Econometrics for Policy Analysis

Jared Greathouse

5/27/24

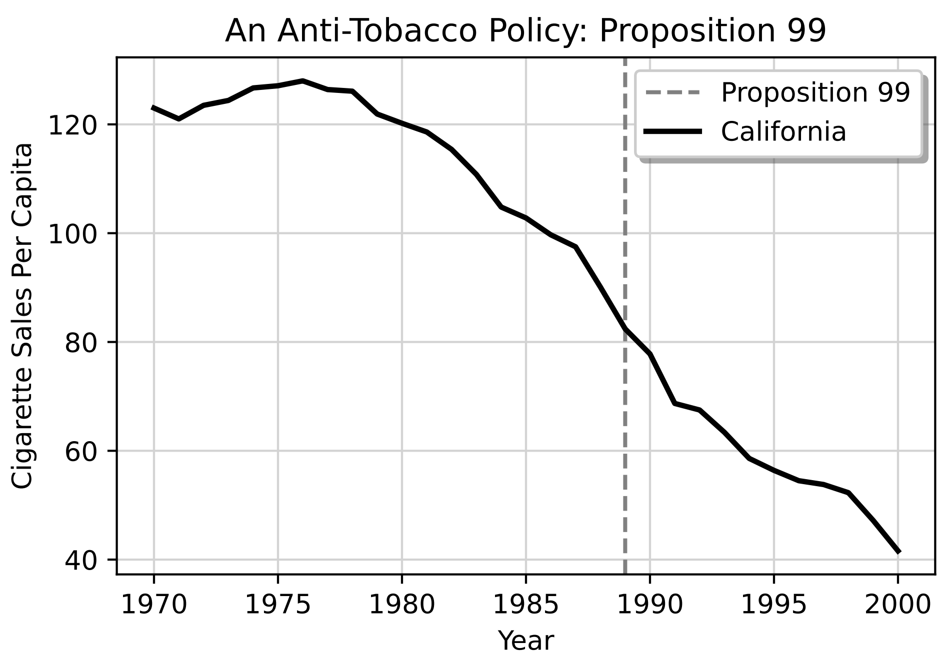
Table of contents

1 Syllabus: PMAP (4041), Fall 2024: Econometrics for Policy Analysis

> **Note**
>
> This is an ongoing project. Comments and suggestions are welcome. Jared Greathouse. Office Hours: By Request.

Every day, governments pass laws/public policy to affect some outcome of interest. Policy usually touches thousands if not *millions* of people. From traffic-circles to pop/sugar sweetened beverge taxes, vaccine mandates and universal pre-k programs, cannabis legalization to minimum wages, public policy impacts us all from birth to death.

Policy is never self justifying. It demands evaluation. If California bans tobacco smoking in public, or if New York City implements gun control, presumably we would agree these *likely* impact outcomes like tobacco use or homicide rates, ideally decreasing both of them.



If California's anti-tobacco policy didn't affect smoking rates at all (or worse, if more people began to smoke) or if gun control has 0 impact on homicide rates (or increased them, paradoxically), then surely these could not be justified in the very first place. Before we continue, understand fundamentally these outcomes being affected *are* the point. The only reason that we, as a society, do policy is precisely **because** we think policy affects (or should affect) people somehow. If political science studies "who gets what where", one summation of policy studies might be "what works?" But what policies should we care about? How can we know if they work? This is the starting point for empirical policy analysis. This class discusses the theory and process for how statistical analysis of data may be used to answer policy questions.

1.1 Course Philosophy and Structure

I believe the best way to demonstrate knowledge of policy analysis is through *writing*. As such, there will be no quizzes or in-class exams. Why? It is unrealistic. In real life, rarely do we have an hour and 30 minutes or a ten minute quiz window on the internet to write a full summation of our ideas or think through a question. Typically, we have much more time and resources to help us. In fact, proper use of resources is what makes a good analyst: good analysts don't need to remember everything, but they do need to be good at *finding answers* and using them sensibly. In this spirit, your sole assignment is to write a paper *applying* the statistical concepts/ideas we cover here to answer questions about a real policy. Here is the breakdown of your grade.

- 35% of your grade comes from the first draft 15% question and 20% draft.
- 60% for the final paper and presentation (respectively, 30 percent each), and
- 5% for attendence.

