Empirical Microeconomics: Inferring Causality

University of Alabama

September 13, 2016

Last Class

- We introduced the most commonly used estimator in econometrics: The Ordinary Least Squares estimator.
- We talked about how OLS is used to perform the Classical Linear Regression.
- Under a set of certain assumptions, OLS is the preferred estimator.
- Showed some examples of OLS regressions of ADHD on various different controls including Total Family Income, race, sex, and age.
- It is important to understand that the models we looked at last week presented only correlations.

Inferring Causality

- One thing that distinguishes econometrics from various other fields is the desire to show that x causes y, as opposed to merely showing that x is correlated with y.
- In econometrics, we ask the question "What is the causal relationship of interest?"
- The establishment of a causal relationship can be useful for making predictions about the consequences of changing circumstances or policies; it tells us what would happen in alternative (or "counterfactual") worlds.
- In health economics, for example, we might seek to answer the question "Are people with health insurance healthier?"
- To show a causal mechanism, a researcher must go beyond simple descriptive analysis, and must come up with what is known as an "identification strategy."

The Experimental Ideal

- Ideally, we would like to not rely on studies using observational data at all.
- In an ideal world, we could construct an elaborate experiment that could be used to capture the causal effect of interest.
- What do you imagine the experimental ideal would look like?
- An ideal experiment would clearly identify a "treatment" group and a "control" group, hold all other factors that may confound results constant, and remove any potential sources of bias in its design.
- Researchers have performed health economics experiments in the past, for example the RAND Health Insurance Experiment and the Oregon Health Experiment

Controlled Experiments are Costly

You may wonder why we don't always perform controlled experiments. Here are a few reasons why experiments are difficult:

1. Experiments typically cost a lot of money
   - The RAND HIE was a multi-million dollar project in the 1970s.
2. Experiments take a lot of time.
   - The RAND HIE began in 1971, and finally came to a conclusion in 1986.
3. Economic experiments involve human subjects, which brings about a host of additional difficulties.
4. Experiments can go wrong quite easily.
5. Even under ideal circumstances, some still question the external validity of experiments.
6. It is extremely difficult to mimic the ideal experimental situation.

So what can we do instead of experiments?

Since it is not feasible to perform a controlled experiment to study every hypothesis in the world, we must be a bit clever with statistical inference.

To establish causality, we look for:

1. Natural Experiments
2. Discontinuities
3. Instrumental Variables

What is a Natural Experiment?

- A natural experiment is an empirical study in which the control and treatment variables of interest are not artificially manipulated by researchers but instead are governed by nature or by other forces outside of the researchers' control.
- Natural experiments look very much like controlled experiments in that there is a clearly defined exposure of some condition in a clearly defined population and the absence of that exposure in another similar population for that comparison.
- Under ideal natural experiment conditions, observed outcomes can feasibly be credited to the exposure meaning that there is some cause for belief in a causal relationship as opposed to simple correlation.

Natural Experiments

There are a ton of natural experiments in economics, many the results of some form of a policy change. Here are three examples:

1. The Broad Streed Cholera Outbreak of 1854
   - In 1854, there was a huge outbreak of cholera in London, England.
   - By the end of the outbreak, 616 people living near Broad Street had died.
   - By using a map and noticing that deaths were clustered around a specific public water pump, a physician named John Snow (seriously) realized that the pump was contaminated.
   - In terms of defining Treatment and Control Groups, those that lived near Broad Street were part of the Treatment Group and those that lived far from Broad Street were in the Control Group.

Natural Experiments

2. The Six Month Smoking Ban in Helena, Montana in 2002
   - Smoking was banned in all public spaces for six months.
   - The area was serviced by only one hospital, which saw heart attack deaths fall by 40% during the period.
   - Opponents to the policy successfully got bans removed after six months.
   - This decline in heart attacks is likely not attributable to the bans, however, as heart attack deaths were already on the decline prior to the policy change. Moreover, the population is too small to make any real generalizations.
   - In this experiment, the treatment group might be those that live in Helena, while the control group may be those living in surrounding cities.

