Killing by lung cancer or by diabetes? The trade-off between smoking and obesity

Federico A. Todeschini, Sergi Jimenez and José María Labeaga*

This draft: January 2009

Abstract

At more or less the same time that the anti-smoking campaign is regarded as one of the "major public health victories" in the US, health practitioners interest shifted towards the prevalence of obesity, becoming one of the major health challenges in decades to come. In this paper we study the impact that the final product of the anti-smoking campaign, that is, people quitting the habit of smoking, had on average weight in the population. To these ends, we use data from the Behavioral Risk Factors Surveillance System, a large series of independent representative cross-sectional surveys. We construct a synthetic panel that allows us to control for unobserved heterogeneity and we exploit the exogenous changes in taxes and regulations to instrument the endogenous decision to give up the habit of smoking. Our estimates, are very close to estimates issued in the '90s by the US Department of Health, and indicate that a 10% decrease in the incidence of smoking leads to an average weight increase of 2.2 to 3 pounds, depending on choice of specification. In addition, we find evidence that the effect overshoots in the short run, although a significant part remains even after two years. However, when we split the sample between men and women, we only find a significant effect for men. Finally, the implicit elasticity of quitting smoking to the probability of becoming obese is calculated at 0.58. This implies that the net benefit from reducing the incidence of smoking by 1% is positive even though the cost to society is $0.6 billions.

*Federico A. Todeschini is affiliated to UPF and FUNIF, Sergi Jiménez to UPF and FEDEA and José María Labeaga to UNED and FEDEA. We would like to thank the participants of the FEDEA seminar, UPF Applied Microeconomics Seminar, Inter-American Development Bank, the UPF Health Economics Seminar, Simposio de Análisis Económico España and Vrije Universiteit Amsterdam for helpful comments. Special thanks to Eduardo Cavallo, Juan Manuel Puerta, Bus van der Klaauw, Shikeb Farooqui and Antonio Cabrales for their comments. The usual caveats applies. Corresponding author: todeschini.federico@upf.edu

1 Introduction

In the last 40 years the percentage of US adults who smoke regularly dropped from above 42% in 1965 to below 20% in 2007, according to the Center for Disease Control and Prevention. This drop has been regarded as one of the most important "health victories", Gruber and Frakes (2006). The logic behind these words is that even today cigarette smoking is calculated to kill 438,000 people per year. On top of that, smokers are up to 40% more expensive for the health care system than non-smokers. In between 2000-2004 cigarette smoking was estimated to be responsible for $193 billions in annual health-related losses (Armour et al. (2005), Barendregt et al. (1997), Miller and Rise (1998) and Adhikari et al. (2008)). These are only some of the direct effects of smoking. Indirect effects range from lower labor productivity to 49,000 deaths per year due to secondhand smoking. The social cost of smoking, calculated in $11 per pack, almost doubles its private cost (CDC (2006)).

As the battle against smoking started to show very positive outcomes, health practitioners began to notice a new problem: the negative correlation between smoking rates and the prevalence of obesity. As we can see in Table 1, in 1985 the average American man was 1.78 mt. tall and weighed 80 Kg. 22 years later he weighs almost 10 Kg. more, representing an increase of 12%, even though he is as tall as before. The picture is even worse for women. During the same period they faced a similar average absolute weight gain and consequently a larger relative growth rate[1]. This trend is in contrast with historical evidence from the past 150 years where weight increases were not as abrupt and pronounced and were accompanied by increases in height, (Costa and Steckel (1997)).

Table 1: United States Average Weight and Height: 1985 versus 2007

	Men		Women	
	1985	2007	1985	2007
Height (in mt.)	1.78	1.78	1.63	1.63
	(0.076)	(0.075)	(0.067)	(0.070)
Weight (in Kg.)	80.06	89.52	63.30	73.31
	(13.09)	(18.11)	(12.35)	(17.69)
BMI*	25.25	28.25	23.82	27.58

Source: Behavioral Risk Surveillance System. Standard Deviation between parenthesis.
*Body Mass Index equals weight in kilograms divided by height in squared meters

The increase of the intensive margin went hand in hand with an increase of the extensive margin. Indeed, during the past two decades obesity rates have jumped dramatically among the US population, becoming one of the biggest health concern for policy makers. Before 1980 only 14% of its population was obese, yet nowadays 38% of the men and 34% of the women classify as obese (that is approximately 65 million people). In fact, nowadays American are more likely to be overweight than to pay federal income tax[2]. But not only the number of obese have increased. An increasing proportion of the obese population now belongs to the classes of obesity deemed more troublesome[3].

The negative correlation between the conditional number of cigarettes smoked and average body mass index (Figure 1a and Figure 1b) is a temptation to conclude that net calorie intake has substituted smoking as a habit and consequently that the decrease in the incidence of smoking is responsible for the increase in weight. In addition, because smoking affected a large share of the adult population, it is a natural suspect to analyze. The growing concern about

[1] According to the NHS data, in the UK for the period that goes from 1993 to 2007 men increased their height by 1 cm and their weight by 4.6 Kg. while women stayed the same height and gained 3kg.

[2] The Economist, Jan 21 2010.

[3] Childhood obesity has also been rising. Its prevalence has nearly tripled. Almost 19% of children aged 6 to 11 years and 17.4% of adolescents aged 12 to 19 are now obese, according to the National Center for Health Statistics. America is not only getting fatter, it is doing it at a younger age.

obesity on the one hand and the impulse the anti-smoking campaign has all over the world on the other, make it critical to examine whether the two processes are causally connected. If quitting smoking has a positive effect on weight, that is, smoking is a substitute for eating, some costs related to the anti-smoking campaign were not fully internalized in the law making process. On the contrary, if the effect happens to be negative, then there would be benefits that were probably not taken into consideration. This would put additional pressure over the states that have not passed a tough legislation on smoking. Unfortunately, economic theory remains vague in providing testable predictions about how individuals react to the elimination of a habit and this problem is further compounded by the fact that there is hardly any conclusive empirical evidence on this issue.

In addition, there are many economic and biological reasons why quitting smoking and increasing weight might be correlated in an idiosyncratic and non-causal way. First, common omitted factors such as, risk aversion, preferences or variation across individuals in the Basal Metabolic rate provides one motivation for suspecting the presence of individual-specific effects. For instance, if people that quit the habit of smoking are potentially more concerned about health (McCaul et al. (2006) and Clark and Etile (2006)) then they should be less prone to weight gains than continuers. In contrast, if it is true that quitting smoking leads to weight increase, then quitters are less concerned by the risks derived from the increase, than the other two groups [WHICH 2 GROUPS?]. In such a scenario, lack of a priori knowledge about the individual specific directional bias can easily generate non-causal correlations. Second, reverse causality posses a similar problem to the analysis. Overweight individuals may use smoking as a weight control method. As a consequence even after controlling for unobserved heterogeneity the error term will still be correlated to the decision of quitting smoking. Similarly, both processes, that is smoking and weight changes, might just be the consequence of a third common factor, for instance the stress due to a harsher labor market. Finally, weight adjustment does not happen instantaneously. On the contrary, weight adjustment costs create an autoregressive process where present weight depends on past weight realizations. Thus, failing to incorporate lagged BMI in the estimation might cause a bias in the estimation.

