2011

# Public Policy and Smoking Prevalence in High Schools

Jerel Xavier San Gabriel
*Trinity University*, jsangabr@trinity.edu

Jerel Xaver San Gabriel

Dr. John Huston

Econometrics

3/21/2012

**Public Policy and Smoking Prevalence in High Schools**

## *I. Introduction*

In the past two decades, the smoking prevalence for US adults has declined, but national youth smoking prevalence has not (Nakajima 2007, 897; Sloan & Platt 2011, 161). According to YRBS data, current smoking prevalence is lower than it was in 1991 (peaking in 1997), but it has remained stagnant since 2003. Research has shown that smoking initiation tends to occur before the age of 18 (Liu 2010). Therefore, if youth smoking doesn't decline, adult smoking may also cease to decline. Since public health officials aim to reduce the youth smoking prevalence to 16% (DeCicca *et al* 2008, 733), the current stagnation poses a concern for public health officials. Health officials seek to reduce youth smoking prevalence because smoking at an early age can cause many long-term health issues later in life (Nakajima 2007, 897), and thus lead to higher healthcare costs in the future. Additionally, once a person has started smoking, the process of quitting is difficult, and the negative health effects are difficult to reverse. According to research by Adda & Cornaglia (2010), a high smoking prevalence has many negative externalities due to the highly negative effects of second-hand smoke (2). The implication is that reducing the smoking population would in turn reduce the negative externalities associated with smoking.

Since youth smoking is a major public health concern, a large body of empirical economic research has been conducted to determine factors that affect youth smoking prevalence. A paper by Nakajima (2007) has attempted to develop a model of youth smoking prevalence that takes into account peer effects. He bases his research on the idea that more young individuals might be more inclined to start smoking in order to be perceived as “cool” or “popular.” He believes that the demand for smoking is highly susceptible to bandwagon effects. That is, he believes that higher levels of usage lead to an increase in the demand for smoking among youths.

In the realm of effects of public policy, research by Adda & Cornaglia (2006) indicates that smoking bans don’t have a significant impact in terms of reducing the negative effects associated with smoking. They reason for this is that smoking bans just lead smokers to smoke at home instead of in public places where smoking is now banned. As a result, the smoker’s family experiences a greater degree of second-hand smoke, and hence they might develop more severe health conditions (23). Additionally, research by Liu (2010) finds that smoking bans do not have any effect on youth smoking decisions (i.e. smoking initiation or cessation), since smokers are likely to just substitute one place for another (2932).

However, the research on the effect of cigarette excise tax yielded mixed results. Research by Liu (2010) indicated that the recent trend of increases in cigarette taxes have had a significant contribution on the decline in youth smoking prevalence from 1992-2002. However, research by DeCicca *et al* (2008) indicates that the price of cigarettes has a weak and statistically insignificant influence on youth smoking participation (745). However, in an NBER Working Paper, DeCicca, Kenkel, & Alan (2008) found that higher cigarette taxes have a weak effect on youth smoking decisions by encouraging cessation, although they did not find cigarette taxes to have a significant effect on youth smoking initiation.

*II. Model*

In this paper, my goal is to estimate the impact of anti-smoking public policy such as cigarette taxes and smoking bans on the smoking prevalence in high schools. Therefore, I propose a model of youth smoking prevalence as a function of cigarette taxes, the existence of smoking bans, while controlling for other factors that might affect the prevalence of smoking amongst high school students. The basic form of my model is as follows:

$$Y = \beta_1 + \beta_2 ALCOHOLPREV + \beta_3 MJPREV + \beta_4 BAN + \beta_5 CIGTAXSTATE + \beta_6 CIGTAXFED + \beta_7 PE + \beta_8 TIME + \boldsymbol{\beta_{ST}} \bullet \mathbf{ST} + e$$

The dependent variable, Y, is the youth smoking prevalence (SMOKEPREV), measured by the percentage of high school students who have smoked at least once within the 30 days prior to the date the survey was conducted. While there are other measures of youth smoking prevalence are available in the YRBS dataset, I chose this one since other measures were too broad (such as percent of high school students who have smoked at least once in the past) or too narrow (such as percent of high school students who smoked on 20 or more days in the 30 days prior to the survey).

The first two independent variables included in the model are alcohol prevalence (ALCOHOLPREV) and marijuana prevalence (MJPREV). Both are defined as the percentage of high school students that have used alcohol or marijuana (respectively) at least once in the 30 days prior to the survey. According to research by Göhlmann *et al* (2008), alcohol and tobacco are complementary goods (22). Since I was unable to find a good source of alcohol prices that suited my dataset, I am using alcohol prevalence as a proxy. By the same logic, since research by Clements *et al* (2010) claim that marijuana and tobacco are also complementary goods (12), I am using marijuana prevalence instead of marijuana prices since marijuana prices are unavailable. However, the use of marijuana prevalence in this model comes with some caveats that I must discuss. There is ongoing debate regarding the direction of causality between marijuana and tobacco use.