You must find a real policy that exists which you expect to affect an outcome. You will discuss the justification for the policy (including why we should care about its effects). You'll gather data on the policy of interest. Finally, you'll use the statistical tools we cover (probability theory, descriptive statistics, and regression) to discuss its effects. Typically, we ask causal questions where only one intervention is of interest, but associative papers are not ruled out inherently.

In many senses, public policy is a catch all term covering various disciplines. Public health scholars may care about how banning of abortion in Texas affected fertility rates, or how COVID-19 vaccine/mask mandates affected the COVID-19 case rate per capita compared to other jurisdictions that did not enact these policies. Criminologists may care about how the building of Cop City affected how many people are shot by police, or how a state legalizing cannabis affects crime rates or the consumption of alcohol. Policy historians may care about how Pinochet's 1973 economic policies affected the GDP of Chile or about how Britain's National Health Service of 1948 affected infant mortality. Environmental scholars may ask how Hurricane Katrina affected the economy of New Orleans. These are just some fields;

increasingly, empirical methods of analysis are used in the business sector as well as other fields. Given the array of areas and topics, I don't care about what policy you choose. You may study whatever is is 1) quantifiable with **accessible** data and 2) interesting to you. To quote Noam Chomsky (who was quoting another MIT professor), the important part isn't what we cover in class; it is about what we discover.

1.2 Helpful Notes from Me

1. Sun Tzu said every battle is won before it is fought. To reverse the perspective, as Ben Franklin said, if you fail to prepare, prepare to fail.
2. Please, do contact me if you have questions. Policy data analysis is what I do in my research every day. I love what I do, and I love discussing this topic with others. If you have any questions about the ideas we cover in class or have any difficulties, you may always meet with me or contact me otherwise.
3. Main Takeaways: In addition to statistics, the main goal of the course is to provide the reasoning skills scientists use to understand the world better. These skills will be useful not just in academia or even the professional workforce, but every day life.

1.3 Additional Requirements

1. If I feel the concept is important, it'll be in the lecture notes or we will discuss it. I will also assign external readings to be done before class.

2. There is no required textbook (aside from this one!) for this course. Various free textbooks exist such as Introductory Econometrics with R, Introductory Statistics, Intro to Modern Statistics, Regression and Other Stories, INtro to Econometrics, Intro to Political Science Research Methods, and many others. The Policy Department at Georgia State also recommends Introduction to Research Methods or Research Methods for the Social Sciences. I will recommend chapters from each book to read (of course, you need only read one chapter, from either text). The corresponding lecture will focus on the content that each respective chapter covers. Note that these books cover different aspects of the course in different levels of depth (Gelman's book *Regression and Other Stories* is obviously mainly about regression, one of the last topics we cover, whereas the others are more rudimentary).

3. The same is true for software– I don't care which of these you use, but the only ones I know well are Stata, Python, and R. For Stata users, Statalist is a great resource for Stata. R also is backed by a vast statistician community.