Summing up, the observed correlation among the two process could be completely spurious. In order to measure the causal effect of quitting smoking on weight then, it is necessary to control for unobservables and the sources of exogenous changes in the individuals decision for quitting smoking.

The economic literature that analyzes the effect of smoking on BMI is relatively recent and so far the results remain inconclusive. Chou et al. (2004), Baum (2009) and Rashad (2006) find that individuals that stop smoking increase their weight. In contrast, Gruber and Frakes (2006) and Courtemanche (2009) arrive to the opposite conclusion. All of these studies concentrate on the effect of increases in the price of cigarettes, whether it be the final price or the excise tax on tobacco. Instead of concentrating on the reduced form regression, we use the exogenous changes in cigarette prices to focus on the impact of quitting smoking on weight. Our decision to to evaluate the final product of the anti-smoking campaign, that is, the decision to quit smoking instead of focusing on just one dimension, for instance the pecuniary cost of smoking, is substantiated by the following observation. The anti-smoking campaign has many highly correlated dimensions, from information on the consequences of smoking and limitations on the advertisement of cigarettes in TV to smoking prohibitions in public places[4].

We contribute further to the literature as well as to the debate by applying cohort data techniques to the cross sectional data in order to construct a synthetic panel. This allows us to control for unobservables, and at the same time take into account the dynamic nature of the problem by incorporating the lags of BMI. We instrument the decision to give up smoking using lags of the excise taxes on tobacco, regulations regarding tobacco use in closed spaces and family characteristics. We find these instruments compelling since, conditional on a set of controls, it is difficult to argue that policy makers decided tobacco taxes and regulated its use with the purpose of controlling voters' weight

[4]For instance, nowadays it is much harder to see a person smoking in a Hollywood movie.

and consequently conditional on certain characteristics of the population we have to control for, they are exogenous. Finally, in order to analyze how quitting smoking affects the probability of becoming obese, we propose and estimate a logistic model for obesity prevalence. The logistic model applied to cell data can be log-linearized, so standard panel and IV methods can be directly applied to the data without loosing the properties of the logistic formulation.

According to our results a 10% decrease in the incidence of smoking leads to an average weight increase of 2.2 pounds to 3 pounds for the average cohort, that is, a 2% weight increase assuming constant height. We also find that the effect overshoots in the short run. However, a significant part of it remains even after two years. We also find that quitting smoking affects the extensive margin as well, with an implied elasticity of quitting smoking to obesity of 0.58. According to the CDC, an obese individual costs $1,400 more to the health system than a healthy person and a smoker costs 3,200$ more than a non-smoker. Taking this into consideration implies that, on average, a 1% decrease in the incidence of smoking has a a net gain of $1.4 billions: the cost of $0.6 billions is offset by the gross benefit of $2 billions.

The paper is organized as follows. In Section 2 we provide a list of alternative explanations for the increase in obesity rates and a review of the economics literature linking smoking to weight increase. In Section 3 we discuss the proposed methodology and analyze the data that we will use in the empirical analysis. In Section 4 we estimate the static and dynamic models of the effect of quitting smoking on Body Mass Index using the constructed pseudo-panel. In Section 5 we present an alternative to study the impact of quitting smoking on the probability of becoming obese. Section 6 concludes the paper.

2 Key Facts, Alternative Explanations and Literature Review

While smoking is the leading cause of death in the U.S., with up to 435,000 adult deaths each year, excess body weight is the third most important risk factor contributing to the burden of disease, most notably type II diabetes, hypertension, cardiovascular disease and disability (WHO, 2006). Flegal et al. (2005) calculate that in 2000, obesity caused 112,000 excess deaths in the US, while Mokdad et al. (2004) estimate 365,000 deaths due to obesity in that same year. Life expectancy of a 40 year old obese male is 6 years shorter than his non obese counterpart and for females the figure jumps to 7 years while for younger adults the effect is even higher (General Surgeon's 2001 report). Moreover, American life expectancy is projected to decrease due to obesity, for the first time since Civil War (Olshansky et al. (2005)).

As it happens with smoking, obesity carries with it several negative externalities and therefore the social cost of being obese is higher than the individual's. One of those externalities is the increase in health care utilization. Finkelstein et al. (2004) found that in 2003 weight problems represented a medical expenditure of $75 billions in the U.S.. The Urban Institute updated this figure to 200 billions for 2008, half of which comes from Medicare and Medicaid. Andreyeva et al. (2004) find that an obese person generates an average of $700 more in health expenditures than a comparable non obese, a figure that is even larger than the increase in health costs due to smoking. Nowadays obesity accounts for 9.1% of all medical spending in the United States, up from 6.5 % in 1998, an average of $1,400 more a year, although these costs are not distributed uniformly among the obese: as the degree of obesity worsens, the associated burden increases almost exponentially. Unfortunately the categories that account for the larger part of the burden are the ones rising at the highest rate (Andreyeva et al. (2004)). Labor productivity is another cost that is shared with the non obese. The U.S. Health and Human Services secretary estimated that obesity related problems costs $13 billions to U.S. businesses[5] and another study finds that on average, of every 100 workers, obese ones had

[5]The total cost is the result of health insurance costs related to obesity ($8 billion), paid sick leave ($2.4 billion), life insurance ($1.8 billion), and disability insurance ($1 billion).

lost 190 days per year, while normal weighted's 14 (Stevens (2004))

2.1 Alternative explanations

As we have seen above, the policy maker's concern about obesity is not unjustified[6]. But while most of obesity costs have been documented, we still lack a broad consensus about what caused the contemporary increase in weight and obesity rates. People put on weight when they consume more calories than they are burn off. Therefore, there are only three channels that can explain the mentioned increase. The first channel is that society, on average, started to consume more calories per day than before, keeping the same physical activities as in the past. The second one is that agents decreased the rate of calorie burning, while consuming the same amount of calories. Finally it could be due to a change in the equation relating the ins and outs of calories.

Genetics is one possible explanation of the third channel. Those that were born weighing above a certain threshold are more prone to develop obesity problems (Baird et al. (2005), Serdula et al. (1993) and Whitaker et al. (1997)). Moreover, if both of the parents are obese, it is more likely that their child would reach the obesity threshold (Wrotniak et al. (2004), Whitaker et al. (1997)). So, as the proportion of obese increased in the adult population, more and more children were born at a higher risk of becoming obese during adulthood. This process might have induced a change in people's metabolism, making the burning of calories harder than before. But were this to be the case, a simple fixed effect regression would take this channel into account, unless this effect exhibited time variance. Metabolism does change through time, as adults find it harder to burn calories than youngsters. Consequently it is necessary to control for factors correlated with these changes.

