

Problem Set #3

Prof. Matt Reimer*

Econ A390: Methods for Public Policy Evaluation

Please answer all the questions below within this Microsoft Word document. For all parts, your answer should consist of a concise explanation, along with any *relevant* STATA commands and results. See [these instructions](#) for importing STATA output and graphs into Microsoft Word documents. You are encouraged to work as a group; however, there may be no more than three students per group. Everyone is responsible for submitting one typed file for grading on Blackboard under Assignments. Please indicate the members of your group at the top of the file. You might find it helpful to read through the assignment, then read your class notes and textbook, and only then try and do the questions. *Attempting this assignment without consulting your notes and the text may be hazardous to your health and your grade!!*

Motivation

Since the release of the 1964 Surgeon General's report, private and public health officials have waged an aggressive campaign against tobacco use. This campaign has included policies as diverse as restrictions on advertising, antismoking public service announcements, education programs, youth-access restrictions, and smoking bans in public places. In aggregate, these policies appear to have achieved some measure of success. Over the past 20 years, per capita cigarette consumption and the fraction of adult smokers have declined considerably.

One policy instrument that has received a considerable amount of attention in the last two decades is higher excise taxes on cigarette. For instance, between 1990 and 2003, there were approximately 113 increases in state excise taxes on cigarettes. Over that time, the average state tax on cigarettes has increased from 23 to 73 cents/pack. A number of these tax increases have been incredibly large including 45 hikes of a quarter or more per pack and 15 hikes of 50 cents per pack or more. The interest in trying to reduce smoking through higher taxes is spurred on in part by a large body of economic research showing that cigarette consumption falls after price or tax hikes. A typical study correlates changes in cigarette consumption overtime within a state to changes in price or tax rates.

In this problem set, you will investigate the claim that a public policy that increases the excise tax on cigarettes will reduce smoking. You will make use of `smoking.dta`, which is a state-level panel dataset that includes information on excise taxes imposed on cigarettes and the prevalence of smoking within a state, along with many other important variables. Type `describe` at the STATA command prompt to see a full description of the `smoking.dta` dataset. A codebook that describes the variables in `smoking.dta` in more

*Telephone: 1-907-786-5430, e-mail: mreimer2@alaska.edu.

detail can be found at the [ImpacTeen website](#), from which this data was obtained.

Problem 1: Warm up—Alaska vs. Wyoming

In 1998, the State of Alaska implemented a significant increase in the excise tax imposed on cigarettes, from an average of \$0.39 in 1997 to an average of \$1.09 in 1998. You have been asked to determine the causal effect of this excise tax increase on tobacco use in Alaska, measured by the proportion of Alaskan residents that use tobacco products. Use the information in Table 1 to answer the following questions.

State	Year	
	1997	1998
Alaska	0.267	0.260
Wyoming	0.240	0.228

Table 1: Tobacco use—measured as the proportion of residents using tobacco products.

A. Explain the counterfactual outcome that you would ideally use to determine the causal effect of the tax increase on tobacco use in Alaska.

B. Let $Y_{AK,97}$ and $Y_{AK,98}$ be the proportion of Alaskans that use tobacco in 1997 and 1998, respectively. Suppose that you estimate the causal effect of the excise tax as $Y_{AK,98} - Y_{AK,97}$.

i) What is the value of your estimate? Explain clearly what you are using as your estimated counterfactual.

ii) What assumption must hold for your estimated causal effect to be true? Is it likely that this assumption holds? Why or why not? Provide an example.

C. Your friend tells you that your estimated counterfactual in Part B is not very good. Instead, your friend suggests that you estimate the causal effect of the increased excise tax as $Y_{AK,98} - Y_{WY,98}$, where $Y_{WY,98}$ denotes the proportion of residents in 1998 that use tobacco products in Wyoming, where excise taxes were relatively low and unchanging over this period of time.

i) What is the value of your estimate? Explain clearly what you are using as your estimated counterfactual.

ii) What assumption must hold for your estimated causal effect to be true? Is it likely that this assumption holds? Why or why not? Provide an example.

D. Your economics professor tells you that the assumptions you outlined in Parts B and C are unlikely to hold. Luckily, these assumptions can be relaxed under a difference-in-difference (DnD) estimation procedure.

i) What is the DnD estimate of the causal effect of an increased excise tax on tobacco use in Alaska? Explain clearly what you are using as your estimated counterfactual.

ii) Is the DnD estimation procedure a better method for estimating the causal effect of an increased excise tax on tobacco use in Alaska? Why or why not? What assumption must hold for your estimated causal effect to be true? Provide an example of why this assumption may or may not hold.

E. Suppose that you would like to implement a DnD estimation procedure using more than one year before-and-after the increased excise tax. It turns out that this is easily implemented by estimating the following regression model:

$$Y_{s,t} = \alpha + \beta TREAT_s + \gamma POST_t + \delta POST_t \times TREAT_s + \varepsilon_{s,t}$$

where $TREAT_s$ is a dummy variable that equals one if the excise tax in state s increase, $POST_t$ is a dummy variable that equals one if year t occurs after the excise tax was increased, and $\varepsilon_{s,t}$ is a residual that is independent of receiving the treatment—i.e. $E(\varepsilon_{s,t}|TREAT_s, POST_t) = 0$.

i) Show that δ corresponds to the DnD estimand (*Hint: use the definition of DnD and manipulate the above equation accordingly*)

ii) Estimate the regression model using the data provided in `smoking.dta` for the years 1996-1999, the states Alaska and Wyoming, and the variable `smoking_total` as your outcome variable $Y_{s,t}$. Is your estimate different from the one you found in Part D? Why or why not?

iii) Plot the outcome variable $Y_{s,t}$ over several years for each state. Does your plot provide evidence for or against your assumption in Part D for the DnD estimator? Explain.