2 To Do

1. Precise Scheduling (e.g, chapters to read, dates, etc under development)

3 How Data Can Be Used For Policy Studies

3.1 What is This Thing Called Science?

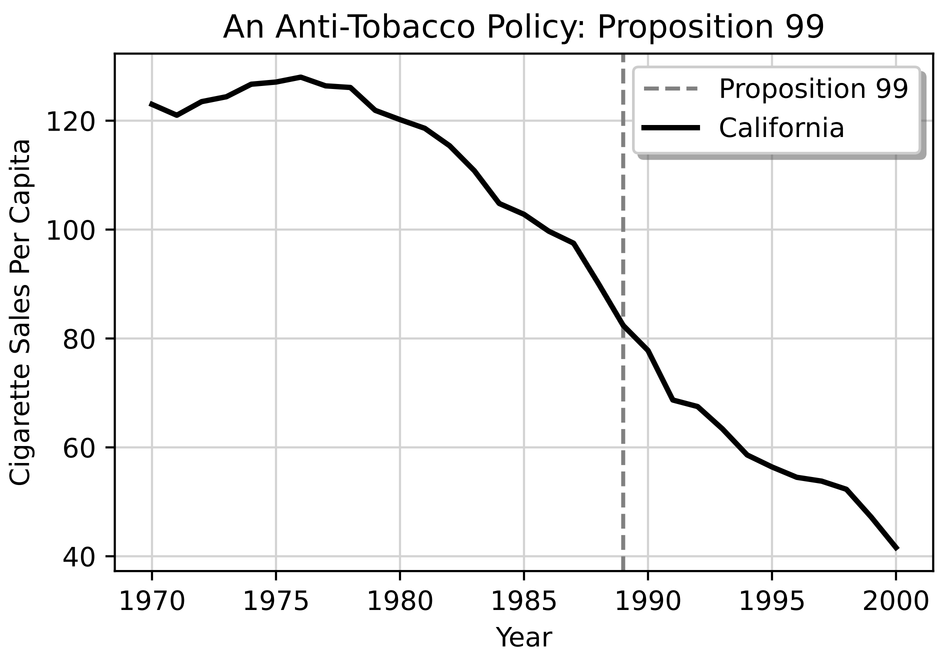
Science at its core is a process we use to understand observable phenonmena. It is based on using logic and observations of the senses to form coherent and simple understandings about the world. Data, or a collection of observations, is fundamental to being able to conduct scientific research. We use data in our daily lives to make conclusions; we don't call it as such, but we do. Note here that data is not a living, breating concept: it requires interpretation by us. We use principles of science to interpret data and the analyses we conduct upon data. As we learn in middle and high school, science typically begins with asking questions or defining a problem.

Suppose our current problem involves commute time to school or work, and we don't wish to walk. In this case, that's our question: "What's the ideal way to get to school/work?" We then gather information. Chances are we may use Google Maps or Waze to guide us. In this context, these tools provide us with the information we need, namely, *estimates* of how long our commute will be. And, assuming we wish to get to our destination as fast as possible, we make *inferences* or conclusions about the ideal way to take based on the GPS' options. If GPS says the highway takes 15 minutes but the backstreets which avoid highways take 35 minutes, we will typically elect to use the highway since that takes us to our destination the quickest.

There's still two more steps to do, though: test our hypothesis and draw conclusions about the actual observed facts. This means that we must, in real life, leave home and take the way we decide to take. When we get to our destination, we form conclusions about how actually taking the highway went. Of course, we repeat this idea multiple times; eventually, we "typically" take a certain direction to work or school precisely because we have the expectation the highway way will, on average, be preferable to *alternative* ways. This is a simple example, yet it illustrates the central point: in scientific inquiry, we ask questions, draw on available information, form ideas, take actions based on that information, and draw conclusions or plan accordingly based on testing the validity of that observed information. We don't call this science in daily life, but that's exactly what it is. The steps I've outlined so far are present in every field from public policy to physics, albeit with a little more sophisitcation.

