the normal distribution approximation is accurate as neither skewness nor fat tails are apparent. For details on this see McCrary (2008).

[17] In the same token, failing to reject the null hypothesis of zero discontinuity at the threshold is not conclusive of randomization around the threshold. See McCrary (2008) for a great example on this.

Panel A. Discontinuity Estimates for Adults.



Panel B. Discontinuity Estimates for Adolescents.



Notes: This figures provide discontinuity estimates from the test proposed by McCrary (2008, 703).

Standard errors are reported in parentheses.

Figure 3.3 Density and Confidence Intervals of the Poverty Scores that Determined Program Eligibility for Adults and Adolescents: A test of the Validity of the RD Design.

Table 3.7 Assessment of the Validity of the RD Design in the Sample of Adults and Adolescents: Discontinuity Estimates.

	Estimates for:	
	Adults	**Adolescents**
Panel A. Log Discontinuity		
Poverty Score that determined program eligibility	-0.043	-0.277
	(0.087)	(0.102)
Panel B. Discontinuity Estimates of Preset characteristics		
B.1 Correlated with smoking		
Female	0.006	-0.040
	(0.026)	(0.038)
Age	-2.108	0.013
	(0.829)	(0.120)
Literate	0.015	-0.018
	(0.027)	(0.017)
Married	0.036	n.a.
	(0.019)	(-)
Single	n.a.	0.001
	(-)	(0.007)
Indigenous	-0.026	-0.062
	(0.034)	(0.044)
Cigarette prices	-0.046	0.099
	(0.092)	(0.140)
B.2 Correlated with baseline poverty		
Despensa	-0.043	n.a.
	(0.025)	(-)
Niños de Solidaridad	n.a.	-0.025
	(-)	(0.039)
Land property	-0.011	-0.006
	(0.034)	(0.042)

Notes: The entries in Panels A, B and C represent the coefficient on the eligibility dummy of a regression similar to equation 6 in the text. It includes first order polynomial interaction terms of the poverty scores and the dummy for eligibility on both sides of the cutoff. Clustered standard errors of that coefficient are in parentheses.

Panel A reports discontinuity estimates from the test proposed by McCrary (2008, 703).

The discontinuities in Panel B are for preset characteristics measured in 1997 from the ENCASEH-97. For this set of regressions estimates with up to second order polynomial terms are very similar and are available from the author upon request. Cigarette prices are state-level prices in 1997 published by the Central Bank of Mexico. See the notes in Table 3 for the definition of Niños de Solidaridad and Despensas.

Table 3.7 (Continuation)

	Estimates for:	
	Adults	Adolescents
Panel C. Discontinuity Estimates based on all available baseline covariates and excluding participation		
Effect of eligibility on Participation	0.005	0.000
	(0.012)	(0.019)
Panel D. Discontinuity estimates including covariates		
Effect of Program Participation on Smoking	-0.005	0.005
	(0.020)	(0.032)

Notes: The entries in Panels A, B and C represent the coefficient on the eligibility dummy of a regression similar to equation 6 in the text. It includes first order polynomial interaction terms of the poverty scores and the dummy for eligibility on both sides of the cutoff. Clustered standard errors of that coefficient are in parentheses.

Panel C includes age, gender, marital status, years of education, and cigarette prices measured in 1997. There were no tobacco control laws in place at that time.

Panel D reports the effect of participation in Oportunidades on smoking. This is the local average treatment effect (LATE): $\theta = \pi_1/\pi_0$ in equation (4) in the text, which corresponds to the participation coefficient of the second stage of the TSLS. The covariates included in this regression are age, gender, marital status, years of education, cigarette prices and state clean indoor air policies as of 2003.

illustrate this point through an example. Suppose that individuals did not exercise any control over their poverty score, but interviewers did. In an attempt to make eligible to the program as many people as possible, poverty ‘points’ could have been given to those who barely failed the eligibility cutoff. This behavior would have caused the discontinuity of the density of poverty scores that we observe. In spite of that, the identifying assumptions would not be violated so long as points were given randomly to the non-eligible. In contrast, giving points based on characteristics unobserved by the econometrician would damage the research design. The remaining of this section discusses alternative tests that attempt to provide evidence against the hypothesis that the ineligible were made eligible based on unobservables.

Testing whether the baseline characteristics of individuals in the neighborhood of the cutoff are similar is an alternative way to provide evidence in favor of pre-program randomization at the threshold.[18] As Lee (2008) recognizes, the sample average in a narrow neighborhood of the eligibility cutoff would in general be a biased estimate of the true conditional expectation function at the 0 threshold when that function has a non-zero slope. To address this problem, polynomial approximations (as in equation (7)) are used to generate simple estimates of the discontinuity of baseline characteristics at the threshold for eligibility. Failure to reject the null hypothesis of no discontinuity ensures that the conditional expectation of any baseline characteristic in the poverty score is continuous, which implies that the unobservable conditional density of the poverty score, of main interest to us, is continuous (Lee, 2008). Ideally, I would test for discontinuities of smoking behaviors before program intervention, but these data are not available in the baseline survey. Hence, I test the hypothesis of no discontinuities on a set of covariates that are correlated with smoking status, and other characteristics that are associated with poverty, but did not compose the poverty scores.

Each cell in Panel B of Table 3.7 represents the discontinuity estimates at the threshold for eligibility of the baseline variables defined in the first column of the table. For instance, the 0.006 coefficient at the intersection of the "female" row and the "adult" column suggests that the female rate before the program started was similar for observations "just above" and "just below" the threshold. The rest of the discontinuities at the cutoff are small and not statistically significant providing evidence in favor of the "randomization" at the threshold for both the adult and adolescent samples. The only exception to this rule is the age coefficient for adults. However, under the null hypothesis that the covariates around the threshold are balanced and independent, we would expect 5 percent of the discontinuity estimates to

[18] This is equivalent to the standard test of randomization in an experimental design, using a test of the equality of the mean of every variable in covariates across treatment and control groups.

be statistically different from zero. Moreover, as I discuss below, controlling for age in the estimation of program effects on smoking behaviors does not change the main results of this chapter.

The next diagnostic tests predicts smoking status as a function of poverty scores, a dummy for eligibility and all the available baseline covariates related to smoking, but excludes program participation. These estimations contain all the information that could possibly predict smoking behaviors (aside from participation). Therefore, if individuals just above and just below the cutoff are nearly identical, baseline characteristics should not predict a discontinuity in smoking behaviors. These predictions are reported in Panel C of Table 3.7. The size of the discontinuities is tiny and not statistically different from zero in both samples, thus favoring the hypothesis of similarity of individuals around the threshold.

The last test involves estimating program effects with covariates included. In the presence of local random assignment, the point estimates of the impact of the program should be insensitive to the inclusion of any combination of baseline covariates (Imbens and Lemiux, 2007; Lee, 2008). In practice, if the covariates are correlated with the potential outcomes, tossing them in the regression may eliminate the biases that result from the inclusion of observations far away from the threshold, which would ultimately improve the precision of the estimates. The bottom part of Table 3.7 presents the estimates of the effect of *Oportunidades* on smoking participation, θ using a flexible parametric model that includes covariates. It is reassuring to see that the adult estimates are exactly the same as the ones where the covariates were excluded ("participation on smoking" column, Table 3.5). Furthermore, the standard errors in the model that includes covariates are slightly smaller, as expected. In the adolescent sample, not only the standard errors in the regression that includes covariates drop to zero, but also the point estimate.

This section showed that adults and adolescents around the neighborhood of the cutoff are nearly identical in terms of observable characteristics. The various tests supporting this argument also suggest that the discontinuity in the density of adolescent poverty scores was the result of a random sorting. The available evidence furnishes the non-testable hypothesis of equal unobservable characteristics. It furthermore suggests that individuals were locally randomized, and that the zero impact estimates on smoking behaviors reported in this paper are internally valid.

Despite that the results are consistent, the lack of significant effect found between participants and non-participants could potentially be contaminated by the presence of measurement error in self-reported smoking status. Cigarette smoking tends to be underreported in places where there is smoking-related social stigma (Rossiter, 2009). Underreporting due to social stigma is a possibility in the *Oportunidades* context. Smokers might have also had incentives to underreport if they believed that their smoking status was associated with a decreased probability of being eligible to participate in the program in the future. This would be the case, for example, if *Oportunidades* officials expected that the health information sessions would lead to a decrease in smoking. In contrast, cigarette consumption could have been overreported if smoking was associated with a "status" value. A comparison of the self-reported with true smoking status coming from biochemical markers would be necessary to address measurement bias, but information on true smoking status is not available. Therefore, the estimates above rest on the assumption that the self-reported data is accurate.

6. Discussion

The evidence provided in this chapter indicates that program participation in *Oportunidades* caused no effect on the smoking participation decisions of adults, and hence on health outcomes. This conclusion is supported by both the point estimates and the narrow confidence intervals implied by the standard errors. From the local average treatment effect (LATE) on the treated of this paper it is possible to predict what would have happened with the smoking rates of 'compliers' in the absence of treatment. Putting it differently, the LATE tells us what would have happened to the smoking rates of non-participants had the threshold for eligibility being moved to the right in order to cover better-off households. This effect is relevant in programs like *Oportunidades* where the evidence of positive impacts increases the probability of program expansions. [19]

Matching individuals in treatment and control communities based on "pre-program characteristics" coming from retrospective information, Duarte Gómez et al. (2005) find a zero average treatment effect on the treated on smoking. This effect is the same as the LATE estimated in this essay. Consequently, the no effect of *Oportunidades* on smoking generalizes to the participant population.

Previous research regarding the average treatment effect on the smoking behaviors of participant youth is less conclusive. Gutiérrez et al. (2005) find no difference in the smoking rates of short-term (up to 3 years) participants and those of non-participants. Duarte Gómez et al. (2005) find that the smoking rates of short-term program participants were 26 percent lower than those of long-term (up to 5.5 years) participants. Combining this evidence suggests that long-term participation in *Oportunidades* increased the smoking rates of the average adolescent participant. However, Gutiérrez et al. (2005) find the average long-term effect on adolescent current smoking to be not statistically different from zero. The long-term local average

[19] See De la Torre (2005) for the latest summary of these effects.

treatment effects documented in this paper are also not statistically different from zero. Nonetheless, the point estimate is positive and less precisely estimated.

What processes were involved in causing this non-result? As program benefits where simultaneously given, disentangling each of the effects is possible to the extent that similar people received heterogeneous program benefits. For instance, suppose that eligible households were randomly assigned to two groups. Now suppose that both groups received cash transfers but only one was required to go to the information sessions. Applying the logic of difference-in-difference quasi-experimental estimators, program effect differences between these two groups could be causally attributed to the information sessions.

Differential treatments between women and men in *Oportunidades* can be used to isolate the income effect on smoking participation among adults. Households in *Oportunidades* were given cash transfers, but only women were required to attend health information sessions (Table 3.2). Therefore, program effects can be interpreted as income effects among men, and income-information effects among women. Moreover, taking the difference between the program impact effects of men and women participating sheds light on the magnitude of the information effect. This exercise is valid if (a) there are no reasons to believe that poor men and women in rural Mexico adjust their smoking behaviors differently to income and information shocks, and (b) there are no spillover effects of information. For the same reasons, differences in adolescent and women program effects can be interpreted as a schooling effect. However, women might be a poor counterfactual for adolescents given that these populations are at different points of their life cycle and economic incentives are expected to affect their smoking behaviors differently.

Table 3.8 reports the effects of *Oportunidades* on current smoking for men and women. The Schwarz preferred program impacts include only first order polynomials (first column). The point estimates of the effect of *Oportunidades* on male smoking, which can be interpreted as an income effect, are not statistically different from zero. Nonetheless, the standard errors imply an effect between +/- 10 percentage points, approximately. Program effects among women are more precisely estimated, but are also non significant. These conclusions are nearly insensitive to changes in model specification (second column) or the inclusion of covariates (third column). Furthermore, I fail to reject the hypothesis of no discontinuity in the density of the poverty score for men and women (Appendix Figure A3.2), which suggest that the results by gender are internally valid.

Table 3.8 Isolating the Income Effect of the Program:
Effect of Participation in *Oportunidades* on Smoking by Gender.
(Estimations based on the preferred and second best flexible parametric specifications)**

	Polynomial order:			
	First order on both sides of the cutoff	**First order to the left of the cutoff and second to the right**	**First order on both sides of the cutoff**	**Observations**
Men	-0.011	0.014	-0.027	906
	(.063)	(.082)	(0.063)	
Women	-0.002	-0.001	-0.002	2406
	(0.011)	(0.013)	(0.011)	
Regression includes Covariates?	No	No	Yes	-

*Notes: ** Based on Schwarz (1978), the first best specification includes first order polynomial terms on both sides of the cutoff for eligibility (see columns 2 and 4). The second best specification includes first order a first order polynomial term to the left and to the right of the cutoff and a second order polynomial term only to the right of the cutoff (see column 3).*

All entries are local average treatment effects of participation in Oportunidades an average of four years on smoking: $\theta = \pi_1/\pi_0$ *in equation (4) in the text. They correspond to the participation coefficient of the second stage of the TSLS.*

Only the regressions in the fourth column include covariates. Clustered standard errors are reported in parentheses.

The difference between program effects of men and women imply a zero health information effect. If so, information dissemination policies might not be the way to reduce cigarette smoking in Mexico. This result is interesting, but should be taken with caution for at least two reasons. First, because it assumes that men are indeed a good counterfactual for women. Second, because the lack of precision of the income effect of *Oportunidades* on smoking among men translates in information effects lying anywhere between +/- 12 percentage points.

7. Conclusions

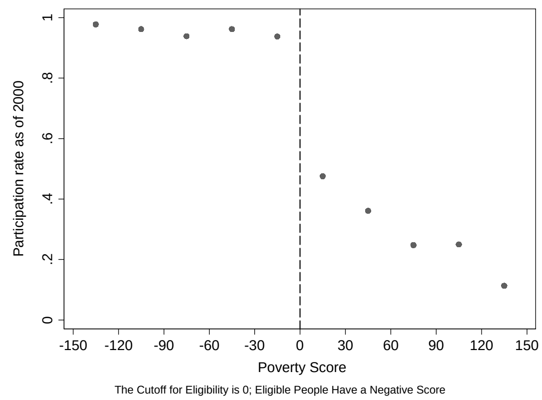
The goal of the *Oportunidades* program is to eradicate poverty in Mexico through investments in human capital in the form of schooling, nutrition and health. Previous research on the health impacts of this program found that short-term participation increased the utilization of public health clinics for preventive care (Gertler, 2000). Along with medical utilization, economic theory predicts that health outcomes are also determined by health behaviors such as smoking, exercising and eating healthy to avoid obesity (see Grossman, 1972).