The other two channels are trickier to measure, since they have several explanations that are definitely time varying. One possible justification for the decrease in the rate of calorie burning is related to technological change. Technology at work has changed dramatically in the last 30 years in favor of less physically intensive jobs (Lakdawalla and Philipson (2002)). Nowadays, the calories that used to be burnt during the labor intensive working hours have to be burnt during spare time. Therefore, the people that worked in physically intensive jobs drastically changed their pattern of physical activity without an equal change in consumption habits leading to an increase in permanent weight.

The other part of the equation has some possible explanations as well. Since 1976 food price has fallen by more than 12% compared to other goods (Lakdawalla and Philipson (2002)). Although this could be a viable explanation for the increase in average BMIfrom 1972 to 1976 and from 1984 to 1991, food prices increased sharply. Indeed, today's price of food relative to price of all items less food is only 5% lower than in 1972. However, what did change is the cost of the lowest quintile of energy density food compared to the highest quintile. Today, the cost of the former is around $18.61/1000 kcal as compared to only $1.76/1000 kcal for foods in the top quintile (Drewnowski et al. (2007) and Monsivais and Jacobsenski (2007)), revealing a disproportionately unequal increase in prices. On top of this, the increase in the relative price of cooking at home, coupled with a reduction in household time, has made it harder for people to eat healthier at home (Lakdawalla and Philipson (2002)). The increasing female participation in the labor market also made eating outside unavoidable for some households (Fokuda (2006) and Jacobsen (2006)). This problem has been confounded by the growth of the fast food industry[7], decreasing the cost in time of eating outside.

Two features are clear from Figure 2a, the prevalence of obesity across income deciles is such that the lowest decile of income has the largest ratio of obese, and, the decrease in the prevalence of obesity is almost monotonic with increases in income. In fact, the lowest income decile of the population has a rate of obesity that almost doubles that of the highest decile. Both the story of food prices and of technology at work are suitable for explaining the distribution

[6] The World Health Organization has qualified obesity as a disease.

[7] 13% in a 10 year period according to the National Retail Census.

of obesity across income.

However, Figure 2b shows that the increase in obesity rates between the 80's and the 00's was similar for all deciles and even slightly larger for the richest ones. This suggests two issues. First, obesity increase shares some common characteristics among the different income deciles. Second and most importantly, the preceding explanations for weight increase are at odds with this stylized fact. To begin with, if technology at work changed the physical intensity of labor, it did so for the poorest deciles and not for the richest ones. The highest income earners, be it professionals or white collars, were already making little or no physical effort in their work. On the other hand, the poorest income deciles are more prone to budget constraints and therefore more affected by changes in the price of unhealthy food. But this is rarely the case with the higher incomes, as they tend to be more sophisticated in their eating habits and incorporate better food into their diet. Thus, it is hard to explain changes in obesity rates in the first deciles of income using arguments that are best suited for the lowest income deciles.

A good story for modern obesity rates has to explain not only the raise in BMI but also the fact that it affected all income deciles similarly, although it had more incidence on the highest income earners. The decline in smoking rates is a potential candidate for two reasons. First, it affected a significant part of the population. Indeed at the beginning of the eighties, almost 30% of the American population was an active smoker. Moreover, while by 2008 that ratio decreased to less than 20%, it is still remains a significant part of the American population (Figure **??**). Second, that decline was due, among other reasons, to a very aggressive campaign to ban smoking for most public places. Society's demands regarding a healthier environment forced the introduction of a number of changes in the regulation regarding tobacco use (Figure 3a and 3b) reshaping the average American smoking habit. Government and private offices, restaurants, recreational facilities, retail stores and educational institutions, all of them suffered some sort of restriction which in some cases manifested in smoking bans within private buildings and their immediate surroundings. These restrictions, however, had an unequal impact. They affected the most those that worked in offices and ate frequently in restaurants. As a matter of fact, already by 1993 nearly 82% of indoor workers faced some restriction on workplace smoking and 47% worked in a 100% smoke-free environment (Farrelly et al. (1999))

Information policy regarding smoking was yet another reason that helps explain the decline in smoking rates. Smoking advertisements were banned from TV and other mass media and supported by an increase in published information focusing on the causal links between smoking and adverse health[8]. High income and educated individuals were at least as likely to be affected by the anti-smoking campaign as they were in a better position to accumulate, process and understand this information and correctly update their costs of smoking.

[8] A policy that took full strength during the last two decades, beginning with the 1980 General Surgeon's report on the subject.

2.2 Literature Review

Several authors have studied the weight impact of smoking in recent years. However, most of them have investigated the question in a reduced-form, that is, assuming that increases in tobacco prices reduce smoking rates and through this channel impact weight. While it is frequently assumed in the literature that this is actually the case (see for instance Chaloupka (1999)), price increase is not the only mechanism to induce people to quit nor is it the most relevant one. Chou et al. (2004) is the first paper we know in the economic literature to link the increase in BMI to smoking. The authors adapt a behavioral model of the determinants of obesity to pooled individual-level data from the Behavioral Risk Factor Surveillance System, matched with the prices of food cooked at home and fast food, tobacco and alcohol prices, and the number of per capita restaurants and fast foods chains, as well as indicators for the regulation regarding tobacco use in private offices and restaurants. They analyze the determinants of BMI in a reduced-form OLS regression with state level fixed effects and a quadratic time trend. Other regressors include the demographic characteristics of individuals, the prices of several commodities and the number of fast food restaurants. The authors account for the fact that certain regressors are likely to be related to BMI in a non-linear fashion. Among the main conclusions of the study, they find that increases in cigarette prices significantly increases BMI as well as obesity rates[9] and also that it helps explaining a significant proportion of the increase in BMI (up to 20%). Therefore, the authors find that tobacco consumption substituted net calorie ingestion as a habit. This result is very important since it says that part of the increase in BMI is due to policy decisions.

However, the study has some potential flaws. Firstly, the channel of identification is that increases in cigarette prices and the tightening of the regulation regarding tobacco use induced people to quit smoking, reduced the frequency of smoking or deterred the starting of the habit. As mentioned above, there are other reasons, such as health problems, that could explain why some individuals quit the habit. Moreover, tobacco companies could be raising prices in response to a diminishing pool of smokers or authorities could be responding to tobacco derived health problems by raising taxes on its use. In addition, if smoking and eating are substitutes, then the significant effect should be found among perennial smokers or former smokers, but not among those who have never smoked. Unfortunately, the framework the authors use is unable to discriminate among the different subgroups.

Secondly, the authors do not attempt to control for unobserved heterogeneity among individuals. Because eating and smoking are closely connected to preferences then raises concerns about the consistency of the estimation. In particular, risk attitudes might will operate through the exclusion restriction. Thirdly, the study does not account for dynamics in the dependent variable. In the case of BMI this is certainly a problem since adjustment costs are non-negligible. Moreover, the estimated equation does not separate the short run effect from the long run one. A fourth issue is nonlinearities. The average effect might be significantly different from the effect in the obese and overweight sample. This is relevant because the health consequences of the tobacco policy would be higher if the effect is larger for the obese. Moreover, and perhaps more importantly, the reported elasticity of smoking to BMI and obesity is too large to truly believe in the results.