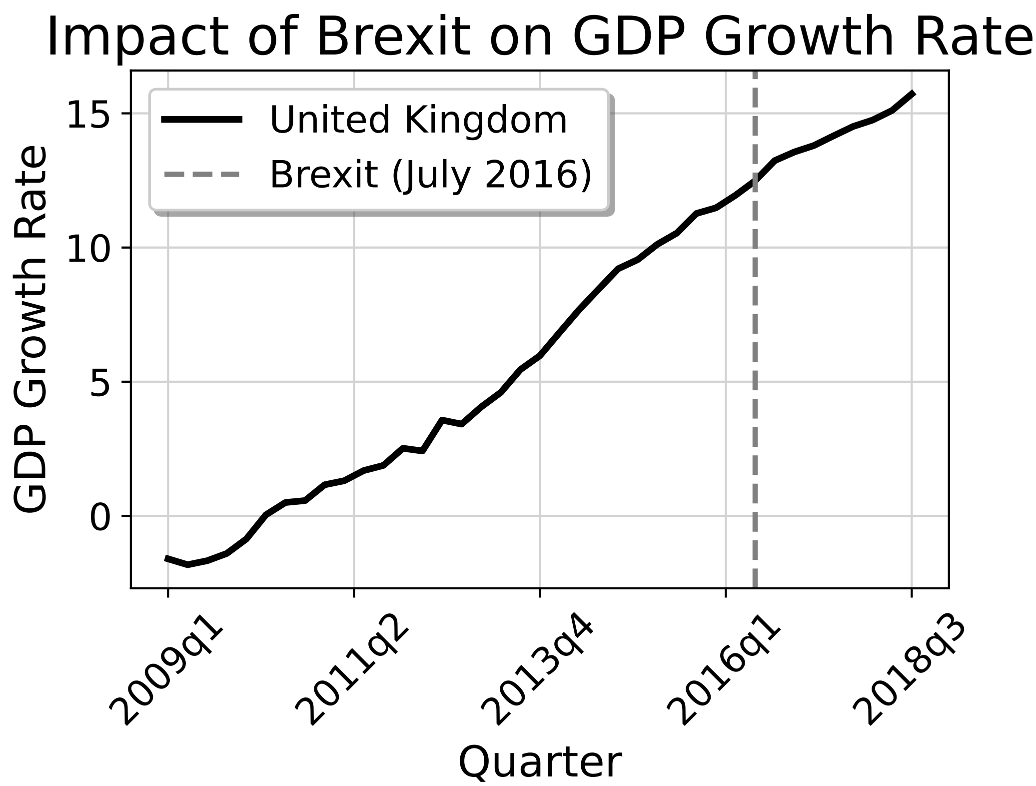
3.2 Data for Policy Analysis

As I've mentioned above, a collection of observations about a set of phenomena is what we call data. Thus, in public policy analysis, data is central to all that we do. One may ask why using data matters at all; the simple reason is that it allows us to resolve disagreements. While people may conduct different data analyses and obtain different results and even reach different conclusions, the main idea is that we can look into the real world and obtain concepts that map on to metrics that we think are important and test them against our expectations. After all, everyone can have opinions or views on things, but the useful part is *testing out* our expectations against reality. That way, we can have a better sense of what's more likely to be true if a certain policy happens/is passed. Indeed, policy analysts are frequently concerned with the impact of some intervention on some outcomes, or sometimes a simple relationship between one variable and another. Recall the plot from the preface



The plot shows the cigarette pack sales per 100,000 for California from the years 1970 to 2000 (our dependent variable). The thick black line denotes California, and the thick dashed line shows the year that Proposition 99 (the independent variable/treatment) was passed. Proposition 99 was an anti-tobacco program that California passed in 1988 and enacted in 1989. It raised the price of tobacco by 25 cents per cigarette per pack and enacted a broader state-wide public health campaign which discouraged people from smoking.

This intervention raises an immediate question for policy analysis: namely, "what was the *effect* of this intervention on the actual smoking rates we see?" This is a question we may collect data on. After data collection (or even prior, in this case), we can form hypotheses. A hyopothesis is a declarative/interrogative, testable statement about the world. It is like a hypothetical in the sense that we try to imagine the effect of a policy on an outcome so that we can answer questions about it. Here, we can hypothesize that Proposition 99 has a *negative* impact on tobacco smoking. Negative here is not intended in the normative sense; presumably most people reading this do not smoke (tobacco, anyways). Instead, here "negative" means that the policy might decrease the tobacco sales per capita compared to what they would have been otherwise. To test this, we can use statistical analysis to compare California to other states that didn't do the policy. We typically wish to produce an estimate of California's cigarette consumption in the years following 1989, had Proposition 99 never been passed. After we do our analyses, we can discuss what the implications are. In other words, was the policy effective by some appreciable margin? Are there other outcomes concerns to consider?). We can do this for other things too.



This is a plot of GDP growth rates for the United Kingdom, taken from this report. The authors were concerned with the impact of Brexit (Britain leaving the European Union) on the GDP in the U.K. economy. The effects of leaving the E.U. is also an empirical question: we can obtain data, as these authors did, about GDP growth rates across a variety of economies/countries. We can also hypothesize about why Brexit may have affected GDP growth. We can compare the U.K. to other areas which didn't have similar economic shocks at the time and try to estimate, using statistics and data analysis, how British GDP would have evolved otherwise.