According to Chaloupka *et al* (1999), previous research suggests that tobacco has been regarded as a gateway drug to marijuana (2). However, other research suggests that some marijuana users use tobacco to enhance their "high" (17). Nevertheless, I believe that the use of marijuana prevalence in this data also captures some peer effects. Since research by Nakajima (2007) suggests that peer pressure have an effect on a youth's tendency to start or continue smoking (900-901), I decided to include MJPREV in my model.

Since both alcohol and marijuana are economic complements to tobacco, I hypothesize that the coefficients of ALCOHOLPREV and MJPREV to be positive. That is, an increase in alcohol or marijuana prevalence in high schools by 1% will yield an increase in the smoking prevalence in high schools.

In this study, the policy variables that I'm interested in are BAN, CIGTAXSTATE, and CIGTAXFED. BAN, is an indicator variable where BAN = 1 if there is a statewide smoking ban in effect. This variable does not take into account local smoking bans. CIGTAXSTATE is the excise tax (in cents) per pack of 20 cigarettes levied by state governments. CIGTAXFED is the excise tax (in cents) per pack of 20 cigarettes levied by the federal government. Later on, I will use the variable CIGTAXFED2 which is a modified version of CIGTAXFED2.

For all three policy variables (BAN, CIGTAXSTATE, CIGTAXFED), I hypothesize that their effects on smoking prevalence to be negative. Since all three policies impose greater costs on smokers, I expect these policies to reduce the prevalence of smoking amongst high school students. That is, holding other variables constant, I expect that the existence of a statewide smoking ban will reduce the percentage of current youth smokers. Moreover, I also expect that a 1 cent increase in the tax on cigarettes (both state and federal) to reduce the smoking prevalence in high schools.

The PE variable was not initially part of the model. However, while looking at the YRBS questionnaire, I noticed questions on youth physical activities. In this model, PE represents the percentage of high school students who attend a physical education class at least once on an average week. I wanted to determine whether having physical education classes would lead students to make better health decisions such as not smoking.

The TIME variable is the number of years since 1990. That is, TIME = YEAR – 1990. Since my dataset contains data from about 18 years of YRBS results, I wanted to control for changes over time that might have an impact on smoking prevalence. In other words, this particular variable is being used as a proxy for other determinants of youth smoking that are difficult to measure such as changes in attitudes towards smoking, etc. I'm not entirely too interested in whether this variable has an effect or not since I'm using it to measure other things that I have not included in my model.

The final component of my model is **ST**, which is a vector of state dummy variables. $\boldsymbol{\beta}_{\mathbf{ST}}$ is the vector of corresponding state indicator variable coefficients. In my dataset, some states did not conduct YRBS as much as others. As a result, the data from those states might exhibit behavior that should be controlled for. Furthermore, some particular states might have some special characteristics that make them more impervious or sensitive to antismoking policy. For example, states that grow tobacco might be less likely to increase state cigarette taxes.

## *III. Data*

The data for physical education participation and youth smoking, alcohol, and marijuana prevalence were collected from the Youth Risk Behavior Surveillance System (YRBSS) which can be found on the website of Centers for Disease Control and Prevention (CDC). This portion of my dataset consists of the estimates of smoking, alcohol and marijuana prevalence between 1991-2009, categorized by survey year and by state. The Youth Risk Behavior Survey (YRBS) is conducted at the national level every other year since 1991. Unfortunately, the national data do not have indicator variables for state. Instead, I had to use results from statewide YRBS data, which does contain physical education participation, and smoking, alcohol and marijuana prevalence estimates—which are the data points that I need. However, the statewide data comes with a few caveats. First, some states do not conduct a statewide YRBS such as California or Virginia. Moreover, some states that do not conduct YRBS as frequently as the national YRBS is conducted. That is, there are some years between 1991 and 2009 where the national YRBS was conducted, but some states did not conduct a YRBS. As a result, some states have more data than others. Moreover, in some years, there may be a missing piece of information because some states did not collect information on physical education participation in the early years of the survey. As a result, I could only really use 252 out of 270 data points in my dataset. I felt uncomfortable trying to estimate the missing information since my sample size per state was 10 at the most.

I collected statewide smoking ban information American Lung Association's Tobacco Policy Project/State Legislated Actions on Tobacco Issues (SLATI) website. The website provides detailed information on smoking bans by state. I used this website since they also list the dates of when the smoking bans came into effect. In collecting data for BAN, I assigned the value of 0 if a statewide smoking ban was not in place on or before the January 1 of that year. For example, for a data point in 2005, I would put a 0 for BAN if there was no smoking ban in that state prior to January 1, 2005, and 1 if there is a smoking ban before January 1 of that year. I chose January 1 as a cut-off date since the YRBS is conducted in school, so any smoking bans passed before the spring semester starts might have an effect on the YRBS results. This particular variable is somewhat shaky since some states that don't have smoking bans do have many localities, or even some large urban areas, that have smoking bans in effect. Ideally, I should incorporate this data into my model, but it's not easy to find data on when localities enacted their smoking bans.