Finally, the regression is not satisfactorily robust. Using the same data set and a similar specification, Gruber and Frakes (2006) finds the opposite effect, that is, increases in tobacco prices significantly decreases BMI ($1.00 rise in taxes lowers BMI by 0.151 and the probability of becoming obese by %0.015). The main differences between the two papers arise due to the use of state excise tax on tobacco instead of tobacco prices and through differences in time effects estimation (the latter authors introduce year dummies rather than a quadratic time trend) . Gruber and Frakes (2006) also instruments the smoking decision by means of a 2SLS regression, where in the first stage they regress the smoking odds against the tobacco excise tax. However, even after correcting for the potential endogeneity, the

[9] A unitary increase in cigarette price leads to an increase of 0.486 in BMI and a 10% increase in the cigarette price would raise by 0.445% the probability of becoming obese for an individual.

resulting coefficients again are too large to be plausible. As in the Chou et al. (2004) their estimation considers neither unobserved heterogeneity and error clustering nor the dynamic problem.

Rashad (2006) performs a similar analysis of Chou et al. (2004) extending the dataset to food and caloric intakes. He separates the analysis to see how the increases in prices and regulation affected caloric intake and tobacco use. Contrary to what should be expected, increases in tobacco prices and regulation did not affect smoking but it did increase caloric intake while changes in food prices did not change caloric intake but changed smoking decision. Nevertheless, he does nothing to correct the mentioned problems in the previous specifications.

Baum (2009) addresses some of the issues of the previous papers through a difference in difference approach, using changes in cigarette prices as the treatment. People that smoked at least 100 cigarettes before the age of sixteen are assigned as the treated group and people that didn't smoke before this age as the control group. He finds a similar result to Chou et al. (2004), that is, a rise in either prices or taxes increases BMI and the likelihood of becoming obese, regardless of the time controls. However, two caveats should be mentioned. First, the study does not use the same dataset as the previous authors and therefore, comparison is limited. Second, he relies on the same assumption as the two previous papers, that raising cigarette costs will lead to a decrease in smoking.

In a very interesting and recent exercise, Courtemanche (2009) revisits Chou et al. (2004), Gruber and Frakes (2006) and Baum (2009) and puts them together using not only the contemporaneous cigarette price/tax but also their lags. He finds that while in the short run increases in cigarette prices might lead to opposite results, in the long run each and every one of the three specifications leads to a decrease in BMI. Moreover, he finds that the decrease in weight is due to both better eating and more exercise. However, because the data he uses is cross-sectional, matching the individual with previous period taxes might lead to an error, in particular if it is done at the state level. Also, the survey used has self reported answers and the error in reported food consumption and exercise should not be overlooked. In addition, a very small fraction of the observations was given the food complementary survey and sample size drops substantially. Last but not least, he uses only increases in the price/tax of cigarettes and does not look at the other dimensions of the anti-smoking campaign.

Eisemberg and Quinn (2006) is the only paper we found that does not rely on the assumption that changes in cigarette prices affects cigarette consumption. In the study the authors use the Lung Health Study, a randomized smoking cessation trial with 5,887 smokers. Unconventionally, however, the authors use weight instead of BMI as the dependent measure and find that the effect of quitting smoking is a weight increase of 10 kg. This paper solves some of the issues mentioned before, however, it is not clear if the entire smoking cessation sample do indeed quit permanently.

3 Data and Methodology

3.1 Data Set Description

The main source of data that we use is the Behavioral Risk Factor Surveillance System, for the period spanning 1984-2007. This is the same data set as in Chou et al. (2004) Gruber and Frakes (2006), with additional waves. The BRFSS is a phone survey designed as a series of independent cross sections with the intention of obtaining information regarding the prevalence of unhealthy habits and behavioral risks among the US population above 18 years and living in family households[10]. The BRFSS survey started in 1984 and since 1995 all states have been participating continuously. The number of yearly interviews has been constantly increasing and by 2007 it was more than 270,000. The survey is a rich source for demographic and economic status variables including state of residence, number of children, race,

[10]More information is available at www.cdc.goc/nccdphp/brfss.

family income, education, marital status and age. The survey asks the subjects weight in pounds and height in foot and inches. We transform these measures into the metric correspondence and from this we calculate the Body Mass Index, calculated as height over weight squared. From this survey we also obtain information on tobacco and alcohol consumption, including whether the person has smoked more than 100 cigarettes during his/her life, whether he or she currently smokes, the number of cigarettes smoked, whether the individual has ever tried to quit and if the individual drinks regularly. Because the data in the survey is self-reported, in order to avoid extreme self reporting bias, we only include observations for people that reported a BMI above 13 and below 100, the complete valid sample yields us 3,286,800 observations[11].

Other sources of data are the Bureau of Labor for the state unemployment rate, consumer price index, food price and number of fast food restaurants, and the Bureau of Economic Analysis for the quarterly per capita income of the state which are used to control for the business cycle. Finally, we complete the data set with an index of regulations regarding tobacco use and the effective real tax on tobacco that we develop using data from the National Cancer Institute State Legislative Database Program. Following Chriqui et al. (2002) we accountfor all the effective changes in state regulations regarding tobacco use from 1970 to 2007 that affected the ability of a smoker to smoke in his daily activities. However, we only concentrated in those laws that had an effective enforcement.

In order to construct the index we identified seven different categories: Government offices, Private offices, Restaurants, Recreational public places, Hospitals, Educational facilities and Public Transport. The index goes from 0 to 5 for each category, except for transport that goes from 0 to 3. The higher the number, the tighter the regulation. The categories are 0 for no regulation, 1 whenever there is some restriction to smoke but does not impose a high cost on the smoker in terms of his time budget, two if smokers and non smokers have to be in a separated room, three if smoking is banned in certain areas, four if smoking is prohibited within the building and five if it is also prohibited in the surrounding areas of the building. For instance, a category 5 in Private office means not only that smoking is not allowed in private places of work, but also within a certain distance from the entrance to the building. Whenever the law creates an important exception, we subtractone point from the index. Because small and medium firms employ a large proportion of US workers, the deduction was considerably higher in case the law exempted this type of business. Using this regulatory data, we constructa new variable that tries to capture tightness of the regulations regarding tobacco consumption in the state. We addthe punctuation the state received in each category and normalize the new variable by its maximum possible score to make it continuous between 0 and 1.

Table 2 contains summary statistics of all the variables we use in the study. Several things are worth mentioning of this first exploration of the data. As we can see, the average sample individual is overweight. Almost 25% of the sample smokes although with a large variance. This is due to the fact that throughout the years, smoking rates have diminished considerably. The unconditional average amount of cigarettes smoked in the sample is four. In order to control for other habits we include whether the individual drinks regularly and whether they exercise regularly. In the sample, 53% of the individuals reports to drink regularly[12], 66% reports doing exercise regularly, 46% of the sample has kids, 31% has a college degree and almost 60% is married.