This essay used program eligibility as an instrument for participation in *Oportunidades* to estimate the effect of this intervention on adult and adolescent smoking via a fuzzy Regression Discontinuity design. The benefits of the program include sizeable cash transfers, health information sessions and schooling. Based on economic theory predictions and previous empirical findings there were reasons to predict that *Oportunidades* would affect smoking among poor Mexicans through each of these benefits. The findings of this chapter suggest, however, a zero local average treatment effect on adults that participated in the program an average of four years. This effect compares to the average treatment effects found in previous literature, suggesting that the impact of the program is homogeneous across the poor. Income, schooling and information jointly did not change smoking among adolescents. Because these estimates were less precisely estimated, the worst case scenario points to a slight increase of smoking among long-term adolescent participants.

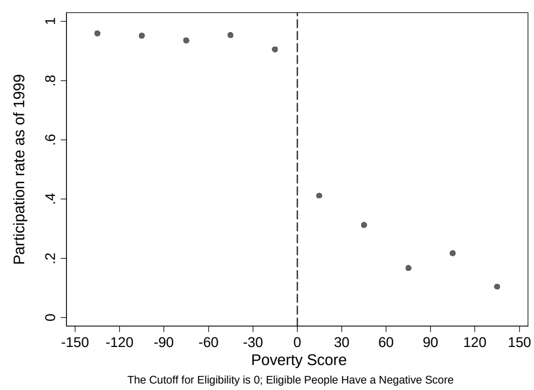
Because of the differential treatments between men and women in the *Oportunidades* program, the income effect was isolated by estimating the program's impact on adult male smoking. The point estimate of the effect of *Oportunidades* on smoking among men was not statistically different from zero. Disappointingly, it was not precisely estimated. Therefore, it is not conclusive of a null income effect on smoking. The analysis by gender, however, indicated that health and income combined did not have an effect on the smoking behaviors of participant women.

APPENDIX

Panel A. Program Participation Rates in 2000 in Communities where the Program Started in 1998 (treatment-98)



Panel B. Program Participation Rates in 1999 in Communities where the Program Started in 1998 (treatment-98)



Panel C. Program Participation Rates in 2000 in Communities where the Program Started in 1999 (treatment-99)

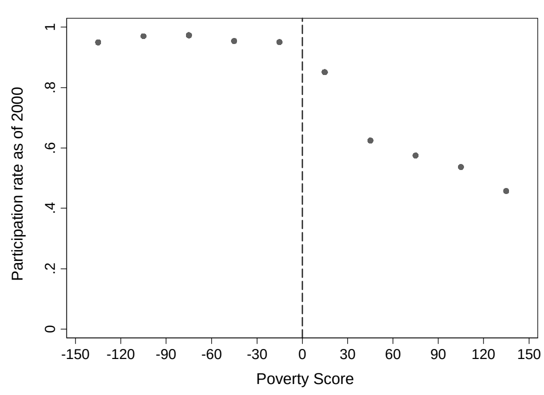
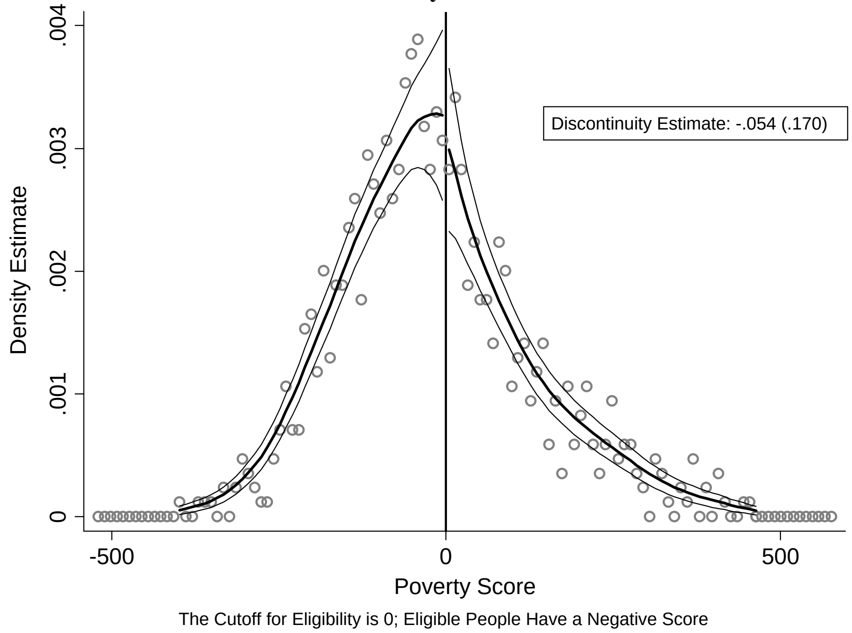
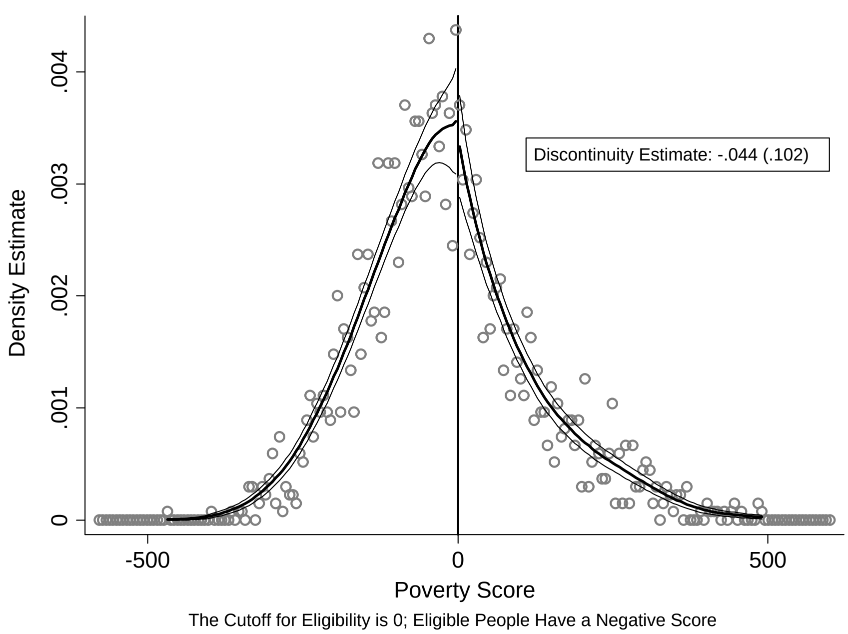


Figure A3.1 Average Participation Rates Over Time in Different Communities as a Function of the Poverty Scores that Determined Program Eligibility.

Panel A. Discontinuity Estimates for Men



Panel B. Discontinuity Estimates for Women



Notes: This figures provide discontinuity estimates from the test proposed by McCrary (2008, 703).

Standard errors are reported in parentheses.

Figure A3.2 Density and Confidence Intervals of the Poverty Scores that Determined Eligibility by Gender: A test of the Validity of the RD Design.

REFERENCES

Adato, Michelle, David Coady, and Marie Ruel. 2000. "Evaluación de operaciones de Progresa desde la perspectiva de los beneficiarios, las promotoras, directores de escuela y personal de salud." Sedesol: Mexico City.

Baum, Christopher F. 2006. *An introduction to modern econometrics using Stata.* College Station, Tex.: Stata Press.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago: The University of Chicago Press.

Becker, Gary S. and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96:4, pp. 675-700.

Behrman, Jere R., Piyali Sengupta, and Petra Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change*, 54:1, pp. 237-75.

Buddelmeyer, Hielke and Emmanuel Skoufias. 2004. "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA." *IZA Discussion Paper No. 827*: Bonn, Germany.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics*, 118:4, pp. 1495-532.

Chaloupka, Frank and Kenneth Warner. 1999. "The Economics of Smoking." *NBER Working Paper No. 7047* Cambridge, MA.

DeCicca, Philip, Donald S. Kenkel, and Alan Mathios. 2008. "Cigarette Taxes and the Transition from Youth to Adult Smoking: Smoking Initiation, Cessation, and Participation." *Journal of Health Economics*, 27:4, pp. 904-17.

de la Torre, Rodolfo 2004. "External Evaluation of the Impact of the *Oportunidades* Human Development Program," in *Evaluación externa de impacto del Programa Oportunidades 2004. Salud.* Hernández-Prado Bernardo and Mauricio Hernández-Avila ed. Cuernavaca: Instituto Nacional de Salud Pública.

de Walque, Damien. 2007. "Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education." *Journal of Health Economics*, 26:5, pp. 877-95.

Dinardo, John and David S. Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001." *The Quarterly Journal of Economics*, 119:4, pp. 1383-441.

Djebbari, Habiba and Jeffrey Smith. 2008. "Heterogeneous impacts in PROGRESA." *Journal of Econometrics*, 145:1-2, pp. 64-80.

Duarte Gómez, María B., Sonia Morales Miranda, Alvaro J. Idrovo Velandia, Sandra C. Ochoa Marín, Siemon Bult van der Wal, Marta Caballero García, and Mauricio Hernández Ávila. 2004. "Impact of *Oportunidades* on knowledge and practices of beneficiary mothers and young scholarship recipients. An evaluation of the educational health sessions." *Instituto Nacional de Salud Pública*: Cuernavaca.

Farrell, Phillip and Victor R. Fuchs. 1982. "Schooling and Health: The Cigarette Connection." *Journal of Health Economics*, 1:3, pp. 217-30.

Fields, Gary S. 2001. *Distribution and Development: A New Look at the Developing World*. New York: Rusell Sage Foundation.

Frölich, Markus. 2007. "Regression discontinuity design with covariates." *University of St. Gallen, Working Paper 2007-32*: St. Gallen Switzerland.

Gertler, Paul. 2000. "The Impact of Progresa on Health." International Food Policy Research Institute: Washington, DC.

González de la Rocha, Mercedes and Agustín Escobar Latapí. 2001. "Primeros resultados de la Evaluación cualitativa basal del Programa de Educación, Salud y Alimentación (PROGRESA) semiurbano." CIESAS: Mexico City.

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy*, 80:2, pp. 223-55.

Grossman, Michael. 2004. "The demand for health, 30 years later: a very personal retrospective and prospective reflection." *Journal of Health Economics*, 23:4, pp. 629-36.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *Quarterly Journal of Economics*, 116:4, pp. 1261-303.

Gutiérrez, Juan Pablo, Paul Gertler, Mauricio Hernández, and Stefano Bertozzi. 2004. "Impact of *Oportunidades* on the risky behaviors of adolescents and their immediate consequences. Short term results in urban areas and medium term results in rural areas," in *Evaluación externa de impacto del Programa Oportunidades 2004. Salud*. Hernández-Prado Bernardo and Mauricio Hernández-Avila ed. Cuernavaca: Instituto Nacional de Salud Pública.

Hahn, Jinyong, Petra Todd, and Van der Klaauw Wilbert. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica*, 69:1, pp. 201-09.

Hammond, David, Geoffrey T. Fong, Ron Borland, K. Michael Cummings, Ann McNeill, and Pete Driezen. 2007. "Text and Graphic Warnings on Cigarette Packages Findings from the International Tobacco Control Four Country Study." *American Journal of Preventive Medicine*, 32:3, pp. 202–09.

Heckman, James J. 2008. "Econometric Causality." *NBER Working Paper No. 13934*: Cambridge, MA.

Hoddinott, John, Emmanuel *Skoufias*, and Ryan *Washburn*. 2000. "The impact of PROGRESA on consumption: A final report". *International Food Policy Research Institute*: Washington, DC.

Ibáñez Hernández, Norma A. 2005. "Disposiciones jurídicas federales sobre la prohibición de fumar," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco, México*. Raydel Valdés-Salgado, Eduardo C. Lazcano-Ponce and Mauricio Hernández-Avila eds. Cuernavaca: Instituto Nacional de Salud Pública.

Imbens, Guido and Thomas Lemieux. 2007. "Regression Discontinuity Designs: A Guide to Practice." *NBER Working Paper No. 13039*: Cambridge, MA.

Instituto Nacional de Estadística, Geografía e Informática (INEGI). 2004. "Encuesta Nacional de Adicciones, ENA-2002." INEGI: Aguascalientes.

Jiménez-Ruiz, J. A., Belén Sáenz de Miera, Luz M. Reynales-Shigematsu, Hugh. R. Waters, and Mauricio Hernández-Avila. 2008. "The impact of taxation on tobacco consumption in Mexico." *Tobacco Control*, 17:2, pp. 105–10.

Kenkel, Donald S. 1991. "Health Behavior, Health Knowledge, and Schooling." *Journal of Political Economy*, 99:2, pp. 287-305.

Kenkel, Donald, Dean Lillard, and Alan Mathios. 2006. "The Roles of High School Completion and GED Receipt in Smoking and Obesity." *Journal of Labor Economics*, 24:3, pp. 635-60.

Lee, David S. 2008. "Randomized experiments from non-random selection in U.S. House elections" *Journal of Econometrics*, 142:2, pp. 675-97.

Mayer-Foulkes, David. 2001. "The Long-Term Impact of Health on Economic Growth in Mexico,1950-1995" *Journal of International Development*, 13:1, pp. 123-26.

Matsudaira, Jordan. 2008. "Mandatory summer school and student achievement." *Journal of Econometrics*, 142:2, pp. 829-50.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142:2, pp. 698-714.

Mullahy, John. 1985. "Cigarette Smoking: Habits, Health Concerns and Heterogeneous Unobservables in a Micro-econometric Analysis of Consumer Demand." Ph.D. Dissertation. University of Virginia: Charlottesville.