I collected cigarette tax information from an annual compendium of tobacco revenue, which was produced by the economic consulting firm Orzechowski and Walker (2010). I kept the tax data in cent form so that the coefficients for CIGTAXSTATE and CIGTAXFED will represent the effect of a one cent increase in the excise tax. As mentioned in the previous section, I created a variable called CIGTAXFED2, in which I assumed that the 2009 tax increase had not taken effect at the time of the 2009 survey. I will discuss the reasoning behind this in the next section.

## *IV. Results*

My initial regression did not include any state dummy variables: Y = $\beta_1$ + $\beta_2$ALCOHOLPREV + $\beta_3$MJPREV + $\beta_4$BAN + $\beta_5$CIGTAXSTATE + $\beta_6$CIGTAXFED + *e*. I wanted to see how good the model was before I included any state indicator variables. I ran this regression in order to determine if any of the variables I'm interested in will have a significant effect. By looking at the p-value for the F-statistic is 0.000000, I can reject the hypothesis that $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$. Moreover, by looking at the p-values for the coefficients in this regression, all of them are significant at the 5% level, with PE and CIGTAXFED the only variables that are insignificantly different from zero at the 1% level. Additionally, the $R^2$ in this original regression is .729307 which means that the model (in its current form) explains roughly

72.9% of the variation in smoking prevalence in high schools. However, this regression has one glaring problem. My hypothesis was that the coefficient on CIGTAXFED was negative. However, I failed to reject the null hypothesis that $\beta_6 = 0$ against the alternative that $\beta_6 < 0$. This poses a problem since economic theory suggests that increases in the price should decrease consumption.

Dependent Variable: SMOKEPREV
Method: Least Squares
Date: 11/27/11   Time: 21:46
Sample: 1 270
Included observations: 252

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 9.222690 | 2.452076 | 3.761176 | 0.0002 |
| ALCOHOLPREV | 0.432313 | 0.045048 | 9.596735 | 0.0000 |
| MJPREV | 0.271134 | 0.046161 | 5.873595 | 0.0000 |
| PE | -0.040123 | 0.018529 | -2.165394 | 0.0313 |
| BAN | -2.263360 | 0.856682 | -2.642009 | 0.0088 |
| CIGTAXSTATE | -0.037102 | 0.006196 | -5.987678 | 0.0000 |
| CIGTAXFED | 0.038652 | 0.016478 | 2.345642 | 0.0198 |
| TIME | -0.578921 | 0.083170 | -6.960671 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.729307 | Akaike info criterion | 5.811745 |
| Adjusted R-squared | 0.721541 | Schwarz criterion | 5.923791 |
| Sum squared resid | 4627.654 | Durbin-Watson stat | 0.922755 |
| F-statistic | 93.91268 | Prob(F-statistic) | 0.000000 |

A quick inspection of my data on the federal excise tax on cigarettes showed an enormous tax hike from $0.39 to $1.01 in 2009. As it turns out, this tax hike came into effect on April 2009, and thus, the tax hike may not have been in effect yet when the 2009 YRBS were conducted. So, I created a new federal tax variable, CIGTAXFED2, wherein I assumed that the federal tax on cigarettes in effect in 2009 was still $0.39. When I ran the above regression with CIGTAXFED2[1] instead of CIGTAXFED, the coefficient for CIGTAXFED2 is negative, and I rejected the null hypothesis that $\beta_6 = 0$ against the null hypothesis that $\beta_6 < 0$. Moreover, this new regression yielded both a higher $R^2$ and adjusted $R^2$. That is, this model is better able to explain

[1]$Y = \beta_1 + \beta_2$ALCOHOLPREV + $\beta_3$MJPREV + $\beta_4$BAN + $\beta_5$CIGTAXSTATE + $\beta_6$CIGTAXFED2 + $e$

the variation in SMOKEPREV compared to the initial regression. Nevertheless, this regression is not without its problems. The coefficient for TIME is positive in this regression. I find this odd because the trend over time seems to be a decline in youth smoking, however, this result suggest that holding everything constant, smoking will increase in the following year. Furthermore, upon checking for multicollinearity, it appears that time and CIGTAXFED2 and TIME are highly correlated ($\rho = 0.91878$). Due to the collinearity of the two variables, the estimates of the coefficients are highly sensitive to changes in the model. Moreover, the residual graph for this particular regression is anomalous, especially in the North Carolina area of the dataset (see Graph A in the Appendix). Nevertheless, I would argue that the improved numbers for $R^2$ and adjusted $R^2$ make this new regression better.