Table 2 also reports the different summary statistics for the never-smoker, current smoker and past smoker groups. The first group is the one that shouldn't be affected by changes in taxes on tobacco or regulation regarding tobacco use. The second group is the one we would be interested in using in a randomized experiment of quitting smoking. Because that is not available, we are going to compare it to the third group, that is, quitters. As we can see, the last group is the one with the largest BMI, while smokers are the group with the lowest, this difference being statistically

[11] In the first survey, information was only available for 15 states and the number of useful observations was around 23,882. Although the survey has been growing in scope and coverage, unfortunately the number and quality of questions changes through time. For instance the question on the number of cigarettes smoked is not available after 2000 and the drinking variable is not asked every year.

[12] Although only 13% are binge drinkers, results available on request.

Table 2: Summary Statistics: Individual variables. 1985 - 2007

	Full Sample		**Never-Smokers**		**Current**		**Past**	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Smoke Currently	0.23	0.42	0	0	1	0	0	0
Smoke Ever	0.49	0.5	0	0	1	0	1	0
BMI	26.49	5.59	26.47	5.63	25.78	5.46	27.14	5.54
Obese	0.20	0.403	0.21	0.41	0.17	0.38	0.23	0.42
Drink	0.53	0.50	0.48	0.50	0.59	0.49	0.57	0.50
Women	0.58	0.49	0.64	0.48	0.56	0.50	0.5	0.50
Age	47.29	16.22	45.71	16.56	43.8	14.54	53.41	15.26
White	0.81	0.39	0.78	0.41	0.81	0.39	0.86	0.35
Black	0.08	0.27	0.09	0.29	0.08	0.27	0.05	0.22
Hispano	0.06	0.23	0.07	0.25	0.05	0.22	0.04	0.20
Married	0.59	0.49	0.61	0.49	0.50	0.50	0.64	0.48
Divorced	0.16	0.37	0.12	0.33	0.24	0.43	0.16	0.37
Widowed	0.09	0.28	0.09	0.28	0.07	0.26	0.11	0.31
Kids	0.46	0.50	0.48	0.50	0.50	0.50	0.38	0.48
School Dropout	0.11	0.31	0.09	0.28	0.16	0.37	0.11	0.31
High School	0.32	0.46	0.28	0.45	0.39	0.49	0.31	0.46
Some College	0.27	0.44	0.26	0.44	0.28	0.45	0.27	0.44
College	0.31	0.46	0.37	0.48	0.17	0.37	0.31	0.46
Real Income	40.8	22.65	42.77	22.91	34.6	20.8	42.17	22.72
Unemployed	0.04	0.20	0.03	0.18	0.07	0.26	0.03	0.17
Exercise	0.66	0.47	0.69	0.46	0.58	0.49	0.68	0.47
General Health	0.85	0.36	0.88	0.32	0.79	0.4	0.82	0.39

Note: All variables have 3,286,800 useful observations except for the Drink variable which has 2,892,973
The sample contains 1,685,770 never-smokers, 743,216 current smokers and 856,418 former smokers
Source: Behavioral Risk Factor Surveillance System 1985-2007

significant. The average profile of a quitter is usually is someone who is in their fifties, married, white and enjoys a higher level of education and real income than the average smoker. These statistics confirms our initial beliefs. It is hard to argue that technological change or food price changes are the main forces behind the increase in weight. Moreover, it shows that the group with the largest BMI is also the group that has stopped smoking. Table 3 reports the summary of the environmental variables.

3.2 Econometric Methods

Individual weight is a stock variable and it is the result of the combination of genetic, metabolic, behavioral, environmental, cultural, and socioeconomic influences. Accordingly, weight increment is the result of consuming more calories than what is burnt. A natural question then is which of the factors is more relevant to explain the increase in U.S. obesity rate. Weight at birth, weight of the parents, gender and ethnicity are among the most relevant genetic variables that influenceit. This group of variables are invariant trough time and as a result, more related to the steady state weight and not to changes per se, although they might be deeply related to how calories are processed. In the group

Table 3: Summary Statistics: Enviromental Variables. 1985 - 2007

	Mean	Std. Dev.
Price Food Away	162.94	3.67
Price Food Home	160.41	4.69
Number Fast Food Restaurant per capita	1.83	0.32
Tobacco Price	2.8	0.8
Excise Tax on Tobacco	0.54	0.43
Alcohol Price	165.49	5.67
Tax on Beer	1.72	0.49
Clean Indoor Air Regulation Index	0.31	0.31
Restaurants	1.54	1.74
Private Offices	1.17	1.61
Government Offices	1.94	1.77
Recreational Places	1.81	1.78

of cultural, behavioral and environmental variables the main determinants are civil status, family composition, education, place of residence, veteran of war, employment situation, industry, tenure, hours of work, household income, wife work status, health status, previous period weight and relevant habits. This second group of variables contains variables both constant in time and some that exhibit time variation. The third group of variables, socioeconomic, consists primarily of food prices, sin goods prices and regulations, all of them time variant. So, in order to investigate the effect of quitting smoking on the individual weight we have to control for this thee groups of variables. Finally, in order to make weight comparable across individuals it is necessary to normalize it. This is usually done by dividing it by height squared, which is called Body Mass Index. This will be the outcome variable in our study.

So for individual i, who resides in state j at year t, the effect of quitting smoking on weight ideally would be estimated through the following equation:

$$BMI_{ijt} = \alpha + \beta_1 X_{ij} + \beta_2 Z_{ijt} + \gamma \text{Quit Smoking}_{ijt} + \delta_j + \delta_t + \eta_i + u_{ijt} \tag{1}$$

Some of the variables mentioned above are not present in the BRFSS dataset but unfortunately correlated with both weight and quitting smoking. A second problem with the dataset is its cross-sectional structure which does not allow us to control for unobserved heterogeneity with fixed effects. Because, preferences for health can explain both BMI and smoking, not being able to include fixed effects for the individuals might bias the estimation. An additional problem, not related to the dataset, is the fact that quitting smoking, as a decision, might be influenced by BMI, as was previously explained in the introduction. As a result, the coefficient of interest in Equation (1) , that is γ, will suffer from conditional bias and there is no a priori idea of the direction.

Fortunately, the fact that most of the non available variables are fixed through time allows us to control for them by means of fixed effects. That is, if panel data would be available, this problem could be solved by treating η_i as a fixed effect, using a transformation of the model or parameterizing the conditional expectation of the individual effects as a function of the explanatory variables. Therefore, solving the unobserved heterogeneity problem means also solving the missing variables one. In order to do that we use cohort analysis. This technique, developed by Deaton (1985) and further improved by Browning et al. (1985), Moffit (1993) and Collado (1997) among others, allows us to control for fixed effects, use lags of variables as instruments. The basic idea of this procedure is to construct population means of the cohorts, in order to form a panel structure for the data. To do that, Deaton (1985) recommends to divide the

population in cells with homogenous individuals and to form cohorts according to one or several characteristics which remains constant in time. For that purpose, in our dataset we could consider date of birth, sex, race and residential location.