Problem 2: Multiple States and Treatment Periods

The previous analysis limited its scope to Alaska's increased excise tax in 1998. A quick glance at the dataset `smoking.dta` shows that multiple states increased their excise tax on cigarettes between 1991 and 2007, suggesting that we can learn a lot more about the causal effect of an increased excise tax on tobacco use if we expand the scope of our analysis to all states. Doing so, however, introduces some complications that our DnD regression equation in Problem 1 cannot handle, which we will address in the following questions.

A. The first problem in expanding our analysis to multiple states is that there is no common “treatment” variable that discretely switches on and off when a state increases its excise tax on cigarettes. Changes in excise taxes range between big and small, and should not be treated equally across all states and years. We can easily handle this by replacing $POST_t \times TREAT_s$ with a single measure of exposure to the policy of interest; namely, the size of the excise tax itself ($excise_{s,t}$).

i) Using `smoking.dta`, regress `smoking_total` on `ex_tax`, using all states and years in the dataset. Interpret the coefficient on `ex_tax` in terms of the effect of excise tax on the prevalence of smoking.

ii) Do you think the coefficient on `ex_tax` in part (i) is a good estimate of the causal effect of excise taxes on tobacco use? Why or why not? Explain your answer in terms of the “assignment” of excise taxes to states and the potential for selection bias.

iii) Your answer in part (ii) may lead you to conclude that we should include some other explanatory variables in the regression equation. Rerun the regression in part (i) with a set of explanatory variables that you think are important to include. Explain why you included the variables that you did. Is the estimated coefficient on `ex_tax` different from the one in part (i)? If so, how did it change?

iv) Do you think the coefficient on `ex_tax` in part (iii) is a good estimate of the causal effect of excise taxes on tobacco use? Why or why not?

B. Your answers to Part A may lead you to conclude that including a set of explanatory variables in the regression equation may not be enough to capture the causal effect of excise tax increases on tobacco use. Suppose a friend suggests that you include a set of dummy variables indicating each year in the sample (with one omitted as a reference year), commonly referred to as *time effects*. These dummy variables take the place of $POST_t$ in the regression equation in Problem 1.

i) Explain the reasoning for including the time effect dummies. What are they capturing in the regression equation that was missing in Part A?

ii) Rerun the regression from Part A and include a set of year effect dummy variables (*Hint: include the term `i.year` in your regression command*). Interpret the coefficient on `ex_tax`. How has its value changed relative to Part A?

iii) Do you think the coefficient on `ex_tax` in part (ii) is a good estimate of the causal effect of excise taxes on tobacco use? Why or why not?

C. Your answers to Part B may lead you to conclude that including a set of explanatory variables with time effects in the regression equation may not be enough to capture the causal effect of excise tax increases on tobacco use. Suppose a different friend suggests that you include a set of dummy variables indicating each state in the sample (with one omitted as a reference state), commonly referred to as *state effects*. These dummy variables take the place of $TREAT_s$ in the regression equation in Problem 1.

i) Explain the reasoning for including the state effect dummies. What are they capturing in the regression equation that was missing in Parts A and B?

ii) Rerun the regression from Part B and include a set of state effect dummy variables (*Hint: include the term $i.state$ in your regression command*). Interpret the coefficient on ex_tax . How has its value changed relative to Part B?

iii) Do you think the coefficient on ex_tax in part (ii) is a good estimate of the causal effect of excise taxes on tobacco use? Why or why not?

D. Let's revisit the states Alaska and Wyoming for the years 1997 and 1998. Suppose that Alaska increased its excise tax by \$1 between 1997 and 1998, and Wyoming did not change its excise tax at all. Also suppose, for simplicity, that all explanatory variables for both Alaska and Wyoming are equal to zero.

i) Show that the population regression coefficient on ex_tax in your regression equation in Part C has a DnD interpretation to it.

ii) Despite its DnD interpretation, the estimated coefficient on ex_tax is likely considerably different than your estimate of the causal effect of excise taxes in Problem 1. Why?

E. At some point during this Problem Set, you may (should) have discussed the necessary assumption that must be true for a DnD estimation procedure to be estimating the causal effect of excise taxes on tobacco use. It turns out that if there is a sufficient amount of years and states in our sample (as is the case here), we can relax this assumption (to some extent) by allowing each state to have its own trend over time.

i) Explain why you would want to allow each state to have its own trend. What is this capturing that was missing in Part C?

ii) Rerun your regression from Part C, including state-specific annual trends (*Hint: include the term $c.year\#i.state$ in your regression command*). Interpret the coefficient on ex_tax . How has its value changed relative to Part C?

iii) Do you think the coefficient on ex_tax in part (ii) is a good estimate of the causal effect of excise taxes on tobacco use? Why or why not?

F. Discuss the "validity" of your analysis in this Problem Set. Are excise taxes an effective approach to reduce tobacco use? Are your results believable? Would you do anything differently than the approach you followed here? Explain.