Dependent Variable: SMOKEPREV
Method: Least Squares
Date: 11/27/11 Time: 21:48
Sample: 1 270
Included observations: 252

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 22.63727 | 2.847578 | 7.949658 | 0.0000 |
| ALCOHOLPREV | 0.412996 | 0.041407 | 9.973957 | 0.0000 |
| MJPREV | 0.231373 | 0.042931 | 5.389359 | 0.0000 |
| PE | -0.040629 | 0.017039 | -2.384512 | 0.0179 |
| BAN | -2.733196 | 0.786499 | -3.475143 | 0.0006 |
| CIGTAXSTATE | -0.034683 | 0.005694 | -6.091280 | 0.0000 |
| CIGTAXFED2 | -0.628476 | 0.090251 | -6.963635 | 0.0000 |
| TIME | 0.316592 | 0.131532 | 2.406955 | 0.0168 |

| | | | |
|---|---|---|---|
| R-squared | 0.769093 | Akaike info criterion | 5.652774 |
| Adjusted R-squared | 0.762468 | Schwarz criterion | 5.764820 |
| Sum squared resid | 3947.487 | Durbin-Watson stat | 0.926097 |
| F-statistic | 116.1002 | Prob(F-statistic) | 0.000000 |

To complete the model, I added all the state dummy indicators in my regression[2] since it is unclear which states will have a coefficient significantly different from zero. Upon running this new regression, not only were many state indicator variables insignificant, but TIME and PE

[2] I won't put the regression on this paper because it is too long. You can see the results of this particular regression in the Appendix B.

also became insignificant. Because of this, I ran multiple joint hypothesis tests to determine which variables—as a group—are insignificant. Although the coefficient for PE is insignificant in this regression (p-value of 0.3014), I could not include it in any F-tests since removing the PE variable from the regression changed the sample size from 252 to 270. Using an F-test, I failed rejected the null hypothesis that BAN = TIME = AZ = CO = CT = DE = FL = GA = ID = IL = LA = MD = ME = MT = NH = NJ = NM = NV = NY = OK = UT = VT = 0.[3] This test yielded an F-statistic of 1.2688, with d.f. 22, 199, and a p-value of 0.1961. The 5% critical value is 1.590. Therefore, this null hypothesis is rejected at the 5% level.

Although I was expecting TIME to be a significant variable, this result is not terribly concerning for me given the collinearity between TIME and CIGTAXFED2. It seems that CIGTAXFED2 captures the national attitude towards smoking over the years. Moreover, it seems that after adding controls for different states, there doesn't seem to be a national trend over time since smoking attitudes vary across states. The more interesting result is than BAN is insignificant once the state indicator variables are added in. As I mentioned before, BAN = 1 only if there is a statewide ban in effect. Therefore, this particular variable doesn't take into account the smoking ban instated by local governments. For example, a state such as Texas where there is no statewide smoking ban, will have BAN = 0 despite having many local smoking bans throughout the state, and most notably, Houston. The flimsiness of the BAN variable makes this result not at all surprising since this particular variable doesn't accurately capture the state of smoking ban laws in a state. Moreover, the existence of smoking bans may also reflect antismoking sentiment in states. Therefore, by introducing state indicator variables, the significance of the BAN variable is diminished.

---

[3] The regression for the restricted model is Table 2 in Appendix B and the unrestricted model is Table 1 in Appendix B.

The new regression (Table 2 in Appendix B) with all the irrelevant variables removed presents significant improvements compared to the model without the state indicator variables in terms of both goodness of fit and specification. The $R^2$ for this new regression is 0.921352—a large improvement compared to the 0.769093 in the previous regression. Moreover, the adjusted $R^2$ for the new regression is 0.910676, which indicates that the added state indicator variables improve the model. This indicates that the current version of the model explains roughly 92% of the variation in SMOKEPREV. The improvements in the model can also be seen in the residual graph (Graph 2 in Appendix A). Moreover, both the Akaike and Schwarz criteria went down once the indicator variables were added to the model which indicates this particular regression is better specified than the previous one. However, this model rejected the RESET Test at both the 5% and 1% levels. Therefore, I can't assume that there are no specification errors in this model.

Ramsey RESET Test
Equation: UNTITLED
Specification: SMOKEPREV C ALCOHOLPREV MJPREV PE CIGTAXSTATE CIGTAXFED2 AK AL ARK HI IA IN KS KY MAS MI MO MS NC ND NE OH OK PA RI SC SD TN TX WI WV WY
Omitted Variables: Powers of fitted values from 2 to 3

| | Value | df | Probability |
|---|---|---|---|
| F-statistic | 8.796239 | (2, 218) | 0.0002 |
| Likelihood ratio | 19.55733 | 2 | 0.0001 |

Due to the nature of the question at hand, it's not surprising to see misspecification in the model. There are many important variables wherein (1) the data is unavailable or difficult to access or (2) the data is difficult to measure. Variables included in the first case include demographic data such as race and gender. While demographic data is available at the nationwide YRBS results, that dataset does not separate the data by state. Therefore, using nationwide YRBS results would prevent the integration of state cigarette tax and statewide cigarette ban data into the model. Part of the problem is that I'm using YRBS estimates of smoking, alcohol, and marijuana prevalence in my dataset. I believe that access to raw YRBS data might prove to be more useful in terms of both separating data by demographics as well as achieving estimates with higher degrees of freedom. The YRBS results also have other shortcomings. It lacks data on other factors that affect risky behavior such as parent's income and level of education. Research by Saffer *et al* (2007) indicates that these factors affect the smoking.