3.3 4 Steps of Data Analysis

Broadly speaking, we can think of data analysis being broken into 5 distinct concepts. I summarize them below, but these chapters will comprise the rest of the notes

1. Identifying Policy Problems
2. Gathering Data
3. Cleansing Data
4. Analyzing Data
5. Presenting Results

3.3.1 Identifying Policy Problems

As we've discussed above, the first step in this process is simply asking questions. What kind of questions? Policy questions. Knowing what specific questions to ask though can be tricky. Policy is a giant field. Of the thousands of questions we could ask, how do we know which ones will be the most pressing or timely? In other words, how do we know that this is a problem that policy *needs* to be enacted for? How can we identify programs whose analysis benefits the citizenry or other interested parties? Put simpler, who cares? Why do we want to do this study or answer this question? Who stands to benefit?

3.3.2 Gathering Data

Even once we've identified the problem, how do we go about gathering real data to answer questions? If we can't get data that speaks to the issues that we're concerned about, we can't obtain answers that are useful.

3.3.3 Cleansing Data

In real life, datasets do not come to us wrapped in a pretty bow ready for use. Cleaning data (or organizing it) can be a very messy affair in the best of times. In order for us to answer our questions, the data we obtain must be organized in a coherent way such that we can answer questions at all. If your data are not sorted correctly by place and time, trust me, the plot you'll get will not just look terrible, but you can't glean any trends or patterns from it. This and the next three bullets will comprise most of this class's material.

3.3.4 Analyzing Data

Next, we do analysis. We apply statistical analysis in order to answer the questions we're asking, using the dataset we've now cleaned. Such techniques can range from simply descriptive statistical analysis to complex econometric models.

3.3.5 Presenting the Results

Now that we've done analysis, we can finally interpret what the findings mean. We attempt to draw conclusions based on our results and come up with avenues for future research or other elements of interest.

4 Identifying Policy Problems

4.1 Justifications For Policy

Before we can do any analysis though, we have to take a step back. We have to ask ourselves how we know a problem exists in the first place. There are two broad justifications that policy is based on: negative externalities and social good, but the main point of both justifucations is "*harm*".

4.2 Externalities

The idea of externalities comes from microeconomic theory, which says that efficient markets will affect only those parties who willingly participate in transactions. Particularly in the case of negative extrnalities, or externalities which harm others, we could use public policy to rectify this.

Consider a very simple example: seatbelts. In physics, any force that is not stopped by an equal, oppostite force will keep going. So, if you're in a car crash while driving at 60 miles per hour while unbuckled, the car stops. You, however, don't stop: you keep going, 60 miles per hour through the windshield. No public policy is needed just yet. So far, any cost that comes from a transaction has been borne by you, the driver. By the way, I'm not kidding: one of the arguments against seatbelts was literally that using seatbelts should be a personal decision *if* it does not put others at risk. Additionally, industry also argued against mandatory seatbelt laws on the grounds that it was the government interfering between the transactions of a consumer and the seller.

However, there are a few issues with the externality argument. Firstly, being unbuckled turns you into a human projectile. You can hit your passengers or even others outside your vehicle if you're unbuckled. Your market exchange (you buying the car and driving it) is now potentially having second-order effects on others by you not using a seatbelt. So, the government may wish to mandate seatbelts while driving in order to prevent these negative externalities which come in the form of medical bills or death. To address the argument of indsutry above, that seatbelt laws would raise costs of production, this raises an important moral dilemma: does the harm caused to the business of having to install seatbelts matter more than the human harm caused by a society where seatbelts are optional? Also, we are human beings. We have imperfect knowledge. We know for fact that we don't have all the answers, to paraphrase