Equation (1) then will be transformed into (2), where now BMI_{cqt} stands for the BMI of cohort c at quarter q and year t.

$$BMI_{cqt} = \alpha + \beta_1 X_c + \beta_2 Z_{cqt} + \gamma \text{Quit Smoking}_{cqt} + \delta_t + \eta_c + \epsilon_{cqt} \tag{2}$$

When analyzing cohort data we must bear in mind that all cohort variables[13] are error ridden measurements of the true cohort population means. The advantage with respect to standard errors-in-variables models is that we can estimate the variances of the measurement errors using individual data. Moreover, if the size of the cohort is large enough, sample means approximate well enough their population counterparts.

We define n_{cqt} as the size of cohort c in quarter q of year t. Every element of $\bar{X}_{cqt}$, for example a dummy for education, is the average (proportion) of individuals in that category of eduction observed for individuals belonging to cohort c in quarter q of year t, and analogously for other variables in the model. The main estimation problem is that $\bar{\eta}_c$ is unobservable and likely correlated with some variables in $\bar{X}_{cqt}$. Therefore, Equation (2) does not constitute an appropriate base for obtaining consistent estimates, unless the size of the cohorts is large enough. In this case, $\bar{\eta}_c$ is a good approximation to η_c, and we can replace $\bar{\eta}_c$ by a set of binary variables (fixed effects) one for each cohort. A natural estimator then, is the covariance or within groups estimator based on the weighted means of the cohorts, where the weights take into account potential heteroskedasticity between cohorts.

Let $\bar{X}_c = (\sum_{q=1}^{Q} \sum_{t=1}^{T} n_{cqt})^{-1} X_{cqt}$ be the average of the observed means for cohort c, and define $\bar{Y}_c$ analogously. Then $\hat{\beta}_{WG}$ will be biased in small samples but it will be consistent as n_{cqt} tends to infinity if standard assumptions about second order moments are met. There exists a trade-off between variance and bias of the estimator. That is, the bigger is the number of cohorts (C), the smaller is their size (n_{cqt}).The trade-off has to be solved in such a way that the variation within cohorts is small, i.e. homogenous individuals, while the variation between cohorts is large, i.e. heterogenous cohorts. Identification of the true parameter requires that the expectation of each element conditional on the cohort identifying variables varies with time. On the other hand, as we have pointed before, enough people in each group or cohort is necessary for the average within a group to be an unbiased estimator of the population mean. Browning et al. (1985) mention that 150 individuals per group is a relatively good number to avoid sampling bias. In this study, we are going to use only those cells with more than 100 individuals within.

3.2.1 Dynamic Specification

Adjusting ones weight is a costly procedure that takes time. As a consequence, past period weight can be a determinant of today's. The BRFSS does not ask about weight in earlier time periods. Thus, previous authors were unable to incorporate dynamics into their estimations, with the resulting potential omitted variable bias in their estimations. Using cohort analysis also gives us the possibility of estimating dynamic models from individuals observations at a single point in time. In this case, the equation to estimate is:

$$BMI_{cqt} = \alpha + \rho BMI_{c(q-1)t} + \beta_1 X_c + \beta_2 Z_{cqt} + \gamma \text{Quit Smoking}_{cqt} + \delta_t + \eta_c + \epsilon_{cqt} \tag{3}$$

[13] This includes the cohort specific effect.

Unfortunately, including lagged BMI might also lead to a bias if the panel is too short (Arellano and Bond (1991)). Therefore, dynamics posses yet another methodological issue to solve. The methodology of Blundell et al. (1998) can be used to address that problem by means of a system GMM when individual data is used. Furthermore, Collado (1997) proposes an instrumental variables estimator based on first differencing the model, which corrects the error-in-variables problem for dynamic models in the context of cohort data.

The estimation procedure in those cases relies on the idea that internal lagged instruments can be found, if they are not correlated with future error terms. While the lagged dependent variable is correlated with past error terms and uncorrelated with the current and future error terms, some of the other variables are potentially endogenous given that they are correlated with the current error. Though, if we assume that they are uncorrelated with future error terms, the system GMM includes a restriction which assumes that although lagged BMI might be correlated with the unobservable, the first differences are uncorrelated with $\eta_c + \epsilon_{c,q,t}$, which implies that deviation from long term trends in BMI are not correlated with individual effects.

Fortunately, when the number of available periods is large enough, the error-in-variables problem tend to disappear as shownin Nickel (1981), Browning et al. (1985) and Jiménez et al. (1998). Since we have data on almost 100 quarters, we can estimate the dynamic specification without instrumenting BMI's lag. Therefore, we have two potential methods to estimate consistently the effect of quitting smoking on weight in a dynamic setup.

3.3 A first exploration of the data

As a first attempt to understand the issues at hand, we replicatethe results of both Chou et al. (2004) and Gruber and Frakes (2006), with some minor differences. The only correction we make for self-reporting bias is to restrict BMI to lie within the range of 13 to 100. In addition we replace tobacco prices as used in Chou et al. (2004) with data from the Tax Burden on Tobacco[14].

The first two columns of Table 4 show our replication of the original formulation of Chou et al. (2004) with the original sample years and with the full sample years. The third and fourth column refer to the Gruber and Frakes (2006) specification with the years used in the published paper and with the complete waves respectively. The sixth column is the specification we will test.

A first thing to check is whether the estimated coefficients are sensitive or not to the number of waves included. As we can see, the Gruber and Frakes (2006) finding that BMI decreased with increases in tobacco prices is no longer significant once we use the 1985-2007 waves. Accordingly, the only result that does not depend on the sample is Chou et al. (2004), that is, raising tobacco prices leads to an increase in BMIWhen we include quitting smoking as one of the determinants of BMI, the effect of the tax on tobacco and regulations regarding tobacco use is not significant. However, quitting smoking it is.

This means that Chou et al. (2004) and our specification have similar conclusions, although our specification does not rely on a reduced form assumption such that increases in tobacco prices leads people to stop smoking. Our next priority then, is to replicate the exercises of Chou et al. (2004) and Gruber and Frakes (2006) with cohort data and see the impact that unobserved heterogeneity has on the estimated coefficients.

[14]Chou et al. (2004) source their data from the ACCRA cost of living index, which is not publicly available, unlike our measure current measure.

4 Cohort Analysis

4.1 Cohort Definition

Cohorts are defined using the following characteristics: Year of birth, Gender and Geographical region of residence and data is aggregated by quarter and year. Since each cell is the average of individual observations within the cohort, dummy variables will be transformed into the proportion of people within a cell that have a certain characteristic. For instance, currently smoking is defined as either 0 or 1, therefore the transformed cohort variable will tell us the proportion of people among the cohort that smoke.