With regards to the variables that are difficult to measure, some research has been done to identify and quantify other factors that might affect youth smoking prevalence. Research by DeCicca *et al* (2008) has attempted to create an index to measure both anti-smoking sentiment and the strength of youth access laws. Under ideal circumstances, I would have included their indices in my data, but they only calculated data for three years, all of which were pre-2000. Given the high $R^2$ and adjusted $R^2$ values of the current regression, I did not want to include their anti-smoking sentiment and youth access indices because it would have slashed a significant portion of my sample size, and my model would lose its forecasting ability since I won't have any data from the last 9 years (i.e. data points from 2001-2009). Nevertheless, their data still proved to be somewhat useful when explaining the state indicator variables that were included in the model. It's plausible that the state indicator variables were able to capture characteristics that are particular to the state such as the general anti-smoking sentiment in the state (which is high in Massachusetts and low in Kentucky). Moreover, 7 out of the top 10 tobacco producing states showed up as significant state indicator variables. Therefore, despite the misspecification errors present in the model, it appears as if the some of the information contained in the missing variables were somewhat captured by the variables that were included in the model.

## *V. Conclusion*

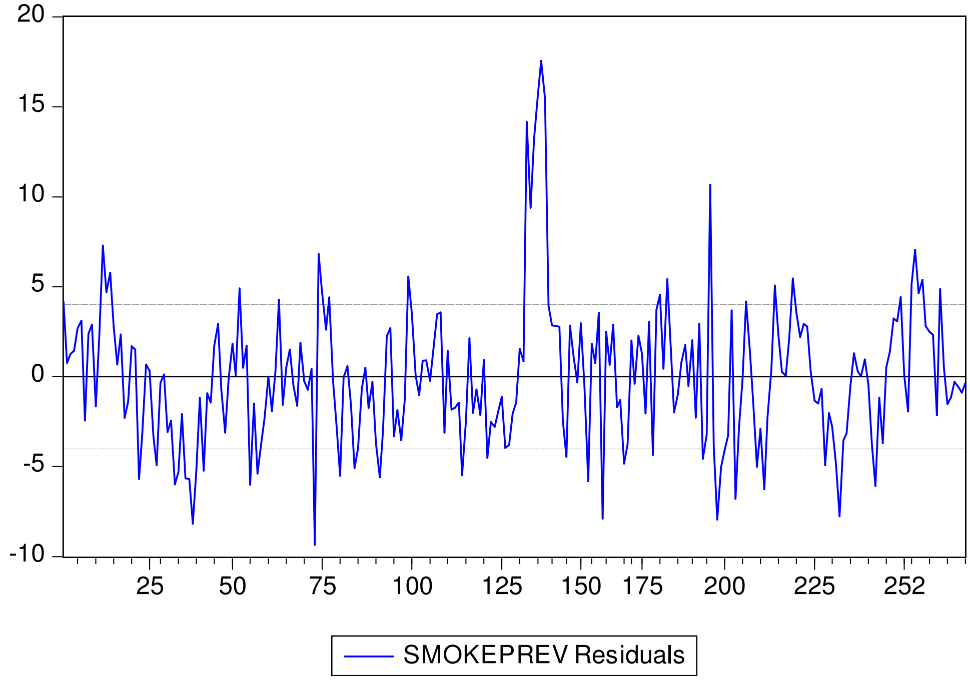
The results of this study show that cigarette taxes significantly reduce the prevalence of youth smoking. This result supports the findings in Liu's (2010) study. Additionally, after controlling for individual states, I also found that statewide smoking bans do not have a statistically significant effect on youth smoking prevalence. This result is also similar to that of Liu's (2010) and Adda & Cornaglia's (2006). However, in the future, a more substantial measure of smoking bans should be incorporated into the model—that is, a measure that would allow the inclusion of local smoking bans into the equation.

In my study, I find that physical education participation has a statistically insignificant impact on youth smoking prevalence. However—should the data become available—it would be interesting to see how effective health classes are in reducing youth smoking prevalence.