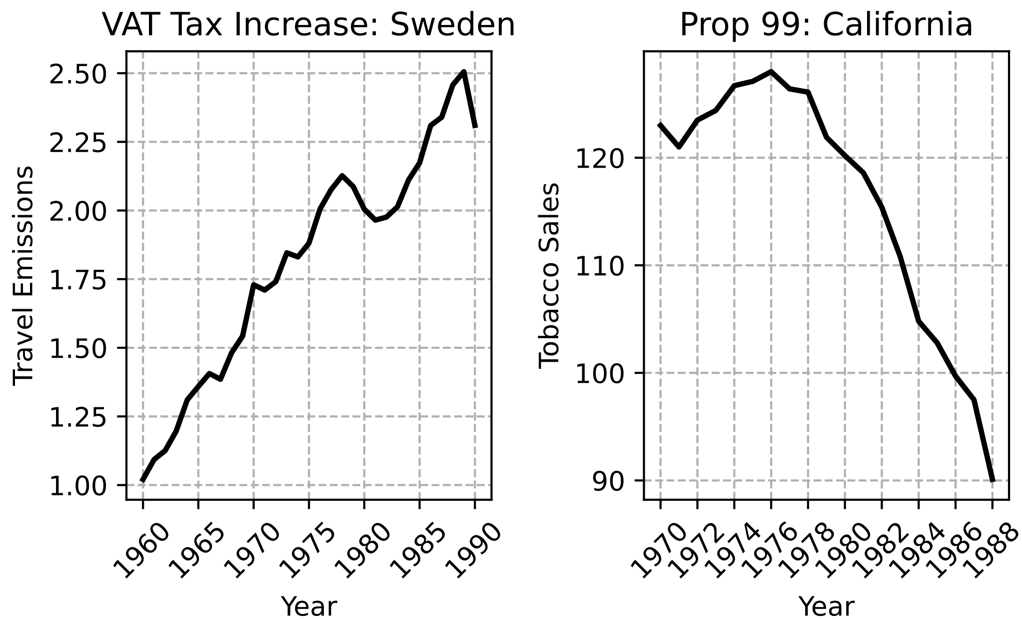
Socrates. We also don't know if the actions we do will ultimately hurt someone else. We live in a probabalistic world (which we will return to later). Indeed, we could argue against laws banning DUI in precisely this manner, saying that we don't know if the intoxicated driver will harm someone until they do. But, as with seatbelts, we never know if there will be another passenger on the road or a child playing in the street. So, we rarely know if we're *actually* putting peoples' lives in danger by driving drunk or unbuckled. We can't know if an externality will occur until it does, usually. Thus, the next view (social good) adopts a different form of reasoning.

4.3 Social Good

Moreover, the externality justification isn't typically the way we think about things from a public policy perspective. Usually, we have social welfare goals in mind. This can come in the form of harm reduction or prevention measures. When we argue for public education, for example, we typically don't do so because we think that the private schools won't educate citizens enough (even though they won't), and that public school will be to decrease inefficient education markets. In fact, we typically don't think of education (in our formative years anyways) as a market at all. We usually argue for public education because we think that education has *inherent* benefits, and that being denied a certain level of education necessitates an inherent harm. Imagine for a moment how the literacy rate of the United States would look if school was completely optional. We likely would not complain about GDP loss, we'd likely complain about a society where lots of people can't read the cereal box or function within society in a decent manner. In other words, society has a vested interest in keeping people safe, educated, and healthy to some degree. So we mandate seatbelt laws, basic schooling, and other laws/regulations in service of these ends. Importantly, "these ends" *does not* have a right or wrong answer. The goals of policy are ultimately decided by people within the society. However, knowing the goals of a policy and reasons for its existence helps us ask meaningful questions about it. Following the above discussion, a natural research question that follows is "How did seatbelt laws affect the rate of car accident injuries and deaths?"

5 Empirical Examples

We can consider two empirical example from the public policy literature (Abadie, Diamond, and Hainmueller 2010; Andersson 2019).