Nuño-Gutiérrez, Bertha L., José Álvarez-Nemegyei, and Eduardo A. Madrigal-de León. 2008. "Efecto de una intervención antitabaco en estudiantes de enseñanza media superior en Guadalajara, México." *Salud Mental*, 31:3, pp. 181-88.

Parker, Susan W., Petra E. Todd, and Kenneth I. Wolpin. 2006. "Within-Family Program Effect Estimators: The Impact of *Oportunidades* on Schooling in Mexico." *Mimeo*. University of Pennsylvania: Philadelphia.

Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." *Mimeo*. Harvard University: Cambridge, MA

Rossiter, John C. 2009. "A comparison of social desirability bias among four widely used methods of data collection as measured by the impression management subscale of the balance inventory of desirable responding ". Ph.D. Dissertation. Kent State University: Kent, OH.

Rubalcava, Luis N. and Graciela Teruel Belismeli. 2003. "Análisis sobre el cambio en variables demográficas y económicas de los hogares beneficiarios del Programa *Oportunidades* 1997-2002." *Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades*. Instituto Nacional de Salud Pública and CIESAS: Mexico City.

Sáenz de Miera, Belen, Jorge A. Jiménez, and Luz M. Reynales. 2007. "La economía del tabaco en México." *Instituto Nacional de Salud Pública*: Cuernavaca.

Schwarz, Gideon 1978. "Estimating the dimension of a model." *The Annals of Statistics*, 6:2, pp. 461-64.

Sebrie, Ernesto. 2006. "Mexico: backroom deal blunts health warnings." *Tobacco Control* 15:5, pp. 348-49.

Secretaria de Desarrollo Social. 2003. "Acuerdo por el que se emiten y publican las Reglas de Operación del Programa de Desarrollo Humano *Oportunidades* para el Ejercicio Fiscal 2003." Diario Oficial de la Federación.

Skoufias, Emmanuel, Benjamin Davis, and Jere R. Behrman. 1999. "An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico." International Food Policy Research Institute: Washington, DC.

Skoufias, Emmanuel and Bonnie McClafferty. 2001. "Is Progresa Working? Summary of the Results of an Evaluation by IFPRI " *International Food Policy Research Institute, FCND Discussion Paper No. 118*: Washington, DC.

Trochim, William M. K. 1984. *Research design for program evaluation: the regression-discontinuity approach*. Beverly Hills: Sage Publications.

van der Klaauw, Wilbert 2008. "Regression–Discontinuity Analysis: A Survey of Recent Developments in Economics." *Labor*, 22:2, pp. 219-45.

van der Klaauw, Wilbert. 2008b. "Regression-Discontinuity Analysis " in *The New Palgrave Dictionary of Economics*. Steven N. Durlauf and Lawrence E. Blume eds. London: Palgrave Macmillan.

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

CHAPTER 4

THE IMPACT OF SCHOOLING ON SMOKING

1. Introduction

Education is widely considered to be critical for development. Aside from its effect on labor outcomes in Mexico, the returns to schooling might also include better health through less unhealthy behaviors (Grossman, 1972; 2004).[20] For Mexicans in their twenties, smoking is more prevalent among men than women, but schooling is associated with less smoking among men and more smoking among women (see Table 4.1). For both men and women the gradient occurs at relatively low levels of education. In this chapter I measure and disentangle the correlation between schooling and smoking, separately for men and women.

Table 4.1 Percentage of People who are Current Smokers and Healthy by Schooling Level and Gender.

Schooling Level	Current Smoker:		Healthy:	
	Men	**Women**	**Men**	**Women**
Elementary	25.5	2.2	56.3	54.2
Middle school	23.5	4.8	64.8	60.3
High school	19.9	7.3	73.7	66.8
College	15.7	8.6	86.2	75.9

Notes: Current smokers are people who reported to ever have had the habit of smoking cigarettes and smoked at least one cigarette a day when they were interviewed.

Healthy people are those who reported to be in 'Excellent' or 'Very good health the day they were interviewed.

Source: Own calculations based on MxFLS-2, individuals aged 20 to 29 years.

In the economics literature, explanations for the correlation between school attainment and health behaviors are not only limited to a causal effect of schooling on less unhealthy behaviors. Two other not necessarily mutually exclusive hypotheses

[20] See Mayer-Foulkes (2006) for a review of the literature on the returns to schooling in Mexico.

have emerged. The first poses the existence of a reverse causal effect, from health behaviors to schooling. The second suggests that the correlation between education and investments in health is caused by unobserved heterogeneity. That is, difficult-to-observe 'third variables' that affect both schooling and health behaviors, such as family background and time preferences (Fuchs, 1982; Farrell and Fuchs, 1982).

Most of the recent work has exclusively focused on obtaining consistent estimates of the effect of schooling on smoking and other health behaviors. One exception is Cutler and Lleras-Muney (2010) who analyze an array of factors explaining the relationship between schooling and health behaviors in the U.S., but do not investigate if the correlation is causal. Two other exceptions are Kenkel et al. (2006) and Fletcher and Frisvold, (2009), who have tested both the 'third-variable' hypothesis and a causal effect of schooling on different health behaviors using a single source of data. Both papers find that family background and ability attenuate, but do not eliminate, the schooling-health behaviors correlation in the U.S, thus favoring the hypothesis of a causal effect.

Table 4.2 summarizes the literature that has taken steps to deal with the potential endogeneity of schooling when estimating the effect of schooling on health behaviors.[21] U.S.-based papers, which constitute the majority of the studies, do not unanimously find significant effects of schooling on health behaviors. Negative and significant effects have been found on smoking during pregnancy (Currie and Moretti, 2003) and on male smoking (de Walque, 2006; Grimard and Parent, 2007; Kenkel, et al. 2006). Interestingly, using a very similar instrument but different datasets de Walque (2006) finds a significant effect on smoking cessation, but Grimard and Parent (2007) do not. Moving to other health behaviors, Fletcher and Frisvold (2009) find

[21] A concise survey of the literature examining causal effects of schooling on other dimensions of health can be found in Albouy and Lequien (2009).

Table 4.2 Summary of the Recent Literature that has Addressed the Potential Endogeneity of Schooling when Estimating the Returns to Schooling on Health Behaviors.

Authors	Year	Goal*	School margin[&]	Health measure	Country	Sample Size	Year	Data	Identification Strategy	Causal effect?
de Walque	2003	CE	Years	Smoking	Indonesia	3,000	1993	IFLS	"School supply" based on community surveys.	Yes/ No
Spasojevic	2003	CE	N.A	Health Index, BMI	Sweden	N.A	1981-91	N.A	Change in compulsory education laws.	Yes/ No
Currie and Moretti	2003	CE	College	Smoking during pregnan.	USA	166,183	1970-99	Vital Statistics Natality	Colleges in the woman's county in her seventeenth year.	Yes
de Walque	2004	CE	College	Smoking	USA	370,000	1978–2000	NHIS,	Vietnam era draft lottery.	Yes
Arendt	2005	CE	Years	BMI, Smoking	Denmark	3,300	1990, 95	SRH WECS	1958 decrease test barriers & increase school proximity. 1975 sch. leaving age 7 to 9.	No
Kenkel, Lillard, and Mathios	2006	CE/ TF	High School/ GED	BMI, Smoking	USA	6,500	1998	NLSY79	5 education policy variables, state level (expenditures, etc.)	Yes/ No

*Notes: *CE is the causal effect of the schooling margin listed in the table on the health behavior of interest in each article. RC refers to the effect of the health behavior on schooling.TF stands for third factor hypothesis, which suggests that schooling and health behaviors are caused by unobserved heterogeneity.*

[&] Years indicates to Years of Schooling.

Source: Please refer to the References.

Table 4.2 (Continued)

Authors	Year	Goal*	School margin[&]	Health measure	Countr y	Sample Size	Year	Data	Identification Strategy	Causal effect?
de Walque	2006	CE	College	Smoking (Cessation)	USA	80,000	1983–95	NHIS	Vietnam era draft lottery.	Yes
Grimard and Parent	2007	CE	Years	Smoking	USA	227,000	1995, 96, 98, 99	CPS Suppl.	Vietnam war draft avoidance.	Yes/ No
Parka and Kang	2008	CE	Years	Smoking, drinking, exercise, health checkups	South Korea	1,611	2001	KLIPS-4	School construction data in the mid-70s and birth order.	Yes/ No
Fletcher and Frisvold	2008	CE/ TF/ RC	College	Preventive care	USA	5,578	1957,64,75, 92-93 2003-04	WLS	Sibbling fixed effects.	Yes
Grabner	2008	CE	Years	Obesity	USA	11,874/ 11,214	1971 to 75 1976 to 80	NHANES 1 & 2	State Laws on Min. Schooling to apply for Work Permit (1914-1978).	Yes
MacInnis	2008	CE	College	Obesity	USA	N.A	N.A	N.A	Pre-lottery Vietnam.	Yes

*Notes: *CE is the causal effect of the schooling margin listed in the table on the health behavior of interest in each article. RC refers to the effect of the health behavior on schooling.TF stands for third factor hypothesis, which suggests that schooling and health behaviors are caused by unobserved heterogeneity.*

[&] *Years indicates to Years of Schooling.*

Source: Please refer to the References.

positive and significant causal effects on various measures of preventive care. In contrast with previous research by Kenkel, et al. (2006); Grabner (2008) and MacInnis (2008) present evidence of a causal impact of schooling on obesity.

Outside the U.S., the controversy of the link between schooling and health behaviors remains as well. Parka and Kang (2008) find little evidence of an effect of schooling on smoking or drinking in South Korea, but do find that education induces individuals to exercise or get regular checkups. De Walque (2003) finds a causal effect of education on smoking in Indonesia, but this result is not robust to the inclusion of geographic fixed effects. Spasojevic (2003) finds positive effects on the probability of healthy body mass index (BMI), but this result only holds when the author uses one-tailed tests. Finally, Arendt (2005) is unable to conclude whether there is a causal effect on self reported health and BMI in Denmark given the lack of precision of his instrumental variables (IV) estimates.

The state-of-the-art of the literature examining the correlation between schooling and health behaviors can be summarized as follows. First, research has mainly focused on developed countries. Second, while the 'third variable' hypothesis has been widely discussed in the theoretical literature, convincing empirical tests of this hypothesis have been hampered by data limitations. Third, the evidence on the schooling effects of schooling on health behaviors is mixed. This conclusion can in fact be extended to other measures of health; in their review of the literature Albouy and Lequien (2009) conclude that the "existence of a causal impact of education on health is not clearly established" (p.156). Fourth, most of the studies that claim to find causal effects do not tease out the mechanisms through which schooling affects health behaviors. Possible channels include income, health-related information and the efficiency to process that information to make health-related decisions, peer effects, occupation, and assortative mating.[22]

In what follows, using data from the second wave of the Mexican Family Life Survey, I contribute to this literature examining the relationship between schooling and smoking in a developing country. In the first part of the analysis I investigate if, and to what extent, usually unobserved factors are responsible for the correlation between schooling and smoking. In particular, I explore if the schooling-smoking relationships among men and women in Mexico are sensitive to the inclusion of variables that are designed to elicit individual's time preferences, expectations, innate ability, and family background in a Linear Probability Model (LPM) regression framework. I thus use a richer set of 'third variables' compared to what has been used before (Kenkel et al., 2006 and Fletcher and Frisvold, 2009). Also, and in contrast with many previous studies, conducting the analysis on men and women, gives very interesting new insights. In the second part of the analysis, I use an Instrumental Variables (IV) framework to test whether health-related returns to schooling exist. In line with previous research, I focus on identifying total effects of schooling on smoking rather than identifying pathways.

To uncover causal effects of schooling on smoking, I capitalize on an uneven and huge middle school construction program that took place in Mexico in the early nineties, just after a federal initiative proposed the amendment of the Mexican Constitution to mandate compulsory secondary schooling. Between 1992 and 1999, the number of public middle schools multiplied by 1.72 and the number of classrooms by 1.77. Most of these schools were built in rural areas (Parker et al., 2008). By reducing the costs of school attendance for the marginal person, a higher availability of classrooms in the individual's state and area of birth is expected to increase middle

[22] See Cutler and Lleras-Muney (2006, 2010) for a comprehensive discussion on this.

school attendance, but there are no reasons to believe that it would directly affect people's smoking decisions.

To instrument for middle-school attainment, I use exogenous variation in public-school classroom availability per thousand school-aged children in the state of birth when the individual was 12 years, (i.e. when he/she first became eligible to dropout from middle school). Generated this way, the instrument takes into consideration that people's opportunities to go to school are influenced by the number of schools, their size and the cohort size, which determines the potential demand. By using state of birth, instead of state of residence at age 12, the instrument also eliminates potential biases caused by classroom endogeneity (e.g. self-selection to areas with higher classroom availability). To account for the fact that unobserved state-level characteristics might affect smoking decisions, the preferred specifications include state fixed-effects. Therefore, identification comes from within-state variation in the number of classrooms available across cohorts. Given that the school expansion was intended to affect rural areas, alternative specifications use as an instrument the interaction between classroom availability and a dummy for rural-born.

To preview the results, I find that the positive correlation between schooling and smoking is caused by unobserved heterogeneity among women, but not among men. The school construction program, measured by the availability of classrooms, is proved to exert a large exogenous and positive effect on middle school attainment, particularly on rural-born men. The second-stage results of the effect of schooling on smoking are less precisely estimated than the LPM. Nonetheless, the results suggest that schooling may include less smoking among Mexican men.

The rest of this chapter is laid out as follows. Section 2 describes the data. Section 3 presents the empirical strategy to test the 'third variable' hypothesis, briefly discusses education policies in Mexico during the 1990s and describes how they lend themselves to an instrumental variables approach. Section 4 reports the estimation results and Section 5 discusses these results. The last section concludes and suggests directions for future research.

2. Data

I analyze the schooling smoking relationship using microeconomic data from the second wave of the Mexican Family Life Survey (MxFLS-2). This is a multipurpose survey representative at the national, regional and rural-urban levels (Rubalcava and Teruel, 2006). The baseline, fielded in 2002, covered approximately 35,000 individuals distributed over 141 localities in Mexico. The MxFLS-2 was able to follow-up 90 percent of all respondents from 2002, including those who move within Mexico and those who move to the United States. Most of the followed-up people were interviewed in 2005 and 2006, but given the difficulties in tracking people out, the fieldwork of the second wave was concluded in 2007.