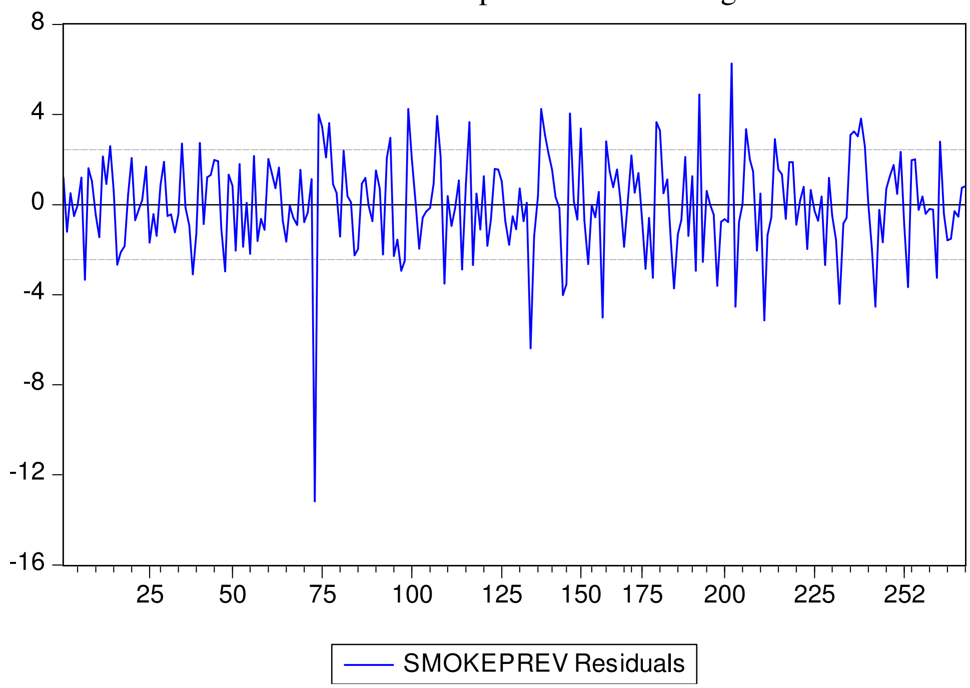
However, due to the possibility of biased results, future research should include more variables, such as race and gender, in order to obtain less biased results. Moreover, the integration of state anti-smoking sentiments and youth access laws might prove to be useful variables in future research. (DeCicca *et al* 2008, 745). Additionally, accounting for other risk factors such as income levels and parent's education might also yield interesting results as suggested in the results by Saffer *et al* (2007).

Appendix A: Graphs

**GRAPH 1:** Residual Graph for:
$Y = \beta_1 + \beta_2 ALCOHOLPREV + \beta_3 MJPREV + \beta_4 BAN + \beta_5 CIGTAXSTATE + \beta_6 CIGTAXFED2$




**GRAPH 2:** Residual Graph for the Final Regression



## Appendix B: Long Tables

**TABLE 1**: Regression with all the state dummy variables

Dependent Variable: SMOKEPREV
Method: Least Squares
Date: 11/27/11 Time: 21:35
Sample: 1 270
Included observations: 252

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 15.16499 | 6.251360 | 2.425870 | 0.0162 |
| ALCOHOLPREV | 0.252446 | 0.074308 | 3.397311 | 0.0008 |
| MJPREV | 0.561483 | 0.056972 | 9.855490 | 0.0000 |
| PE | 0.044550 | 0.043001 | 1.036028 | 0.3014 |
| BAN | -0.279180 | 0.600528 | -0.464891 | 0.6425 |
| CIGTAXSTATE | -0.015327 | 0.005794 | -2.645523 | 0.0088 |
| CIGTAXFED2 | -0.517169 | 0.061189 | -8.451961 | 0.0000 |
| TIME | -0.100258 | 0.100444 | -0.998145 | 0.3194 |
| AK | 3.076550 | 2.748305 | 1.119435 | 0.2643 |
| AL | 6.779433 | 2.705574 | 2.505728 | 0.0130 |
| ARK | 8.637930 | 2.904484 | 2.973998 | 0.0033 |
| AZ | 3.212447 | 2.798305 | 1.147998 | 0.2523 |
| CO | -1.271020 | 2.891825 | -0.439522 | 0.6608 |
| CT | 0.322940 | 2.253503 | 0.143306 | 0.8862 |
| DE | 0.101695 | 2.783340 | 0.036537 | 0.9709 |
| FL | -0.432887 | 2.878344 | -0.150395 | 0.8806 |
| GA | 1.519383 | 2.908951 | 0.522313 | 0.6020 |
| HI | 2.546473 | 2.808896 | 0.906574 | 0.3657 |
| IA | 6.732054 | 1.995636 | 3.373387 | 0.0009 |
| ID | 1.680394 | 2.638394 | 0.636900 | 0.5249 |
| IL | 2.007968 | 1.785613 | 1.124526 | 0.2621 |
| IN | 6.429751 | 3.017570 | 2.130772 | 0.0343 |
| KS | 6.203871 | 2.725332 | 2.276373 | 0.0239 |
| KY | 9.253723 | 3.246287 | 2.850556 | 0.0048 |
| LA | 2.335108 | 2.500770 | 0.933756 | 0.3516 |
| MAS | -2.068607 | 1.797634 | -1.150739 | 0.2512 |
| MD | -0.053362 | 3.014865 | -0.017700 | 0.9859 |
| ME | 1.563014 | 2.676973 | 0.583874 | 0.5600 |
| MI | 4.036800 | 2.735889 | 1.475498 | 0.1417 |
| MO | 4.488994 | 2.501483 | 1.794533 | 0.0742 |
| MS | 5.775509 | 3.179246 | 1.816628 | 0.0708 |
| MT | 1.880010 | 2.188258 | 0.859136 | 0.3913 |
| NC | 18.74775 | 2.766242 | 6.777336 | 0.0000 |
| ND | 8.716859 | 2.504416 | 3.480596 | 0.0006 |
| NE | 6.062422 | 2.637806 | 2.298282 | 0.0226 |
| NH | 2.888275 | 2.839903 | 1.017033 | 0.3104 |
| NJ | 1.518832 | 2.031052 | 0.747806 | 0.4555 |
| NM | 2.573775 | 2.606726 | 0.987360 | 0.3247 |
| NV | -0.338793 | 2.317686 | -0.146177 | 0.8839 |
| NY | -0.540582 | 1.390234 | -0.388843 | 0.6978 |
| OH | 5.191194 | 2.846153 | 1.823933 | 0.0697 |
| OK | -3.842902 | 3.235583 | -1.187700 | 0.2364 |
| PA | 7.788486 | 3.384353 | 2.301322 | 0.0224 |
| RI | -12.80622 | 2.077150 | -6.165286 | 0.0000 |
| SC | 4.142465 | 2.908982 | 1.424025 | 0.1560 |