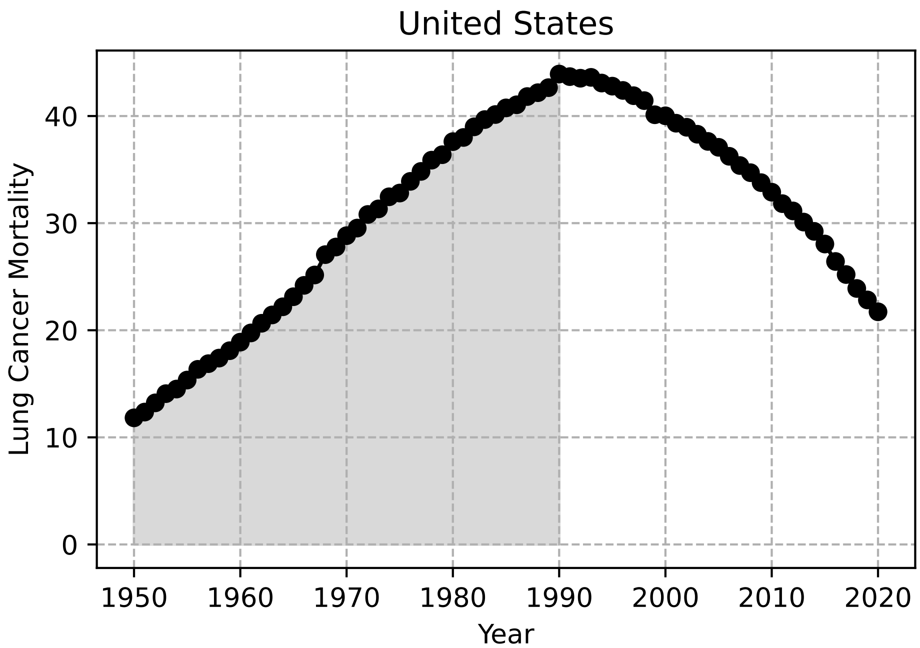


These two plots show the travel emissions trends of Sweden and per capita smoking rates of California, respectively. How might we go about identifying if a problem exists, though?

5.1 Identifying Problems: The Case of Tobacco

As we've discussed above, harm or necessity is typically a standard we look to in order to determine if policy is needed. As I've mentioned, California passed Proposition 99 in 1989 to reduce smoking rates. But, how did we know there was a problem to begin with? To do this, we can grab data on lung cancer morality rates from 1950 until today. Presumably, of course, we view lung cancer as harmful and somehthing we wish to prevent.

The shaded area represents the period before any state-wide anti-tobacco legislation was passed in the United States. We can see quite clearly the age-standarized lung cancer mortality rates



rose in a fairly linear manner in the United States. However, the curve is parabolic; mortality rates were rising every single year until the zenith in 1990. Mortality began to fall when the first large scale anti-tobacco laws were passed. Of course, the *degree* to which these laws were the cause of this decrease is an empirical question. However, given the clear increase in lung cancer rates and other obvious harms of tobacco smoking, policymakers in California and the voters, in fact, became increasingly hostile to tobacco smoking in public and in other crowded areas where a large number of people could be exposed. So, California passed Proposition 99 in 1988 (as did at least a dozen other states from 1988 to 2000, passing similar laws). Just to tie this all together, this plot reinforces why data is needed for policy analysis. Had I not plotted this trend line, people (from the tobacco industry, for example) could simply say "Well, nobody *knows* if lung cancer mortality is a problem. How do we know if there's a problem here? I don't think one exists." A simple plot of data over time can convincingly suggest that a harm is likely being casued by tobacco smoking (which peaked in the U.S. around 1965, but of course lung-cancer occurs more down-the-line). This plot makes a powerful case for how we can *use* data to evaluate if policy problems exist. In other words, data provides intellectual self-defense; if you posit that a problem exists, then this should be demonstrable using datasets that speak to the issue at hand.

5.2 The Swedish Carbon Tax

Follwoing similar arguments above, we can now consdier Sweden. Sweden was one of the first countries in the world to implement a carbon tax (Andersson 2019). The Social Democrats who were in power at the time implemented one, as an addition to the VAT tax on fuel, in 1990 specifically because of climate change. The left panel shows how transport emissions rose in Sweden consistently in the years before 1990, when the tax was passed. In this case, the economic reasoning is simple: the tax increases the price of fuel for consumers, suggesting that less people, on average, will drive as a result, using other forms of transportation. Again, presuming we wish to do demand-side policy to reduce the emissions emitted by a population, a tax is one way of potentially doing this. Much as with the tobacco example above, the plot provides us intellectual self-defense. We can see, for fact, that the emissions were trending upwards in the years before the policy took place.

6 Conclusions

The central takeaway here is that we can use data to idenitfy policy problems by simply looking at relevant datasets which show point to some metric of harm/inequity. Of course, we can also use statistical tools to evaluate good-ness. If a county passes some policy which sees if some local subsidy decreased people not having enough to eat, if we wish to check if a program meant to reduce recidivism actually did decrease it, we identify the effects of policy using existing datasets which we may collect and apply statistical techniques with.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Andersson, Julius J. 2019. "Carbon Taxes and CO2 Emissions: Sweden as a Case Study." *American Economic Journal: Economic Policy* 11 (4): 1–30. https://doi.org/10.1257/pol.20170144.