The structure of the sample in terms of the aggregation variables is the following:

- Year of birth: This grouping has 5 possible categories corresponding to different decades of birth. The first category is for those born before 1940 while the last one is for those born after 1970. The largest proportion of the male population was born during the 50's while the largest proportion of females were born before the 40's.

- Sex: The data set over represents females as they are 58% of the sample.

- Geographical Area: geographical location has been divided into the four categories that the Bureau of Labor uses to produce the CPI. The Southern region is the one more represented, while the Northern East region is the one with the fewestobservations, both for males and females

When we include quitting smoking as one of the determinants of BMI, the effect of the tax on tobacco and regulations regarding tobacco use is not significant. However, quitting smoking it is.

Using this cohort definition and taking into consideration that our dataset goes from the first quarter of 1984 to the fourth semester of 2007 we have a total of 3680 potential observations. Unfortunately some data needed to adjust household income[15] is not available in the 1984 survey, and as a result we dropped that year, leaving a total of 3520 potential observations. Following Blundell et al. (1998), we dropped from the analysis those cells with less than 100 observations in order to avoid sampling bias [16], resulting in 3,439 useful observations.

4.2 Static Specifications

4.2.1 Accounting for Unobserved Heterogeneity in Chou and Gruber Specifications

Since the pseudo panel allows us to apply regular fixed effects analysis, and with that to control for confounders like unobserved heterogeneity, we first explore how sensitivethe results of the Chou et al. (2004) and Gruber and Frakes (2006) specifications are to a fixed effect regression. Recall that Chou et al. (2004) found that cigarette prices significantly increase BMI under a quadratic time trend and a quadratic effect of prices while Gruber and Frakes (2006) used a specification linear in the cost of cigarettes and yearly dummies.

As we can see in Table 6, the linear effect of tobacco prices on BMI[17] in the Chou et al. (2004) specification is now bit higher than in the OLS regression using individual data. Once fixed effects are included the value drops to

[15] Household income is coded as an interval variable so we adjusted the values it by means of a interval regression. In order to do this, several variables were used as predictors, including the number of individuals that live in the house, a question that is asked from 1985 onwards.

[16] Many states were only incorporated after 1995 and the number of interviews has also increased through time and thus some cells have very few observations within. Therefore this is does not represent an endogenous problem between BMI and the number of observations within a cell.

[17] the total effect at the average is 1.35+2*(-0.217)*2.55=0.24.

almost half[18]. The Gruber and Frakes (2006) price effect is also reduced significantly after including fixed effects[19]. This points out that unobserved heterogeneity is an important force behind the results obtained by both papers The immediate question is whether the effect in a structural model is significantor not.

4.2.2 Analysis of Quitting Smoking with Fixed Effects and Instrumental Variables

In this subsection, we conduct regression analyses of the effect of quitting smoking on BMI in a structural model. Using the constructed cohort data, we are now are able to estimate Equation 2 and correct for the potential bias that the simple OLS estimation has. For that, we introduce fixed effects in order to control for unobserved heterogeneity and we instrument the decision to quit smoking in order to estimate the causal effect on BMI. The specification is the same as Chou et al. (2004), except for the inclusion of the decision to quit, the non-parametric time controls, and the use of excise tax on tobacco instead of tobacco prices.

The decision to quit the habit of smoking is instrumented using a one year lag[20] of the tax on tobacco and the numbers of adults within a house. Theseinstruments, from an ex ante point of view, satisfy the exclusion restriction of not being a predictor of contemporaneous BMI, as it is very hard to argue that local governments introduced changes in tobacco taxes in order to modify the weight of the voters. On the other hand, they are relevant for quitting smoking. A 10% increase in tobacco taxes leads to a 4% decrease in smoking prevalence (U.S. Department of Health and Human Services (2000)) and smoke free workplaces reduce smoking incidence by 6%. Nevertheless, we report both Hansen's test for excluded restriction and Cragg-Donald's test for instruments weakness (Stock et al. (2002)), in order to know if they are good from an ex post analysis. The findings for *Quitting Smoking* are reported in Table 7.

Column (A) contains the estimates for the decision to quit smoking using an OLS regression[21]. The effect of quitting smoking on BMI is negative and significant, something at odds with the same regression using the individual data, yet the effect is small in terms of BMI's variability. The specification in column (B) contains cohort fixed effects. As we can see, once unobserved heterogeneity is taken care of, the sign on the coefficient changes and the effect becomes insignificant. This means that the omission of unobserved confounders introduces a negative bias on the estimated coefficient.

Once we add instrumental variables for the decision to quit smoking, column (C), the effect increases and turns significant once again, which is further evidence of the direction of the bias in the OLS regression. Quitting smoking has a positive effect on weight once unobserved heterogeneity has been taken care of and the decision to quit instrumented. The implied elasticity of quitting smoking to BMI is 0.048[22]. That is, a 10% decrease in the incidence of smoking leads to an increase of 3 pounds in the weight of the average cohort, that is, a 2% increase, assuming a constant height. A Hansen J test on the validity of the exclusion restriction fails to reject the null hypothesis, which means that the instruments are not rejected as such. This test is similar to Sargan's test but allows for heteroskedasticity and therefore more suitable for our specification. On the other hand, the Cragg-Donald's test on weak instruments is above 20, which means that the estimated effect is within the 5% bias interval, so we need not be worried that the results are driven by the wrong set of instruments.

Specification (D) reestimates equation (2) using the log of BMI instead of BMI. As we mentioned in the introduction, BMI is the result of dividing weight by height squared. Since the information in the survey is self reported, the measurement error regarding weight and height would not be linear and as a result the standard conclusions of

[18] the total effect also drops almost half to 0.16.

[19] However, it is true that the coefficients are higher in absolute value than when using individual data.

[20] In the present context, that is a four period lag, since our data is aggregated by quarters.

[21] As explained in Table 7, several controls were included. Except for the price of food at home, the availability of fast food restaurants and having children, all the other controls have the expected sign.

[22] Full tables are available upon request.

measurement error in the endogenous variable do not apply here. In that sense, the log of BMI will log linearize the error. The estimated effect in this case, 17%, is the growth rate of BMI after quitting smoking and it has a similar value to the one implied in the linear specification. Finally, column (E) studies the impact of decreasing the intensity of smoking but marginally. As we can see, small changes in rate of smoking does not seemto have a significant effect, although the effect is positive [23]. This means that only the complete abandonment of the addiction has a significant impact on weight but minor therapies do not.