The oldest cohort of individuals in my sample was 12 years by September 1st of 1990. According to Mexico's school entry-age requirements during the nineties, these individuals should have started middle school that year. [23] The youngest cohort in my sample was 20 years old the day of the follow-up survey. Two reasons motivated this age restriction. First, by age 20 most of the people in Mexico have already reached their definitive education level. The individuals in my sample who are still studying are finishing college, so they have been fully exposed to the junior high school construction that took place in the nineties. Second, by age 20 people should also have made their smoking decisions. Own calculations based on the National

[23] Regulations regarding the minimum school entry age were changed in June of 2006. Since then, individuals are expected to start middle school education in the school year in which they turn twelve years old by December 31st (Ministry of Education, 2006).

Health and Nutrition Survey 2006 indicate that the average age at which Mexican men start smoking is 16.9 years. The same figure is 18.5 for women.

Based on the above restrictions, 2298 men and 2808 women were eligible to be included in my sample. They should have started middle school between 1990 and 1999. 117 men and 134 women were removed from the sample due to missing information in schooling, smoking, state-of-birth and area-of-birth. Missing values in time-invariant variables were imputed based on what was reported in 2002. Time varying variables with missing values were also imputed using the information available in 2002 in combination with changes reported by the individual between 2002 and the day of the follow-up interview. I imputed the remaining (small set of) missing values using multiple imputation techniques (see Roylston, 2005). In all specifications, each imputed observation was flagged with an imputation dummy. The final sample of analysis consists of 2181 men and 2662 women. Summary statistics for women and men are reported in Table 4.3.

One of the most important advantages of the MxFLS is that all respondents were directly asked their history of schooling and smoking consumption. Current smoking participation was coded based on the responses to the following questions: "Do you/did you ever had the habit of smoking cigarettes?" and "How many cigarettes do you smoke a day?" People who answered 'yes' to the first question and a positive number to the second question were coded as current smokers. Hence, the omitted category includes never and former smokers. Given the young age of the sample used, former smokers only represent 1.05 percent of the female sample and 1.93 percent of the male sample. By 2005, 76.02 percent of men and 75.5 percent of women in my sample reported to have finished at least one year of middle school. I focus on this margin because, as indicated in Table 4.1, the schooling-smoking gradient arises at relatively low levels of education. Furthermore, Grossman (2004) suggests that the

Table 4.3 Descriptive Statistics of Individuals and State Policy Variables by Gender.

	Men		Women	
	Mean	**SD**	**Mean**	**SD**
Panel A. Individual-level variables				
Current smoker	0.191	0.393	0.057	0.232
Some middle school	0.760	0.427	0.759	0.428
Age	23.500	2.509	23.476	2.485
Currently attends school	0.149	0.356	0.131	0.338
Indigenous	0.094	0.292	0.098	0.297
Rural-born	0.522	0.500	0.532	0.499
Future oriented	0.264	0.441	0.264	0.441
Good future expectations	0.558	0.261	0.533	0.267
Risk averse	0.320	0.467	0.264	0.441
Cognitive ability	-0.024	1.002	0.013	1.002
Family resources at age 12	0.605	0.489	0.614	0.487
Parent with some middle school	0.352	0.478	0.341	0.474
Mother has some middle school	0.230	0.421	0.216	0.411
Panel B. State policy variables				
Classroom availability (per 1000 ch)	15.974	2.634	15.95	2.777
Families in *Oportunidades* in 2005	0.194	0.123	0.20	0.127
Families in *Oportunidades*, age 12	0.008	0.027	0.01	0.029
Families in *Oportunidades*, age 15	0.045	0.069	0.05	0.074
Cigarette price in 2005	14.442	1.039	14.42	1.036
Cigarette price, age 15	11.192	2.041	11.17	2.125
Tobacco control law in 2005	0.220	0.411	0.23	0.423
Observations	2181		2662	

Source: Individual level data comes from the second wave of the Mexican Family Life Survey (MxFLS-2). People aged 20 to 29 years old who started middle school between 1990 and 1999. State-level policies come from different sources: Tobacco Control Laws: State-level regulation (as explained in Chapter 1). Cigarette Prices: Central Bank of Mexico. Families in Oportunidades: Ministry of Social Development.

returns to schooling may be larger for low levels of schooling. Fortunately, middle school is also the level expected to be influenced by the features of the school expansion that I use to instrument for schooling (more on this below).

The MxFLS is ideally suited for this research because it sought to measure individuals' expectations, inter-temporal, and risk-related preferences based on questions that involve hypothetical choices in the financial domain. The internal validity of these questions has been validated through an associated project —the Mexican Family Life Survey –Preferences Pilot— which collected information from both experiments that involved real financial incentives and survey measures of preferences on a subsample of MxFLS respondents (Hamoudi and Thomas, 2006). Still, responses to such questions cannot be taken at face value, because it is not clear that the preferences elicited over money apply to the health domain of interest here. However, they may reasonably be taken as informative about these preference parameters.

To gauge their value of the future, individuals were asked whether they agreed with the following statement: "What time frame is the most important to you when deciding how much money to spend and to save?" I code those who answered "at least some months" as being future-oriented.[24] Individuals who value more the future (i.e. have low discount rates), plan more for the future. Consequently, they are expected to invest more in both schooling and health promotion.

Individuals were also asked: "How probable is it that there will be enough money in three years to cover all your household needs?" People with higher subjective probabilities on this question were coded as having 'good future expectations'.[25] To ensure realization of future utility, people with better expectations

---

[24] Other questions were also designed to elicit inter-temporal preferences. Subjects were, for example, presented with a choice between receiving a payment immediately, or waiting a specified period of time and then receiving a larger payment. Each decision involved different waiting times and implied discount rates. The results presented below are not sensitive to the choice of proxy.

[25] Other similar questions included: How probable is it that there is enough money this year to cover all of your household needs? How likely is that you will be working in 10 years? How likely is that you will be working in 10 years? Using these alternative questions to grasp people's expectations about the future does not affect the results presented in section IV.

might invest more in their education and health. Alternatively, for individuals with good future expectations and strong tastes for unhealthy behaviors, it might be optimal to engage in the behaviors they like and use future resources to ensure a healthier life (i.e. pay for medical expenses or smoking cessation products).

The final domain of preferences includes attitudes towards risk. The instrument presents respondents with different income-earning mechanisms ("lotteries") all of which carry different amounts of risk, so that by choosing a preferred lottery the respondent reveals her degree of tolerance to risk. Following Conroy (2008), risk aversion is constructed as a dummy variable indicating if the respondent chose the risk-free gamble. To guard against the possibility of early mortality, risk averse people will try to live healthier lives. If education is associated with more stable future consumption streams, they will also invest more in it.

Cognitive ability is based on Raven's Advanced Progressive Matrices Test, which involves the matching of patterns. This test is designed to assess a person's capacity for intellectual activity, irrespective of his acquired knowledge (Raven et al., 1993). For this study individual Raven test scores were normalized to obtain a continuous distribution of cognitive ability, with mean 0 and variance 1 that is not truncated at either tail. This procedure was done separately for men and women by cohort. Raven tests from the MxFLS have been used before to explore the role of mother's cognitive ability in procuring her children's health (Rubalcava and Teruel, 2004), and to show that cognitive ability dynamics interact with both individual and local indicators of macroeconomic wellbeing (Mayer-Foulkes, 2009).

Family background variables include a proxy for family resources and parental schooling. The first is a dummy indicating whether or not the household where the individual lived when he/she was 12 had a toilet (versus having a latrine, black hole or no sanitary services). Parental schooling is coded as a dummy variable taking the value of one if either parent completed at least one year of middle school. An indicator for whether the mother is the one who has at least some middle school is also included.

For the reasons explained below, each individual in the MxFLS was matched to the number of schools available in his/her state-of-birth when he/she was 12 years old. This is when individuals in Mexico are expected to start middle-school. I also matched individuals with the characteristics of their current state of residence, and their state of residence when they were 12 and 15 years old. Current state-level characteristics include cigarette prices, a dummy indicating whether the state had enacted tobacco control policies and the number of families in *Oportunidades* in relation to the total number of families in the state. This is a proxy for the intensity of the *Oportunidades* program. Characteristics at age 15 also include the intensity of this program and cigarette prices. At age 12, only the percentage of families in Opportunities is included.

State-level data come from different sources. The school construction data is from the Ministry of Education (*Sistema Nacional de Estadística Educativa*, in Spanish). State-level tobacco control laws "for the protection of non-smokers" were gathered from *Orden Jurídico Nacional* - a repository of laws available online -. Details on how the laws were coded are explained in Chapter 2 of this dissertation. Information on cigarette retail pack prices is derived from barcode scanning in large food stores reported monthly by the Central Bank of Mexico for 46 urban cities and published in the Official Diary. Nominal prices were converted to real prices using the official "Cigarette City Price Index" available online at the Central Bank's webpage.[26]

---

26 http://www.banxico.org.mx/polmoneinflacion/estadisticas/indicesPrecios/indicesPreciosConsumidor.html

State cigarette prices were constructed as simple averages of city prices. Information on the percentage of families who benefit from the *Oportunidades* program is public information made available online by the Ministry of Social Development.[27]

3. Econometric Framework and Identification of the Causal Effect of Schooling on Smoking

3.1 Econometric Framework

To facilitate discussion of the econometric issues involved, consider a simple equation describing tobacco consumption T_i for individual i (in 2005):

$$T_i = b_o + X_i b_1 + S_i b_2 + u_i \qquad (1)$$

where b_o is an intercept, X_i is a vector of observed attributes, S_i is a schooling measure and u_i is a residual. If schooling is strictly exogenous, then OLS estimation of b_2 in equation (1) can be interpreted as a consistent estimate of the effect of schooling on cigarette consumption. The 'third factor' and the 'reverse causality' hypotheses suggest, however, that schooling is endogenous. To see this, let equation (2) represent the schooling decision:

$$S_i = a_o + X_i a_1 + T_i a_2 + \varepsilon_i \qquad (2)$$

where a_o is an intercept and ε_i is a residual. Following Cawley (2004), suppose in addition that each of the residuals in equations (1) and (2) can be decomposed as having a genetic component G_i^j $(j = T, S)$, a non-genetic component NG_i^j $(j = T, S)$, and error terms ω_i and v_i that are i.i.d over individuals:

[27] http://www.oportunidades.gob.mx/

$$\varepsilon_i = G_i^S + NG_i^S + \omega_i \quad (3)$$

$$u_i = G_i^T + NG_i^T + \omega_i \quad (4)$$

A violation of the assumption that b_2 is uncorrelated with u_i and hence, that OLS estimates of b_2 will yield biased estimates of the true effect of schooling on smoking might occur in three circumstances: 1) $a_2 \neq 0$; 2) the genetic factors G_i^S affecting schooling decisions are correlated with genetic factors that exert an effect on smoking consumption G_i^T (e.g. innate ability); or, 3) the non-genetic factors influencing schooling NG_i^S are related to non-genetic factors affecting smoking NG_i^T (e.g. time preferences). That $a_2 \neq 0$ is what the reverse causality hypothesis predicts. The other two situations are considered by the 'third factor' hypothesis.

The richness of the dataset I use in this chapter enables me to control for genetic and non-genetic variables that are widely thought to influence both schooling and smoking. These variables include future orientation, expectations about the future, a measure of risk aversion, cognitive ability, and family background variables measured prior to middle school attendance. The third variable hypothesis will be tested comparing the schooling coefficients of a baseline LPM with those that add gradually each of the 'third factors'. The baseline regression will include a basic set of controls, namely, age, age squared, a dummy coded as one if the respondent is self-reported as indigenous, a dummy reflecting current school attendance, a dummy for rural-born and a set of state-of-birth fixed-effects.[28] This strategy will allow me to assess the extent of the biases and estimate more precise coefficients.

In spite of the rich set of controls that are usually unobservable, the schooling coefficients are not to be interpreted as causal effects. This is because there might still

[28] The heteroscedasticity of the error term in this linear model is corrected using White's robust estimator.

be other time-invariant characteristics driving both health behaviors and schooling. Also, there might be other unobserved characteristics influencing both schooling and smoking that vary over time. Finally, the correlation may be driven by reverse causality. These three concerns are addressed in the second part of the analysis. To present a convincing analysis of the causal link of schooling on smoking, I will use exogenous variation in classroom availability that results from a considerable school expansion that came about in Mexico during the nineties. The context of this supply-side intervention and how it lends itself to an instrumental variables approach are discussed in turn.

3.2 Middle school policy background

Mexico's objective of providing full coverage to all elementary school-age children was nearly fulfilled in the early nineties. At the same time, the *secundaria,* a schooling cycle created in 1923 that corresponds to the U.S. junior high school (grades 7-9), only served as an optional continuation of primary studies for youth aged 12 to 15 years (Levinson, 1999). In fact, it was not until 1989 when the Mexican Government set as a goal "the provision of junior high school educational services through the use of traditional and newer non-conventional options" (*Poder Ejecutivo Federal*, 1989; p.41). This objective was set in the context of an unprecedented educational modernization program that involved a number of administrative and curricular reforms. In 1992 administrative resources were decentralized from the Federal to State Governments. Also in that year, a federal initiative proposed the amendment of the Mexican Constitution to mandate compulsory secondary schooling, which was passed by Congress in 1993.

A sizable secondary school construction seems to have followed the 1992 initiative requiring completion of lower secondary school education. Figure 4.1 shows the annual growth rate of public middle schools and classrooms in Mexico between the school cycles of 1991-1992 and 2007-2008.