| | | | | |
|---|---|---|---|---|
| SD | 4.981634 | 2.813550 | 1.770587 | 0.0782 |
| TN | 8.756651 | 3.129757 | 2.797869 | 0.0056 |
| TX | 4.296023 | 2.517824 | 1.706245 | 0.0895 |
| UT | -0.119020 | 2.937897 | -0.040512 | 0.9677 |
| VT | 0.070104 | 2.527409 | 0.027737 | 0.9779 |
| WI | 5.697273 | 1.964060 | 2.900764 | 0.0041 |
| WV | 8.752050 | 2.967627 | 2.949175 | 0.0036 |
| WY | 5.582154 | 2.241608 | 2.490246 | 0.0136 |

| | | | |
|---|---|---|---|
| R-squared | 0.931027 | Mean dependent var | 25.67369 |
| Adjusted R-squared | 0.913004 | S.D. dependent var | 8.252865 |
| S.E. of regression | 2.434186 | Akaike info criterion | 4.801612 |
| Sum squared resid | 1179.127 | Schwarz criterion | 5.543913 |
| Log likelihood | -552.0032 | Hannan-Quinn criter. | 5.100299 |
| F-statistic | 51.65767 | Durbin-Watson stat | 2.025688 |
| Prob(F-statistic) | 0.000000 | | |

**TABLE 2**: Regression with all the state indicator variables

Dependent Variable: SMOKEPREV
Method: Least Squares
Date: 11/27/11 Time: 21:33
Sample: 1 270
Included observations: 252

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 15.05492 | 1.953097 | 7.708230 | 0.0000 |
| ALCOHOLPREV | 0.347727 | 0.031681 | 10.97577 | 0.0000 |
| MJPREV | 0.444618 | 0.035567 | 12.50085 | 0.0000 |
| PE | 0.025292 | 0.013472 | 1.877389 | 0.0618 |
| CIGTAXSTATE | -0.016702 | 0.003927 | -4.253050 | 0.0000 |
| CIGTAXFED2 | -0.556158 | 0.030697 | -18.11745 | 0.0000 |
| AK | 3.301276 | 1.296223 | 2.546842 | 0.0116 |
| AL | 5.697690 | 1.059821 | 5.376085 | 0.0000 |
| ARK | 7.247799 | 0.998326 | 7.259952 | 0.0000 |
| HI | 2.040544 | 1.018708 | 2.003070 | 0.0464 |
| IA | 5.427438 | 1.487803 | 3.647954 | 0.0003 |
| IN | 5.213745 | 1.288518 | 4.046312 | 0.0001 |
| KS | 4.689635 | 1.468314 | 3.193892 | 0.0016 |
| KY | 7.831960 | 1.190965 | 6.576147 | 0.0000 |
| MAS | -2.303428 | 0.895666 | -2.571748 | 0.0108 |
| MI | 3.233822 | 1.015773 | 3.183608 | 0.0017 |
| MO | 3.381878 | 0.936511 | 3.611145 | 0.0004 |
| MS | 4.168366 | 0.985236 | 4.230830 | 0.0000 |
| NC | 18.15219 | 1.070730 | 16.95309 | 0.0000 |
| ND | 6.432186 | 1.232242 | 5.219904 | 0.0000 |
| NE | 4.155237 | 1.320792 | 3.146019 | 0.0019 |
| OH | 3.873747 | 1.156360 | 3.349948 | 0.0010 |
| PA | 8.883277 | 2.591578 | 3.427749 | 0.0007 |
| RI | -10.28208 | 1.430308 | -7.188721 | 0.0000 |
| SC | 2.997607 | 0.951800 | 3.149410 | 0.0019 |
| SD | 3.912045 | 0.940409 | 4.159941 | 0.0000 |
| TN | 7.807449 | 1.173104 | 6.655378 | 0.0000 |
| TX | 2.941628 | 1.278615 | 2.300636 | 0.0223 |