Several robustness checks have been performed to see how sensible the results are. We have repeated the experiment including in the cohort only those individuals for whom the habit of smoking is already developed, that is, with individuals 26 years or older. Also, we have tried with more lags of the instruments and with other instruments as well. Finally we have tried a different definition for quitting smoking. Instead of using the proportion of former smokers in the cell we used the change in the number of active smokers. In all the cases the result remains relatively unchanged, although the power of the instruments do change and sometimes the Hansen test is not rejected in the margin. [24]

Gender Differences Column (F) and (G) repeats the experiment of (C) splitting the sample between men and women. As we can see, the effect of quitting smoking is significant for men but not for women[25]. This is at odds with the medical literature (Basterra-Gortari et al. (2010), which finds that both women and men gain weight. The potential explanation for this difference is that women are penalized more than men when they deviate from their "optimal" weight. As a matter of fact, the likelihood of an obese or overweighted women getting married or being hired is significantly lower than that of a men of similar demographics (Hamermesh and Biddle (1994); V. Atella and Vuri (2007); Cawley (2000); Cawley and Danziger (2004) and Brunello and D'Hombres (2007)). As a result, women will act in consequence and will probably eat healthier than their men counterparts or do more exercise in order to avoid the negative consequences of gaining too much weight[26].

Persistence in Time Longitudinal data allows us to test in the context of the static model the time persistence of the effect. The evidence so far says that quitting smoking leads to an increase in weight, but there is no evidence of whether such an effect remains in time or if it vanishes after a few quarters. As a matter of fact, it could well be that the weight which is gained after leaving the addiction is lost in the middle run, like a Christmas or Thanksgiving day effect of eating too much. On the contrary, it could be that the effect remains there, changing permanently the weight of the person. To answer that question we have regressed BMI on the lags of quitting smoking, in order to see whether the effect remains significant after several periods.

Table 8 shows the effect of the different lags of quitting smoking on contemporaneous BMI. That is, specification (C) using the contemporaneous variable, the first lag, fourth (one year) and eighth (two years). The first thing to notice is that even after two years the effect remains significant and positive, although it diminishes moderately after one year, leaving the increase to an approximately 14% weight growth. This means that the steady state weight of the quitter increases after leaving the habit but the dynamics are such that the effect overshoots initially.

4.3 Results for the Dynamic Model

As commented in the introduction, adjustment costs make last period weight an important determinant of today's. Cohort data allows us the possibility to include this variable and instrument it using internal instruments. However,

[23] the regressor here is changes in the number of cigarette smoked. As a result, the effect is positive for reductions. It should be noted that the question on the number of cigarettes smoked was discontinued after 2000 and therefore sample size is smaller.

[24] Results available upon request.

[25] As far as we are aware, this is the first study that finds a difference between men and women.

[26] However, there is no evidence whether the health consequences of obesity differs between the two groups.

because the panel is large enough, in principle the usual Arellano-Bond problem should not be present here. Nevertheless, we have estimated equation (3) instrumenting and without instrumenting BMI's lag.

Table 9 presents the results of the effect of quitting smoking in equation (3). Specification (F) includes cohort fixed effects and instruments the decision to quit smoking[27]. Specification (G) instruments lagged BMI using the difference in the lag of BMI, as in Collado (1997). Specification (H) uses the log(BMI) as the independent variable and its lag as one of the regressors.

In the first three specifications, H, I and J, Quitting Smoking is positive and significantand of a similar magnitude as in the static model. In this context, a 10% decrease in the incidence of smoking leads to a weight increase of about 3 pounds. Lagged BMI is positive and significant in all three specifications, although the magnitude substantially changes when it is instrumented. As a result, the effect of quitting smoking is similar even after taking into consideration the initial situation of the stock variable. On the other hand, the static model conclusions about the differential effect between women and men are also present in the dynamic set up. As we can see, quitting smoking has a significant effect only for men.

[27] Non reported controls are the same as in specification (C) while the instruments for decision to quit smoking are Tax on Tobacco (-4, -8 and -12).

4.4 Robustness Check: Different Cohort Definition

To conclude this section we redefinethe structure of the cohort. Cohort definition plays an important role in controlling for unobserved heterogeneity and a valid question is whether our results are driven by a particular definition. To see how sensitive the results are, we introduced race as one of the variables that define the cohort[28].

As a result of this new definition, more cohorts are added which allow us to get more variation and as before we only utilize those cohorts with more than 100 individuals[29]. Table 10 re estimates specification A, B, C and D using the new cohort definition.

With this new definition we reduce the scope for bias at the expense of increased variance of the estimator. Nevertheless, similar results obtain[30].As we can see, the effect of Quitting Smoking in the static specification (C*) is positive and significant as in the previous cohort definition, although the implied weight growth rate is 14% instead of 19%. In the dynamic specification (F*) the effect is again positive and significant and of the same magnitude as in the previous cohort definition. As a consequence, the redefinition of the cohort does not bring any substantial modification to the conclusions. We safely conclude that quitting smoking increases permanently the weight of a person.

5 The Extensive Margin: An Investigation on the Probability of becoming Obese

To conclude the analysis of cohort data in the context of weight and smoking, we should have a better understanding of the impact that quitting smoking has on the increase in the probability of becoming obese.

A natural specification to investigate the effect of a set of variables on the probability of being obese, given a set of covariates X would be $E[p(Obese)|F(X'\vartheta)]$ where the standard choice of F is the logistic function of the form $\Lambda(z) = \frac{e^x}{1+e^x}$[31], that evaluates the expectation by nonlinear least squares. Unfortunately, as discussed throughout the paper, the explanatory variable of interest, quitting smoking, is potentially correlated with the error term. In addition unobserved heterogeneity can bias the estimation. Each of these problems could be dealt with separately. But the two at the same time are much harder to solve.

A more appealing approach to dealwith these concerns simultaneously, is an equation of the form $p(Obese) = \Lambda(X^{'}\beta + u)$, where u is correlated to X but not to set of instruments Z. Using the logistic transformation of the obesity variable into $y = F^{-1}(p(Obese)) = log(\frac{Obese}{1-Obese})$ the model now allows one to linearly instrument the variables and even use fixed effects through the generalized method of moments as in Arellano and Bover (1995b).

Obesity has been traditionally defined as BMI above 30, while overweight is a BMI between 25 and 30. Although thesethresholds are widely used, critics point out that thesetwo measures do not take into consideration different bone structures or different lifestyles. For instance, American football players will weighmore than a person of their same height, yet in general one would not consider them as overweightor obese. Therefore, we implementdifferent thresholds to determine the participation rate, that is, whether an individual is obese or not, going from a BMI of 25 to a BMI of 40. Consequently, for each threshold the aggregation of individuals in each cohort that are obese according to the threshold, gives us the proportion of obese in each category. This set of variables is the one we used to estimate whether quitting smoking affects the probability of becoming obese. Equation (4) represents the dynamic specification

[28] We tried also to aggregate using month instead of quarter and using States instead of Regions. Similar results were obtained, although the number of cohorts with more than 100 observations was considerably lower.

[29] In this case only 67% of the cohorts remains after removing those with less than 100 observations.

[30] The coefficients are slightly smaller than before and the OLS estimation (A*) is now positive, although not significant as before.

[31] Alternatively the probit function can also be used.