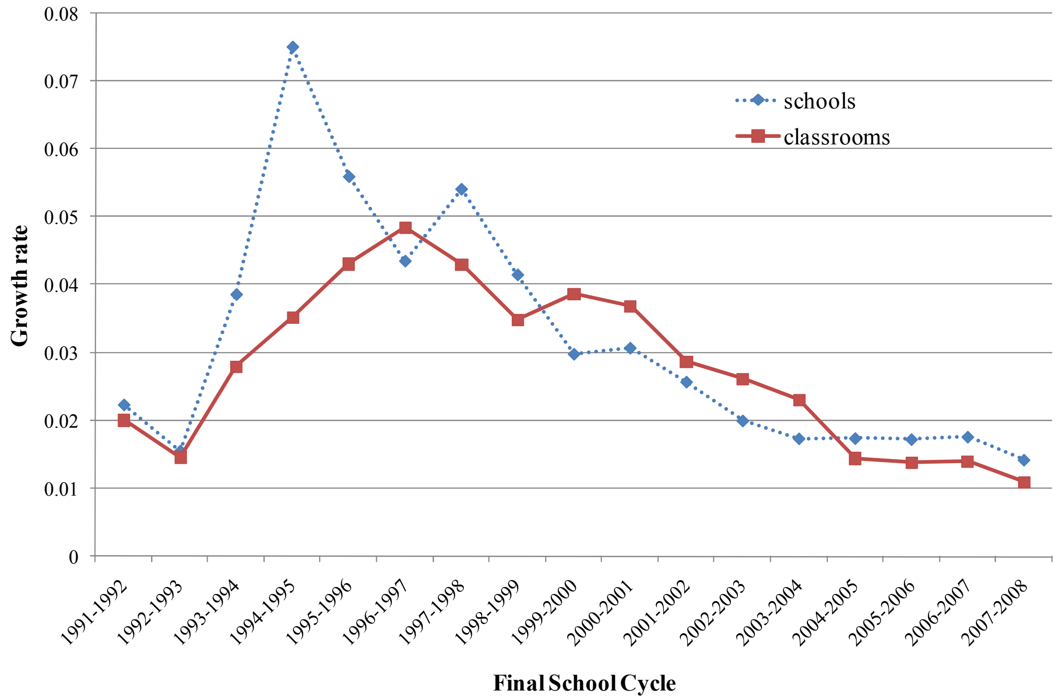
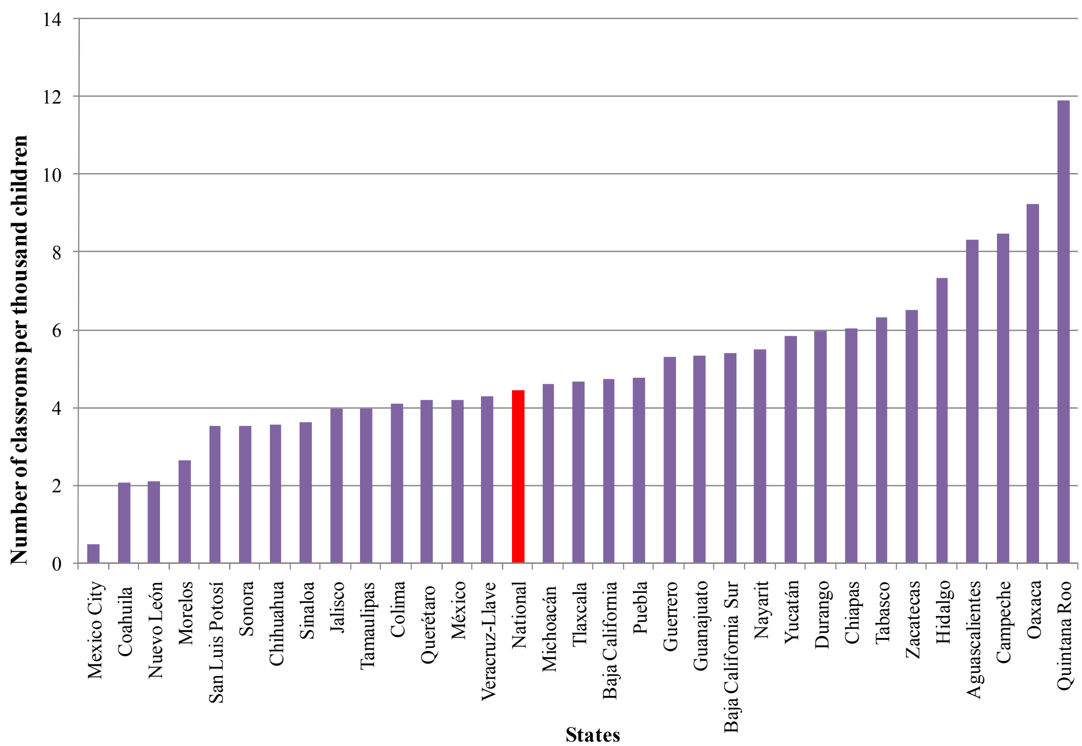


Figure 4.1 Annual Growth Rate of Public School Infrastructure in Mexico.
(Schools and Classrooms)

It can readily be seen that a tremendous rise in the number of both middle school facilities and classrooms started in 1992 and did not settle down until around 2003. In 1992, there were 17,584 public schools and 121,031 classrooms in Mexico. By 1999, the year where the youngest cohort in my sample should have started junior high school, these figures had risen to 24,450 and 157,906 respectively. This increase represents 0.82 schools and 4.4 classrooms per 1000 children aged 12 to 15 in 1992.[29] As it is clear from Figure 4.2, the expansion of classroom availability was very heterogeneous across states. For example, while the increase in the number of

[29] Based on data from the population council (*CONAPO* in Spanish) there were 8,301,905 children aged 12 to 15 in 1992.

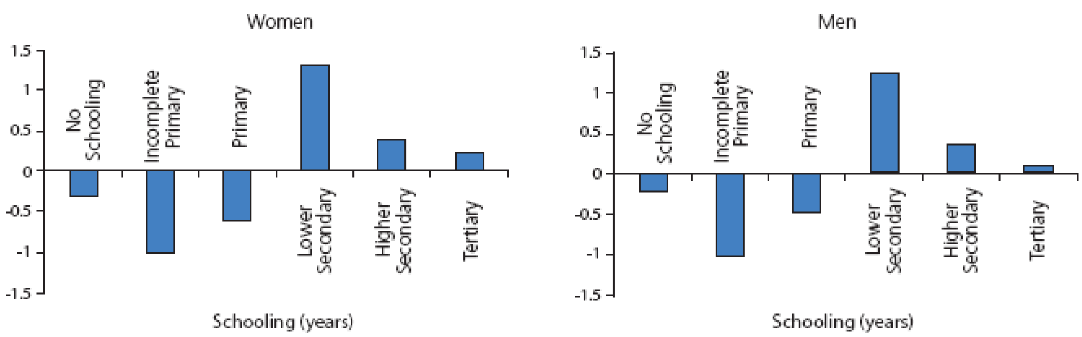
classrooms per thousand school-age children between 1992 and 1999 was of 0.4 in Mexico City, this figure was 11.91 in Quintana Roo.



Source: Own calculations based on data from the Ministry of Education (Sistema Nacional de Estadística Educativa).

Figure 4.2 Change of middle school classroom availability between 1992 and 1999.
(per thousand children aged 12 to 15 years in 1992)

Most of the school expansion took place in rural areas, where at least 17% of the students with elementary school completed had no access to secondary schools before the intervention (*Poder Ejecutivo Federal*, 1989). Many of the schools built were tele-secondary schools, which are part of a satellite system that shows videos during class time and are followed by time spent doing exercises (Parker et al., 2008). By decreasing the price of schooling, the school construction intervention was expected to relax the initial supply-sided binding constraint and increase junior high schooling levels (Kanbur, 2008). Average schooling attainment for people aged 15 and over in Mexico increased from 6.5 grades in 1990 to 8.1 in 2005. Was this caused, at least partially, by the increase of public schools? Based on descriptive evidence, Mayer-Foulkes (2008) suggests that this was indeed the case. Between 1989 and 2000, a higher proportion of men and women completed lower secondary school instead of having no schooling or incomplete or complete elementary schooling (see Figure 4.3). Mayer-Foulkes (2008) emphasizes that the changes in the distribution of schooling occurred "almost exclusively in response to the increased availability of public schooling" (p.16).



Source: Mayer-Foulkes (2006).

Figure 4.3 Profile of Changes in the Distribution of Adult School Attainment 1989-2000.

3.3 *Identification and Empirical Strategy*

Following the literature that uses educational policies as instruments for schooling attainment (see, for example, de Walque, 2003; Lleras-Muney, 2005; Parka and Kang, 2008; Albouy and Lequien, 2009), in this chapter I use plausibly exogenous variation coming from public classroom availability in the state where the individual was born at the time he/she was making his/her decisions to start middle-school. Classroom availability is measured as the number of classrooms that existed in the individual's state of birth when the person was 12, divided by the estimated number of 12-15 year olds in the state in that year (in thousands). The denominator takes into account that, for a fixed amount of classrooms, cohort size might have an impact on the actual availability of schooling for a particular individual. In contrast to previous studies which have used the availability of schools as instruments for schooling (e.g. Duflo, 2001; de Walque, 2003; Currie and Moretti, 2003; Chou et al., 2007; Parka and Kang, 2008), the numerator of my instrument considers that school availability not only depends on their numbers, but also on their size (Card and Lemieux, 2001). By using the individual's state-of-birth, as opposed to state of education, I ensure that my estimates are not contaminated by self-selection. That is, the estimated coefficients are not influenced by the fact that some individuals in my sample could have moved across geographic units to attend middle school in areas where school construction was higher (or lower). This is because all the individuals were born before the school construction took place.

The general framework for my instrumental variables regression, which I estimate via two-stage least squares (2SLS), can be written as follows:

$$S_{iby}^{2005} = a_o + X_i^{'}a_1^{'} + C_{iby}a_2 + \delta_b + \varepsilon_{iby} \qquad (5)$$

$$T_{iby}^{2005} = b_o + X_i^{'}b_1^{'} + S_{iby}\hat{b}_2 + \gamma_b + u_{iby} \qquad (6)$$

Equation (5) describes the first stage of the 2SLS estimation. Individual levels of schooling are regressed on an intercept $\mathrm{a_o}$, a vector $\mathrm{X_i^{'}}$ capturing all observed attributes other than classroom availability, the per-thousand children classrooms available for individual *i*, who was born in state *b* and turned twelve years old in year *y*, $\mathrm{C_{iby}}$, and a set of state-of-birth dummies δ_b. ε_{iby} is the error term.

The key identifying assumption is that the number of clasrrooms available in the state of birth at age 12 is not correlated with unobserved factors of the smoking participation decisions in 2005. That is, C_{iby} in equation (5) can be excluded from equation (6), the second stage. To the extent that classroom availability generates exogenous variation that affects smoking only through its effect on schooling, the predicted levels of education from the first stage will produce unbiased estimates of the effect of schooling on smoking in the second stage. The inclusion of fixed-effects in equations (5) and (6), δ_b and γ_b respectively, wipes out state-level time invariant unobserved characteristics. Identification thus derives from within-state variation across cohorts who attended middle school at different times. As such, the validity of the IV estimation rests on the assumption that the cross cohort difference in individual smoking behaviors follows a smooth-enough trend and that any departure can be attributed to the extra schooling induced by the school construction in their state of birth.

The maximum level of disaggregation of the data on the number of classrooms (or schools) available from the Ministry of Education is at the state-level. This seems an imperfect measure of within-state variation in the number of classrooms in urban areas given that the construction of public schools was intended to affect people in rural areas. Put it differently, for somebody born in an urban area, the school expansion in rural areas in the state where this individual was born should not have affected their schooling decisions, unless this person decided to emigrate to rural areas, which seems implausibly. To account for this fact, I use both the level effect and the interaction between classroom availability C_{iby} and rural-born R_i as instruments for schooling in the first stage:

$$S_{iby}^{2005} = a_o + X_i^{'}a_1^{'} + C_{iby}a_2 + C_{iby} * R_i a_3 + \delta_b + \varepsilon_{iby} \qquad (7)$$

This identification strategy relies on the extra boost to education that school construction in rural areas should have on children of rural backgrounds.

3.4 Instrument validity

The validity of the instrument rests on the assumption that the number of classrooms available when the individual was 12 years old is not correlated with smoking participation decisions observed in 2005. Understanding what motivated the allocation of classrooms is key to assessing the validity of the instrument. Table 4.4 sheds some light on the source of classroom availability comparing an array of observable characteristics for two groups of states: 'low' and 'high' intensity. States are classified as 'low intensity' if the increase in the availability of classrooms between 1992 and 1999 fell below the national average. The rest of the states form the 'high intensity' group.

As expected, the first row of Table 4.4 shows that a higher number of classrooms became available in states where the share of rural population was higher. In the same table, the comparison of national-level socio-demographic variables before the school construction took place reveals that, compared to low intensity states, in high intensity states the percentage of indigenous people and of the number of children born alive was higher, but educational attainment, individual and household income, and the percentage of the economic active population was lower. When only rural areas across states are compared, the same patterns arise; although the differences are no longer statistically significant.

The evidence presented in Table 4.4 suggests that the placement of middle schools was done in a compensatory way (i.e. in underserved areas). The inclusion of state-of-birth fixed-effects should capture pre-school construction time invariant differences. However, policy endogeneity may still be problematic if within-state

Table 4.4 Comparison of State-Level Variables Measured in 1990 by Intensity of Classroom Construction.

	Unit of Measurement	Increase in Availability of Classrooms 92-99		t-statistic	
		Low	**High**		
Rural	%	0.224	0.374	-2.93	***
		(0.036)	(0.035)		
National level:					
Indigenous households	%	0.041	0.162	-2.36	
		(0.012)	(0.044)		
Educational attainment	grade	6.918	5.914	3.29	***
		(0.221)	(0.207)		
Children born alive	number	2.724	2.971	-3.03	***
		(0.062)	(0.053)		
Household income	pesos	1113.643	876.389	2.20	**
		(58.098)	(83.400)		
Individual income	pesos	232.286	176.611	2.18	**
		(13.827)	(19.721)		
Economic active pop.	%	44.115	42.123	1.72	
		(0.500)	(0.942)		
Rural areas:					
Indigenous households	%	0.087	0.236	-1.97	
		(0.025)	(0.064)		
Educational attainment	grade	4.605	4.121	1.76	
		(0.201)	(0.185)		
Children born alive	number	3.369	3.473	-1.28	
		(0.074)	(0.044)		
Household income	pesos	600.357	519.611	1.15	
		(38.711)	(54.276)		
Individual income	pesos	120.000	100.722	1.18	
		(8.416)	(12.746)		
Economic active pop.	%	40.540	39.624	0.69	
		(0.698)	(1.043)		
Observations		14	18		

Notes: 'Low' intensity are states where the increase in the availability of classrooms between 1992 and 1999 fell below the national average. The rest of the states form the 'high intensity' group.

*Standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations based on data from the National Institute of Statistics (INEGI).

Table 4.4 (Continued)

	Unit of Measurement	Increase in Availability of Classrooms 92-99		t-statistic
		Low	High	
Urban Areas:				
Indigenous households	%	0.021	0.115	-2.28 **
		(0.005)	(0.036)	
Educational attainment	grade	7.578	6.913	3.37 ***
		(0.140)	(0.136)	
Children born alive	%	45.073	43.638	1.43
		(0.443)	(0.810)	
Household income	number	2.543	2.696	-2.50 **
		(0.043)	(0.042)	
Individual income	pesos	1247.143	1059.722	1.91
		(50.450)	(77.110)	
Economic active pop.	pesos	263.500	219.167	1.92
		(11.179)	(18.403)	
Observations		14	18	

Notes: 'Low' intensity are states where the increase in the availability of classrooms between 1992 and 1999 fell below the national average. The rest of the states form the 'high intensity' group.

*Standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations based on data from the National Institute of Statistics (INEGI).

unobservable characteristics drive both classroom availability at age 12 and smoking status in 2005. This possibility would arise if, for example, high intensity states are also the states that eventually passed more stringent tobacco control laws or exhibited higher cigarette prices. Table 4.5 shows that cigarette prices and the enactment of tobacco control laws are orthogonal to the intensity of the program. Nonetheless, I include state-level cigarette prices and policies because they are standard controls in a cigarette demand estimation equation. In fact, I also include cigarette prices at age 15. These prices are more representative of the prices the individual faced when he/she initiated smoking, and, due to the addictive nature of cigarettes, those prices might also influence current smoking decisions.

Table 4.5 State Policy Variables by Intensity of Classroom Construction.

	Year	Unit of Measurement	Increase in Availability of Classrooms 92-99		t-statistic	
			Low	**High**		
Tobacco producer state	.	dummy	0.071 (0.071)	0.111 (0.076)	-0.37	
Cigarette price	2000	pesos	10.504 (0.266)	10.207 (0.298)	0.72	
Cigarette price	2006	pesos	15.041 (0.261)	14.871 (0.279)	0.43	
Tobacco control law	2000	dummy	0.000 (0.000)	0.167 (0.090)	-1.62	
Tobacco control law	2006	dummy	0.286 (0.125)	0.500 (0.121)	-1.21	
CIAPI-men	2006	Index(0-1)	0.037 (0.017)	0.090 (0.027)	-1.54	
CIAPI - women	2006	Index (0-1)	0.018 (0.008)	0.037 (0.010)	-1.38	
Families in *Oportunidades*	1997	%	0.010 (0.006)	0.020 (0.008)	-0.98	
Families in *Oportunidades*	1998	%	0.056 (0.014)	0.100 (0.016)	-1.99	**
Families in *Oportunidades*	1999	%	0.070 (0.017)	0.151 (0.022)	-2.76	***
Families in *Oportunidades*	2000	%	0.070 (0.017)	0.159 (0.024)	-2.85	***
Families in *Oportunidades*	2001	%	0.087 (0.019)	0.192 (0.030)	-2.77	***
Families in *Oportunidades*	2002	%	0.124 (0.024)	0.257 (0.035)	-2.97	***
Families in *Oportunidades*	2003	%	0.121 (0.024)	0.252 (0.035)	-2.93	***
Families in *Oportunidades*	2004	%	0.142 (0.026)	0.286 (0.035)	-3.11	***
Families in *Oportunidades*	2005	%	0.140 (0.026)	0.277 (0.034)	-3.06	***
Observations			14	18		

Notes: 'Low intensity' states are the ones where the increase in the availability of classrooms between 1992 and 1999 fell below the national average. The rest of the states form the 'high intensity' group.

*Standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Source: Own calculations based on different sources of data. Tobacco Control Laws: State-level regulation (as explained in Chapter 1). Cigarette Prices: Central Bank of Mexico. Families in Oportunidades: Ministry of Social Development.

A final concern is that states where the number of classrooms increased the most between 1990 and 1999 were also states where the presence of other programs that potentially affected individual's smoking decisions in 2005 was increasing (or decreasing). The lower part of Table 4.5 shows that in high intensity states not only a higher percentage of total families were beneficiaries of the *Oportunidades* program, but this difference widened over time. This is not surprising as the *Oportunidades* program was targeted to rural marginalized communities in 1997 and it was expanded to urban localities in 2002. In chapter three I was unable to reject the hypothesis of no effects on smoking of the *Oportunidades.* Since there is no evidence of a causal effect of *Oportunidades* on smoking, the increasing presence of this program in states where the number of classrooms was also increasing should not cause concerns.

The *Oportunidades* program did have an effect on school attainment (see Behrman et al., 2005). Even though the objective of this paper is not to estimate the causal effect of state policies on schooling, but to use these policies to investigate the effect of schooling on smoking; the intensity of the *Oportunidades* program in the individual's state of residence at age 12 is included in most specifications. By doing that, biases in the estimation of the effect of the school construction on schooling would be reduced. For completeness, I also include a variable capturing the intensity of the program at age 15, when individuals start taking their smoking initiation decisions, and in 2005, when the outcome of interest is observed. Because state-of-birth fixed-effects are included in all specifications, the identification of the *Oportunidades* program, the state-level prices and tobacco control policy variables comes from people's movements across states over time. Additionally, as the prices and intensity of the *Oportunidades* program varied over time, the exposure of non-movers to each of these variables at age 12 and 15 varied across cohorts.[30]

Based on the previous discussion, there are no reasons to believe that classroom availability is not a valid instrument for schooling. To further support this conclusion I formally test the validity of the instrument following Curie and Moretti (2005). That is, I ask whether the impact on schooling of public middle school classrooms differs from the impact of private classrooms. If the instrument captures reductions in the price of schooling, the availability of public classrooms should have a larger impact than the availability of private classrooms. As individuals in rural areas are generally of disadvantaged backgrounds, most of them attend public schools. Hence, the availability of private classrooms in rural areas should have a marginal effect.

4. Results

4.1 Results of Linear Probability Models of Smoking Participation

Before I present the results that test the omitted "third variables" hypothesis, Table 4.6 provides the estimates from Linear Probability Models (LPM) that correspond to the specification that includes only the most basic set of control variables. Columns (1) to (3) present results for men and columns (4) to (6) report results for women. In columns (1) and (4) I only control for state-of-birth indicators, age and age squared. These schooling coefficients thus represent the total association between schooling and smoking for men and women, respectively. In line with Table 4.1, men with some middle school are five percent less likely to smoke, but women are two percent more likely to smoke. In columns (2) and (5) demographic variables are added. When

[30] Most of the tobacco control laws were enacted between 2000 and 2005, after the majority of the individuals in my sample had turned 15 years. For that reason, I only include a measure of tobacco control laws in 2005.

Table 4.6 Linear Probability Estimations of Smoking Participation: Benchmark Models.

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
Some middle school	-0.0360 **	-0.0503 **	-0.0515 **	0.0388 ***	0.0230 ***	0.0209 ***
	(0.017)	(0.020)	(0.020)	(0.010)	(0.008)	(0.008)
Currently attends school		-0.0128	-0.0142		0.0159 **	0.0214 ***
		(0.020)	(0.018)		(0.007)	(0.007)
Indigenous		-0.0351	-0.0466 **		-0.0015	-0.0021
		(0.026)	(0.023)		(0.012)	(0.010)
Rural-born		-0.0727 ***	-0.0778 ***		-0.0532 ***	-0.0504 ***
		(0.011)	(0.012)		(0.008)	(0.007)
State-level time varying controls#	no	no	yes	no	no	yes
R-squared	0.024	0.032	0.045	0.037	0.049	0.069
Observations	2181	2181	2181	2662	2662	2662

Notes: All specifications include the following controls: age, age squared, and state fixed effects.

 # State-Level time Varying Controls are the following: 1) the percentage of families in Oportunidades in the state where the individual lived in 2005, and when he/she was 12 and 15 years of age; 2) a dummy variable for whether the state where the person resided the year of the follow-up survey had enacted tobacco control laws; and 3) cigarette real prices in the state where the individual resided in 2005 and when he/she was 15 years old.

*Robust standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

controlling for indicators for rural-born, indigenous status and for being currently studying, the schooling coefficient the negative association between schooling and smoking increases among men, and the positive association decreases among women. Being indigenous or rural-born is negatively associated with smoking for both men and women. Interestingly, while both some middle school and being currently studying are positively associated with smoking among women, they are negatively related to smoking among men. Columns (3) and (6) add tobacco control laws, cigarette prices and the intensity of *Oportunidades.* The inclusion of these variables further increases the schooling coefficient for men and decreases that for women. The magnitude of this change is small and the schooling coefficients remain statistically significant. The F-test statistic at the bottom of the table shows that the set of policy and price variables is significant.

4.2 Results of the Test of the 'Third Variable' hypothesis

As discussed before, the negative correlation between schooling and smoking among men and the positive association among women may be due to omitted variables. Using information from a single source, the Mexican Family Life Survey, this part of the empirical analysis assesses the importance of usually unobserved measures as determinants of smoking decisions and evaluates its effect on the schooling coefficient. The variables are grouped in three sets: utility function parameters, cognitive ability and family background variables. To see the effects of each set alone, I gradually add variables to a baseline LPM specification that includes the same controls as columns (3) and (6) in Table 4.6. The results for men are reported in Table 4.7 and those for women in Table 4.8.

I start the discussion with the results for men. The inclusion of utility parameters in Table 7 columns (2) to (4) result in a slightly increase in the middle school coefficient. Even though none of the utility function parameters are statistically significant, the

Table 4.7 Linear Probability Estimations of Smoking Participation: Test of the 'Third Variables' Hypothesis for Men.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Some middle school (sms)	-0.0515 **	-0.0532 **	-0.0533 **	-0.0533 **	-0.0418 **	-0.0512 ***	-0.0607 ***	-0.0525 **
	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.019)	(0.021)	(0.022)
Future oriented		-0.0020	-0.0161	-0.0167				-0.0177
		(0.013)	(0.036)	(0.036)				(0.036)
Good future expectations		0.0192	0.0124	0.0125				0.0192
		(0.034)	(0.040)	(0.040)				(0.040)
Future oriented *good future			0.0246	0.0256				0.0288
			(0.054)	(0.054)				(0.053)
Risk averse				0.0062				0.0077
				(0.014)				(0.014)
Cognitive Ability					-0.0193 ***			-0.0234 ***
					(0.006)			(0.006)
Age 12 family resources age 12						-0.0020	-0.0094	-0.0082
						(0.014)	(0.015)	(0.016)
Parent with sms							0.0650 ***	0.0705 ***
							(0.014)	(0.014)
Mother has sms							0.0295	0.0328
							(0.023)	(0.023)
R-squared	0.045	0.045	0.045	0.045	0.047	0.045	0.049	0.052
F, controls				0.21	10.09		7.31	4.56
p-value of F				0.93	0.00		0.00	0.00
Observations	2181	2181	2181	2181	2181	2181	2181	2181

Notes: All specifications include: age, age squared, a dummy for school attendance, a dummy for being indigenous, a dummy for rural born, state fixed effects, state-level time varying controls and dummies that flag for MV. See table 6 for a definition of state-level time varying controls.

*Robust standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

Table 4.8 Linear Probability Estimations of Smoking Participation: Test of the 'Third Variables' Hypothesis for Women.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Some middle school (sms)	0.0209 ***	0.0187 **	0.0186 **	0.0184 **	0.0134 **	0.0188 **	0.0129 *	0.0055
	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)
Future oriented		-0.0148 **	0.0161	0.0160				0.0136
		(0.006)	(0.014)	(0.014)				(0.015)
Good future expectations		0.0389 **	0.0548 **	0.0548 **				0.0478 **
		(0.016)	(0.023)	(0.023)				(0.024)
Future oriented* good future			-0.0561 *	-0.0561 *				-0.0542 *
			(0.029)	(0.029)				(0.031)
Risk averse				0.0091				0.0081
				(0.008)				(0.009)
Cognitive Ability					0.0111 ***			0.0091 ***
					(0.004)			(0.003)
Age 12 family resources age 12						0.0118	0.0066	0.0051
						(0.009)	(0.010)	(0.011)
Parent with sms							0.0331 ***	0.0301 ***
							(0.008)	(0.007)
Mother has sms							-0.0178	-0.0196
							(0.012)	(0.014)
R-squared	0.069	0.071	0.072	0.072	0.071	0.069	0.075	0.079
F , controls				4.30	8.35		11.39	6.31
p-value of F				0.00	0.00		0.00	0.00
Observations	2662	2662	2662	2662	2662	2662	2662	2662

Notes: All specifications include: age, age squared, a dummy for school attendance, a dummy for being indigenous, a dummy for rural born, state fixed effects, state-level time varying controls and dummies that flag for MV. See table 6 for a definition of state-level time varying controls.

*Robust standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

test at the bottom of the table shows that these variables are jointly significant. I next raise the question of whether, controlling for the basic set of variables, the schooling-smoking relationship is mediated by cognitive ability as measured by the standardized Raven test. Column (5) of Table 4.7 shows that cognitive ability is negatively associated with smoking participation among men in their twenties. The apparent smoking-related returns to schooling are 18.8% smaller that those implied by column 1. The next set of regressions control for family background variables. Family resources at age 12 are negatively but not statistically associated with smoking. However, parental schooling increases the probability of smoking. The magnitude of this 'effect' is even higher than the effect of own schooling. As with own schooling, the relationship between parental schooling and smoking decisions of the offspring includes the direct effects of schooling, but also the indirect effects. The later include peer effects, occupation and assortative mating, among others. To the extent that family resources, measured as the household disposing of toilet, are not a good proxy for income, the schooling parental coefficient also reflects income effects. The last column of Table 4.7 list the coefficients of a regression where all the 'third factors' are included. In doing that, both the cognitive ability and parental schooling 'effects' increase, but the magnitude and statistical significance of the middle school coefficient remains fairly robust (compare the last column of Table 4.7 with column 1).