| | | | | |
|---|---|---|---|---|
| WI | 4.394532 | 1.154229 | 3.807331 | 0.0002 |
| WV | 7.625797 | 0.948634 | 8.038714 | 0.0000 |
| WY | 4.128797 | 0.942829 | 4.379157 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.921352 | Mean dependent var | 25.67369 |
| Adjusted R-squared | 0.910676 | S.D. dependent var | 8.252865 |
| S.E. of regression | 2.466545 | Akaike info criterion | 4.758279 |
| Sum squared resid | 1344.529 | Schwarz criterion | 5.192455 |
| Log likelihood | -568.5432 | Hannan-Quinn criter. | 4.932982 |
| F-statistic | 86.29975 | Durbin-Watson stat | 1.890739 |
| Prob(F-statistic) | 0.000000 | | |

References

Adda, J., & Cornaglia, F. (2010). The Effect of Bans and Taxes on Passive Smoking. *American Economic Journal: Applied Economics*, 2(1), 1-32.

American Lung Association. Tobacco Policy Project/State Legislated Actions on Tobacco Issues (SLATI). Retrieved from http://www.lungusa2.org/slati.

Centers for Disease Control and Prevention (CDC). 1991-2009 High School Youth Risk Behavior Survey Data. Available at http://apps.nccd.cdc.gov/youthonline.

Chaloupka, F., Pacula, R.L., Farrelly, M.C., Johnston, L.D., O'Malley, P.M., & Bray, J.W. (1999). Do higher cigarette prices encourage youth to use marijuana?. *NBER Working Papers* (Working Paper 6939). Retrieved from http://www.nber.org/papers/w6939.pdf.

Clements, K. W., Lan, Y., & Zhao, X. (2010). The Demand for Marijuana, Tobacco and Alcohol: Inter-commodity Interactions with Uncertainty. *Empirical Economics*, 39(1), 203-239. doi:http://dx.doi.org/10.1007/s00181-009-0302-x

DeCicca, P., Kenkel, D., Mathios, A., Shin, Y.-J. & Lim, J.-Y. (2008), Youth smoking, cigarette prices, and anti-smoking sentiment. *Health Economics*, 17: 733–749. doi: 10.1002/hec.1293

DeCicca, P., Kenkel, D., Mathios, A.D. (2008). Cigarette taxes and the transition from youth to adult smoking: smoking initiation, cessation, and participation. *NBER Working Papers* (Working Paper 14042). Retrieved from http://www.nber.org/papers/w14042

Göhlmann, S., Requate, T., Schmidt, C.M., & Tauchmann, H. (2008). Tobacco and alcohol: complements or substitutes? A structural model approach. *IZA Discussion Paper Series*, Discussion Paper no. 3412. Retrieved from http://ftp.iza.org/dp3412.pdf.

Liu, F. (2010). Cutting through the Smoke: Separating the Effect of Price on Smoking Initiation, Relapse and Cessation. *Applied Economics*, *42*(22-24), 2921-2939.

Nakajima, R. (2007). Measuring Peer Effects on Youth Smoking Behaviour. *Review Of Economic Studies*, *74*(3), 897-935.

Orzechowski and Walker. (2010). The tax burden on tobacco, 45. Retrieved from http://nocigtax.com/upload/file/148/Tax_Burden_on_Tobacco_vol._45_FY2010.pdf

Saffer, H., Wakefield, M., & Terry-McElrath, Y. (2007). The effect of nicotine replacement therapy advertising on youth smoking. *NBER Working Papers* (Working Paper 12964). Retrieved from http://www.nber.org/papers/w12964.

Sloan, F., & Platt, A. (2011). Information, Risk Perceptions, and Smoking Choices of Youth. *Journal Of Risk And Uncertainty*, *42*(2), 161-193. doi:http://dx.doi.org/10.1007/s11166-010-9111-z

Smoking states and tobacco producing states. Retrieved from http://blog.bestandworststates.com/2010/02/10/smoking-states-and-tobacco-producing-states.aspx.