The results for women are very different from the results for men. Column 1 of Table 4.8 lists the coefficient of the baseline regression. The inclusion of some of the utility function parameters reveals that 10 percent of the schooling-smoking relationship is mediated by future orientation and future expectations (column (2)). As one would expect, future oriented women are less likely to smoke. However, better future expectations are positively associated with smoking. In fact, as shown in column (3), future orientation decreases smoking only when the individual has good expectations about the future. The risk aversion coefficient is small and statistically insignificant (see column (4)). The inclusion of cognitive ability alone decreases the schooling coefficient by 35 percent. In contrast with the results for men, cognitive ability is positively and strongly associated with smoking among women. The family background variables are added in columns (6) and (7). The proxy for family resources is not statistically significant. However, the total 'effect' of parental schooling increases smoking participation. The magnitude of this effect is smaller than that for men. The inclusion of family background variables decreases the schooling-smoking correlation among women by almost 40 percent, and the coefficient becomes only border-line significant. The last column of Table 4.8 includes all the usually unobserved regressors. While the contribution of these factors is similar as when each of them was included separately, controlling for all 'third factors' drives the schooling effect on smoking among women to zero, and the coefficient is no longer statistically significant.

All in all, the hypothesis that the correlation between schooling and smoking is caused by unobserved heterogeneity is borne out for Mexican females, but not for males. In particular, once usually unobserved variables are controlled for, the positive correlation between schooling and smoking among women vanishes. In contrast, the finding that males with more schooling are five percent less likely to smoke is robust to the inclusion of variables capturing the value of the future, individual's expectations, cognitive ability and family background. As discussed in the previous section, despite the stability of the schooling variable across specifications one should be reluctant to interpret the smoking gap between more and less educated men as a reliable estimate of the health returns to schooling through less smoking. In the next two sub-sections, I present the results that attempt to estimate causal effects of schooling on male smoking.

4.3 Instrumental Variable Estimations

The formal analysis of the effect of schooling on smoking starts with the discussion of the first stage results. Each cell in Table 4.9 reports the first-stage coefficient and standard error on classroom availability from separate regressions that are similar to equation (5) above. They are reduced-form estimates of schooling (see Glewwe and Kremer, 2005). In addition to coefficients and standard errors, I also report the F statistic for the test of the joint statistical significance of the excluded instruments and the partial R-squared of the excluded instruments from the first-stage regression of each IV specification. Columns (1) to (4) present the results of the effect of private schools and columns (5) to (8) report the results of the effect of public schools.

When classroom availability is used as an instrument (Panel A, Table 4.9) and only demographic variables are included (column (1)), one private classroom per 1000 children aged 12 to 15 increases the probability of having some middle school by 5 percent. However, as controls are added gradually, the effect diminishes and becomes statistically insignificant (see columns (2) to (4)). The F-statistic for the test of the hypothesis that the coefficient on the instrument in this regression of schooling is zero decreases significantly as controls are added. This suggests that the instrument is not powerful enough. Moving to the effects of public school expansion, column (5) indicates that an increase of one public classroom per 1000 children aged 12 to 15 results in a 1.1 percent increase in the probability of studying at least one year of middle school. Because individuals may have different tastes for both schooling and smoking, column (6) includes variables that capture preferences, expectations, family background and cognitive ability. When these factors are included, the classroom availability coefficient drops down. This suggest that unobserved characteristics are associated with the benefits of the public school construction. In columns (7) and (8)

Table 4.9 First Stage Estimation of the Effect of Classroom Availability on Schooling for Men.

	Private Schools				Public Schools			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Level Effect of Classroom Availability (CA)								
CA	0.050 ***	0.038 ***	-0.015	0.002	0.011 ***	0.010 ***	0.017	0.011
	(0.014)	(0.014)	(0.066)	(0.068)	(0.004)	(0.004)	(0.013)	(0.016)
R-squared	0.097	0.188	0.212	0.214	0.106	0.198	0.221	0.223
F statistic, 1st stage	12.84	7.40	0.02	0.01	8.12	7.35	1.64	0.50
Panel B: Level and Interaction Effect of Classroom Availability (CA)								
CA	0.038 **	0.033 **	-0.0319	-0.0127	0.004	0.002	-0.001	-0.004
	(0.015)	(0.015)	(0.067)	(0.070)	(0.005)	(0.005)	(0.014)	(0.017)
CA*rural born	0.028	0.014	0.0315	0.027	0.015 **	0.017 ***	0.026 ***	0.025 ***
	(0.031)	(0.030)	(0.036)	(0.036)	(0.007)	(0.007)	(0.007)	(0.007)
R-squared	0.097	0.188	0.212	0.214	0.108	0.201	0.227	0.228
F statistic, 1st stage	6.51	3.92	0.44	0.31	5.59	6.15	7.79	6.72
Demographics	yes	yes	yes	yes	yes	yes	yes	yes
Third factors	no	yes	yes	yes	no	yes	yes	yes
State fixed-effects	no	no	yes	yes	no	no	yes	yes
Time varying state controls	no	no	no	yes	no	no	no	yes
Observations	2181	2181	2181	2181	2181	2181	2181	2181

Notes: Dependent Variable is 'some middle school'. Demographics are: age, age squared, dummies for school attendance, being indigenous, and rural born. a Third factors are: dummies for a) risk averse, b) future oriented and c future expectation, and interaction of future expectations and future orientation, Raven's cognitive ability test, a dummy for family resources (having a toilet at age 12), parental middle school, a dummy for whether the mother is the parent with middle school and dummies that flag for MV. See table 6 for a definition of state-level time varying controls.

Classroom availability is the number of classrooms available in the state-of-birth of the individual when he/she was 12 years old.

*Robust standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

state-fixed effects are included. Column (7) excludes and column (8) includes state-level variables. The effect of classroom availability increases with the inclusion of state fixed-effects, but the coefficient is no longer statistically significant. As in the case with private schools, the F statistic of the preferred specification (last column) is not powerful enough. Comparing the preferred specifications of public and private classroom availability (columns (4) and (8)), as expected, the effect of public classrooms is larger than that of private ones. Nonetheless, none of the coefficients are statistically significant.

As previously discussed, the school expansion was targeted to underserved rural areas. The lack of effect of classroom availability on schooling for all the population might be due to an absent effect on individuals in urban areas. Panel B in Table 4.9 presents first stage reduced-form results of regressions where schooling is instrumented by both classroom availability and the interaction between classroom availability in the state of birth and a dummy for rural-born. As before, columns (1) to (4) list the effects of private classrooms and columns (5) to (8) report the effects of public classroom access. The level effect of classroom availability in private schools is only significant when state fixed-effects are not controlled for (see columns (1) and (2), Panel B). Having more private classrooms in rural areas is not associated with a statistically significant increase in schooling at the junior high school margin. In contrast, the reduced form estimates of the effect of public classrooms confirm that being born in a state with better access to schools increases middle school attainment for individuals who were born in rural areas (see columns (5) to (8) in Panel B). As state fixed-effects and additional state-level controls are included, the level effect of classroom availability remains insignificant, but the magnitude and significance level of the interaction term increase.

Finding a positive effect of public classroom availability in rural areas, as opposed to private classroom access, lends some credibility to the hypothesis that the opening of middle schools induced some youths who would have otherwise stopped studying after graduating from elementary school to continue their junior high education. Furthermore overidentification tests indicate that one cannot reject the validity of the exclusion restrictions for the classroom availability level and its interaction with a dummy for rural-born.

Comparing the partial R-squares in panels A and B, columns (8), the preferred specification, shows that adding one more instrument increases the explanatory power of the excluded instrument very much, explaining why the F statistics improves so much between the two specifications. As recognized by Bound et al. (1995) the F-statistic on the excluded instruments in the first stage contains valuable information about the magnitude of the finite sample bias created by weak instruments. Staiger and Stock (1997) show that, in a simple model, 1/F provides an approximate of the finite-sample relative bias of IV to OLS. Testing the joint significance of the two additional instruments yields an F of 6.72, which corresponds to a relative bias of 15%. This F is only slightly smaller than the widely cited threshold of F>10, which is associated with a 10% bias. Based on the test by Bound et al. (1995), where the extent of the bias for given F varies with the number of endogenous regressors and the number of instruments, the F-statistic of 6.72 with two excluded instruments also yields a bias of IV estimates relative to OLS of less than 20%. Therefore, while the instruments are not overwhelmingly powerful, the finite-sample biases of the estimates of the second-stage coefficients are not tremendous.

As I wish to ascertain the impact of schooling on smoking, I limit myself to the discussion of the second-stage results that are interpretable based on the F-statistic for the joint significance of the instrument in the first stage. The results that are valid and (border-line) powerful include the level effect of the public classroom availability and its interaction with a rural-born dummy. As a benchmark for comparison with the IV results, column (1) in Table 4.10 reports the schooling coefficient of a LPM that includes time preferences, state-fixed effects and state-level control variables. As before, columns (2) and (5) in Table 4.10 add controls gradually to the specifications. In all the IV specifications the direction of the effect is reversed, but none of the coefficients are statistically different from zero. The standard errors of the IV approach are relatively large, and one cannot reject the hypothesis that differences between LPM and IV estimates are due to sampling error.

Table 4.10 Second-Stage: Instrumental Variables Estimates of the Effect of Public Schooling on Smoking Participation for Men.

	(1)	(2)	(3)	(4)
Some middle school	0.1382	0.1107	0.1292	0.0539
	(0.248)	(0.244)	(0.252)	(0.267)
Demographics	yes	yes	yes	yes
Third factors	no	yes	yes	yes
State fixed-effects	no	no	yes	yes
Time varying state controls	no	no	no	yes
F Statistic on First Stage	5.59	6.15	7.79	6.72
Observations	2181	2181	2181	2181

Notes: Dependent Variable is 'some middle school'. Demographics are: age, age squared, dummies for school attendance, being indigenous, and rural born. a Third factors are: dummies for a) risk averse, b) future oriented and c future expectation, and interaction of future expectations and future orientation, Raven's cognitive ability test, a dummy for family resources (having a toilet at age 12), parental middle school, a dummy for whether the mother is the parent with middle school and dummies that flag for MV. See table 6 for a definition of state-level time varying controls.

Some Middle School is the predicted value of Schooling in the First Stage. It was obtained from a regression where Schooling was regressed on the level effect of classroom availability, an interaction term of classroom availability and a dummy for rural-born.and all the control variables included in these regressions.

*Robust standard errors are reported in parentheses. * 10% Significance level. ** 5% Significance level. *** 1% Significance level.*

5. Discussion

The evidence provided in this chapter uncovered important features of cigarette smoking in Mexico. A result of striking consistency is that males and females smoke for very different reasons. While unobserved heterogeneity explains the positive association between schooling and smoking among women, the negative correlation among men is not influenced by the addition of these difficult-to-observe control variables.

Smoking seems to occur among sophisticated women of relatively better backgrounds. That is, female smoking participation increases with the price of cigarettes, better future expectations, cognitive ability and better family background. Many of the statistically significant determinants of smoking among women have effects opposite of what one would expect. This 'lack of consistency' is actually consistent with the fact that smoking prevalence among has been increasing over time. Even in the last decade when taxes were raised, health-warning labels and anti-smoking mass media campaigns were launched, and youth access restrictions to cigarettes and clean indoor-air laws were strengthened (Ibáñez-Hernández, 2005), female smoking kept on increasing. While it is difficult to be conclusive on this point, smoking may be part of feminine modernity.

The rest of the discussion focuses on men. To the extent that the measures used in this paper are good proxies of individual's preferences in the health domain, the evidence I found is not consistent with the hypothesis that time preferences are important determinants of smoking.[31] However, in line with previous research by Kenkel et al. (2006), cognitive ability was not only associated with less smoking, but the inclusion of this variable also affected the schooling coefficient. Therefore,

---

[31] Results available from the author upon request reveal, however, that the utility parameters are important determinants of schooling.

excluding cognitive ability in models of smoking yields biased estimates of the effect of schooling on smoking.

The potential influence of family background variables on schooling and health behaviors has been widely recognized (see, for example, Case et al., 2005; Fletcher and Frisvold, 2009). Nonetheless, due to data limitations, most of the studies of smoking participation, initiation or cessation have failed to control for this type of variables. In this study, parental schooling was found to be statistically and strongly associated with smoking among men and women. The magnitude of the effect was twice as large among men. In contrast, research in the U.S. has found a small and generally not statistically significant effect of parental schooling on smoking participation (Kenkel, et al., 2006) or on preventive care behaviors (Fletcher and Frisvold, 2009). An effect of parental schooling on obesity does exist in the U.S. but it is negative (Kenkel et al., 2006).

Following the study by Berger and Leigh (1989) a number of studies attempting to estimate the causal effect of schooling on health and health behaviors used parental schooling as an instrument for schooling. In a country like Mexico where the intergenerational transmission of human capital is strong (Mayer-Foulkes, 2008) parental schooling would probably be a very powerful instrument for schooling in a model of smoking. Nevertheless, the results in this chapter show that this strategy would not be valid. Family background variables, particularly parental schooling, have direct impacts on both male and female smoking, implying that specifications that exclude these variables would be faulty. This might be also true in the analysis of other health behaviors.

Regarding causal effects, the first stage of the instrumental variables strategy indicated that the availability of one classroom per thousand school-aged individuals in rural areas was associated with a 2.6 percent increase in the probability of studying junior high school. This is the effect on relatively disadvantaged men who were constrained by lack of middle school availability and who would have not gone to middle school if it had not been for the fact that a (sufficiently large) middle-school opened in their state of birth. It is also the effect on men who took the opportunity to study as soon as it arrived. Compared to other men who had the same opportunity, they might have higher tastes for schooling relative to work.

In spite of the statistically significant effect of the school expansion on schooling and of the relatively powerful instrument, the second-stage IV estimates of the returns to schooling on smoking are no statistically different from zero. The IV parameters are less precisely estimated than the LPM ones. Based on Imbens and Angrist (1994) and under the assumption of heterogeneity on the returns to schooling this 'lack of impact' is the so-called local average treatment effect (LATE). It only applies to the individual whose education has been affected by junior high school openings (described above) and it may differ from the average effect for the population.

So, while this unusual intervention was effective in increasing schooling levels it is not clear whether it generated health-related returns to this education. At least two reasons might be behind this finding. The school construction was intended to increase the quantity of schooling, but the quality of schooling might have been deficient. Alternatively, the quality of education might have been acceptable, but the fact that individuals were 'obliged' to stay in school longer than those who voluntarily chose to go to school might have had detrimental effects on the acquired knowledge and skills. Deficient health and malnourishment should not have affected these results. These factors do affect student achievement (Maluccio et al., 2009), but they do it through their effects on cognitive development, which was controlled for in the regressions.

The lack of precision of the IV estimates and the robustness of the schooling coefficients to the inclusion of unobserved variables suggests that more education may exert a negative effect on male smoking, but my instrument simply does not create enough variation to pick it up. Moreover, Mincerian estimates on labor income in Mexico show that there are increasing returns to education at levels not achieved by most of the population (Meyer-Foulkes, 2006). Hence, years of schooling at higher educational levels might also have a causal impact on smoking and other health behaviors in Mexico.

6. Conclusions

For the most part, the research examining the relationship between schooling and health behaviors has focused on developed countries. In this chapter, I extended this literature using a rich dataset of individuals in Mexico. In trying to disentangle the correlation between schooling and smoking two hypotheses of the link between schooling and smoking participation were tested; the 'third factor' hypothesis and the hypothesis of a causal effect of schooling on smoking. The analysis was conducted separately for men and women providing very interesting new insights to the literature examining schooling and health behaviors in developing countries. I focused on middle school attendance (grades 7 to 9) as the margin on which schooling was expected to affect smoking. Descriptive evidence suggested that this was the right margin to look at schooling effects. It is also the school level of most of the people in Mexico.

The evidence presented in this chapter suggests that unobserved heterogeneity explains the positive association between schooling and smoking among women. Women from better-off backgrounds are more likely to engage in smoking. The negative correlation among men was found to be very robust to the inclusion of usually unobserved controls such as preferences for the future, cognitive ability and family background. For both men and women, schooling of the parents is positively associated with smoking.

The public school construction in Mexico during the nineties is associated with increases in junior high schooling among men. The estimation of the impact of this increased schooling on smoking among the individuals who were affected by the school expansion was less precisely estimated. Finding equally valid but more powerful instruments in future research would shed some light on whether part of the negative correlation between male schooling and smoking is causal. The results in this paper indicate that family background variables might be powerful, but not valid. Future research may benefit from exploring the effect of other schooling margins on smoking outcomes that capture the dynamics of smoking (i.e. smoking initiation or cessation). Finally, uncovering the mechanism(s) through which educational attainment affects smoking remains an important research question, and one that future research should address.

REFERENCES

Albouya, Valerie and Laurent Lequienb. 2009. "Does compulsory education lower mortality?" *Journal of Health Economics*, 28 1, pp. 155–68.

Arendt, Jacob N. 2005. "Does education cause better health? A panel data analysis using school reforms for identification." *Economics of Education Review*, 24:2, pp. 149-60.

Behrman, Jere R., Piyali Sengupta, and Petra Todd. 2005. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change*, 54:1, pp. 237-75.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2007. "The Impacts of Obligatory Secondary School Education in Mexico." *Mimeo*. Centro de Investigación y Docencia Económicas Mexico City.

Berger, Mark C. and J. Paul Leigh. 1989. "Schooling, self-selection, and health." *Journal of Human Resources*, 24:3, pp. 433-55.

Bound, John, David Jaeger, and Regina Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90:430, pp. 443-50.

Card, David and Thomas Lemieux. 2001. "Can falling supply explain the rising return to collee for younger men? ." *Quarterly Journal of Economics*, 116:2, pp. 705-46.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics*, 24:2, pp. 365-89.

Cawley, John. 2004. "The Impact of Obesity on Wages." *The Journal of Human Resources*, 39:2, pp. 451-74.

Chou, Shin-Yi, Jin-Tan Liu, Michael Grossman, and Theodore J. Joyce. 2007. "Parental Education and Child Health: Evidence from a Natural Experiment in Taiwan " *NBER Working Paper No. 13466*: Cambridge, MA.

Conroy, Hector V. 2007. "The Role of Attitudes Toward Risk in the Decision to Migrate." *California Center for Population Research*: Los Angeles.

Cutler, David M. and Adriana Lleras-Muney. 2006. "Education and Health: Evaluating Theories and Evidence." *NBER Working Paper No. 12352*: Cambridge, MA.

Cutler, David M. and Adriana Lleras-Muney. 2010. "Understanding differences in health behaviors by education." *Journal of Health Economics*, In Press.

Currie, Janet and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics*, 118:4, pp. 1495-532.

de Walque, Damien. 2003. "HowDoes Education Affect Health Decisions? The Cases of Smoking and HIV/AIDS." Ph.D. Dissertation. University of Chicago: Chicago.

de Walque, Damien. 2004. "Education, information, and smoking decisions, evidence from smoking histories 1940–2000." *World Bank Policy Research Working Paper No. 3362*: Washington, DC.

de Walque, Damien. 2006. "Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education." *Journal of Health Economics*, 26:5, pp. 877-95.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91:4, pp. 795-813.

Farrell, Phillip and Victor R. Fuchs. 1982. "Schooling and Health: The Cigarette Connection." *Journal of Health Economics*, 1:3, pp. 217-30.

Fletcher, Jason M. and David E. Frisvold. 2009. "Higher Education and Health Investments: Does More Schooling Affect Preventive Health Care Use?" *Journal of Human Capital*, 3:2, pp. 144-76.

Franque Grimard, Daniel Parent. 2007. "Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking?" *Journal of Health Economics*, 26:5, pp. 896–926.

Fuchs, Victor R. 1982. "Time preference and health: An exploratory study," in *Economic Aspects of Health*. Victor R. Fuchs ed. Chicago: University of Chicago Press, pp. 93–120.

Fuchs, Victor R. 2004. "Reflections on the socio-economic correlates of health." *Journal of Health Economics*, 23:4, pp. 653-61.

Glewwe, Paul and Michael Kremer. 2005. "Schools, Teachers, and Education Outcomes in Developing Countries." *Center for International Development Working Paper No. 122*. Harvard University: Cambridge, MA.

Grabner, Michael J. 2008. "The Causal Effect of Education on Obesity: Evidence from Compulsory Schooling Laws." *Mimeo*. University of California, Davis: Davis, CA.

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy*, 80:2, pp. 223-55.

Grossman, Michael. 2004. "The demand for health, 30 years later: a very personal retrospective and prospective reflection." *Journal of Health Economics*, 23:4, pp. 629-36.

Hamoudi, Amar and Duncan Thomas. 2006. "Do You Care? Altruism and Inter-Generational Exchanges in Mexico." *California Center for Population Research Working Paper No. 008-06.* University of California Los Angeles.

Ibáñez Hernández, Norma A. 2005. "Disposiciones jurídicas federales sobre la prohibición de fumar," in *Primer informe sobre el combate al tabaquismo. México ante el Convenio Marco para el Control del Tabaco, México*. Raydel Valdés-Salgado, Eduardo C. Lazcano-Ponce and Mauricio Hernández-Avila eds. Cuernavaca: Instituto Nacional de Salud Pública.

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62:2, pp. 467-75.

Kanbur, Ravi. 2008. "Intergenerationalities: Some Educational Questions on Quality, Quantity and Opportunity." *Mimeo*. Cornell University: Ithaca, NY.

Kenkel, Donald, Dean Lillard, and Alan Mathios. 2006. "The Roles of High School Completion and GED Receipt in Smoking and Obesity." *Journal of Labor Economics*, 24:3, pp. 635-60.

Levinson, Bradley A. 1999. ""Una etapa siempre difícil": Concepts of Adolescence and Secondary Education in Mexico." *Comparative Education Review*, 43:2, pp. 129-61.

Lleras-Muney, Adriana. 2005. "The Relationship between Education and Adult Mortality in the United States." *The Review of Economic Studies*, 72:1, pp. 189-221.

Maluccio, John A., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing, and Aryeh D. Stein. 2009. "The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults." *Economic Journal*, 119:537, pp. 734-63.

Mayer-Foulkes, David. 2008. "The Human Development Trap in Mexico." *World Development*, 36:5, pp. 775-96

Mayer-Foulkes, David. 2009. "The Cognitive Transition in Mexico: Economic Geography and Local Governance Matrix." *Mimeo*. Centro de Investigación y Docencia Económicas Mexico City.

MacInnis, B., 2008. Does college education impact health? Evidence from the pre-lottery Vietnam draft. Under review at Quarterly Journal of Economics.

Ministry of Education. 2006. "Fue Aprobado el Ingreso Preescolar y Primaria con Tres y Seis Años Cumplidos al 31 de Diciembre " *NotiSep*.

Parka, Cheolsung and Changhui Kang. 2008. "Does education induce healthy lifestyle?" *Journal of Health Economics*, 27 pp. 1516–31.

Parker, Susan W., Jere R. Behrman, and Luis Rubalcava. 2008. "The Quality of Education in Latin America and the Caribbean Region: The Mexican Case." *Mimeo*.

Raven, John C., John H. Court, and Raven J. 1993. *Test de Matrices Progresivas. Escalas Coloreada, General y Avanzada*. Barcelona: Ediciones Paidós Ibérica.

Royston, Patrick 2005. "Multiple imputation of missing values: Update of ice." *Stata Journal*, 5:4, pp. 527-36.

Rubalcava, Luis N. and Graciela M. Teruel. 2004. "The Role of Maternal Cognitive Ability on Child Health." *Economics and Human Biology*, 2:3, pp. 439-55.

Rubalcava, Luis N. and Graciela M. Teruel. 2006. "User's Guide for the Mexican Family Life Survey First Wave". Available online at: http://www.ennvih-mxfls.org/en/mxfls.php?subseccion=ver&session=30495159934#

Spasojevic, J., 2003. "Effects of education on adult health in Sweden: results from a natural experiment." Ph.D. Dissertation. City University of New York Graduate Center: New York.

Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65:3, pp. 557-86.

CHAPTER 5

CONCLUSIONS

Using individual-level data, which is a step toward the correct way to analyze smoking behaviors; this dissertation empirically assessed the extent to which different public policies, through changes in incentives, influenced the smoking decisions of youth and young adults.

The first chapter described the model to be used as the framework of analysis, and provided background on the demand and supply of cigarettes, against the backdrop of anti-smoking policies in Mexico.

One goal of the first substantive essay, which corresponds to Chapter 2, was to develop an index of clean indoor air policies based on economic theory predictions. This index, which I named Clean Indoor Air Policy Index (CIAPI), gives less credit to regulations where a small share of the population is potentially affected or where people spend nearly no time. This is because smoking restrictions in these locations, even if enforced, are not expected to increase the costs of smoking, and thereby decrease smoking consumption. The CIAPI also penalizes differences between the actual and the desired level of enforcement. This index is not without limitations, but it does provide a ‘summary score’ that, compared to previous indices, is based on more credible assumptions.

The second chapter also asked whether stringent limits to smoking are associated with the decision to smoke and, conditional upon smoking, with the intensity of consumption. More stringent laws, measured by the CIAPI, were not found to be correlated with smoking participation for either men or women. Nonetheless, stringent smoking restrictions were shown to be negatively associated with the number of cigarettes smoked by men. Despite that I found no evidence indicating that more stringent laws were caused rather than cause smoking, I prefer to not interpret these results as causal. This is because, with the data available, I was unable to rule out the possibility that the correlation is driven by unobserved state-level heterogeneity.

The first chapter also provided a general sense of the determinants of smoking using, for the first time in Mexico, individual-level data. The stock of cigarettes was found to be a good predictor of the number of cigarettes smoked. Nonetheless, none of the other individual and state-level variables were found to determine the intensity of smoking. In contrast, smoking participation was predicted by a number of individual-level variables, which varied by gender. Interestingly, the decision to smoke was negatively associated with *Oportunidades,* a federal anti-poverty program in Mexico, and with schooling among men. In the case of women, however, the *Oportunidades* program was not found to be correlated with smoking participation. Additionally, the schooling-smoking gradient was found to be positive. These two contrasting findings were explored further in the other two essays of the dissertation.

The third chapter, and second substantive essay of the dissertation, investigated the causal effect of *Oportunidades* on the smoking behaviors of its participants. The benefits of this program include sizable cash transfers, health information sessions and schooling. Affecting smoking is not a goal of this program. However, health economics research suggests that the *Oportunidades* intervention could substantially change smoking among poor Mexicans. Exploiting an exogenous jump in program participation by means of a fuzzy Regression Discontinuity Design, the estimated local average treatment effect on smoking among adults that participated in the program an average of four years was zero. In contrast, *Oportunidades* might have increased slightly the smoking rates of participant adolescents. Finally, differential treatments by sex enabled isolating the income effect of *Oportunidades* by estimating the program's impact on adult male smoking. The evidence suggested a null income effect on adult smoking.

Chapter four measured and disentangled the correlation between schooling and smoking in Mexico. Using rich data from the second wave of the Mexican Family Life Survey (MxFLS-2), I found that the positive correlation between schooling and smoking among women is driven by unobserved heterogeneity. Smoking is more prevalent among more sophisticated women coming from better-off backgrounds. In contrast, neither of the usually unobserved ‘third factors’ was shown to affect the negative association between schooling and smoking among men. The apparent health-related returns to schooling among men were tested using and Instrumental Variables (IV) framework. Within each state, I exploited variations across cohorts in new junior high school classroom openings between 1990 and 1999 to construct an instrument for schooling. The evidence in that chapter suggested that junior high school expansion was associated with higher schooling achievement, particularly in rural areas. The results also indicated that returns to schooling might include less smoking among Mexican men.

This dissertation constitutes the first thorough analysis of cigarette consumption in Mexico from an economic perspective. Using a static framework for the analysis, most of this research focused on understanding the decision to smoke. Suggestions for future research were discussed in each chapter. General extensions to this research include the exploration of the determinants of conditional intensity; and the examination of the dynamics of cigarette consumption (i.e. smoking initiation and cessation).