Natural Experiments

3. Angrist (1990) studied the effect of Military service on lifetime earnings.

- There is a selection bias issue in this question, as individuals that self-select into the military are likely to be highly motivated on average, and hence likely to earn more.
- The study used the Vietnam War draft lottery as a mechanism of random assignment.
- The draft lottery offers the grounds to randomly separate those drafted from those not drafted into "treatment" and "control groups."
- Angrist (1990) found that the earnings of veterans were about 15% lower than non-veterans.

Can you think of any other "natural" experiments?

Selection and Randomization

- In order for a natural experiment to be ideal, we need to have randomization in the formation of treatment and control groups. This idea of Selection can be best understood by looking at an age-old question, "Do Hospitals make People Healthier?"
- Well, of course hospitals make people healthier, right?
- To make this question more realistic, let's imagine we are studying a poor elderly population that uses hospital emergency rooms for primary care. This sort of care is expensive, crowds hospital facilities, and is, perhaps, not very effective (Grumbach, Keane, and Bindman, 1993). In fact, exposure to other sick patients by those who are themselves vulnerable might have a net negative impact on their health.

Selection and Randomization

Suppose we estimate a linear model given by:

$$HS = \alpha + \beta \mathit{Hospital\ Visit} + \varepsilon$$

Where HS is some measure of health status. Specifically, respondents were asked "Would you say your health in general is excellent, very good, good, fair, or poor?"

The Hospital Visit data is a categorical variable asking "During the past 12 months, were you in a hospital overnight?"

Would you expect the coefficient β to be positive or negative here?

Selection and Randomization

Data coming from the National Health Interview Survey (NHIS) shows the mean differences in health status across the two groups:

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No hospital	90,049	3.93	0.003

- Here, the mean health status of the two groups is significantly different with the hospital "treatment" group suffering a worser health status. A linear regression here will result in a negative β coefficient, indicating that hospitals make people worse off.
- This is an example of Selection Bias, i.e. sick people are self-selecting into hospitals. These people were going to have a worser health status regardless of whether they went to the hospital or not.

Selection and Randomization

- Randomization solves the selection bias problem, as randomization makes Treatment status independent of potential outcomes. In other words, in order to observe the true effects of a hospital stay, people going to the hospital and people not going to the hospital should have no systematic differences.
- Randomization, though very difficult to establish in a natural experiment, ensures that the treatment and control groups are formed randomly and not due to preexisting differences.
- Randomization can be best understood by looking at an article titled "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns" (Doyle et al., 2015), published in *Journal of Political Economy.*

Selection and Randomization

(Doyle et al., 2015) study whether hospitals that spend more money actually achieve better outcomes.

Think of those that go to high spending hospitals as parts of the "Treatment" group and those that go to low-spending hospitals the "Control" group.

Can you see the selection bias issue with this particular topic?

(Doyle et al., 2015)

- Uses data from New York City on Medicare claims, patient characteristics, dates of death, and ambulance identifiers and patient locations.

- Is able to achieve randomization through heterogeneous ambulance preferences.

- To explain, if a person is very ill, maybe they are more likely to self-select into the hospital that is known to be the "best." This is textbook Selection Bias.

Selection and Randomization

(Doyle et al., 2015)

- In emergency situations, however, selection bias may be less of a problem, particularly in a large city such as New York.
- New York has many hospitals, both public and private, and many competing ambulance companies, both public and private.
- When an ambulance picks up a patient in an emergency, the delivery to a hospital is in some sense random, perhaps depending on ambulance company preferences, proximity to a hospital, etc.
- They are able to show that ambulance delivery effectively randomizes selection, hence bettering the selection bias problem, and they conclude that increased hospital spending by 10% lowers 1-year mortality by 2.4 percentage points.

Common Identification Strategies in Health

Typically, natural experiments offer the testing grounds to infer causality, but the exact estimation strategy employed can differ across situations. Four different strategies are quite common:

1. Fixed Effects Estimation
2. Difference-in-Differences
3. Instrumental Variables
4. Regression Discontinuity Design

Fixed Effects Estimation

- Often times, we are concerned that selection into treatment or control groups is non-random, such as the case with returns to hospital spending. Moreover, we are concerned with "unobservables" that we are unable to control for in a regression model.
- One technique to attempt to control for these unobservables is known as Fixed Effects Estimation. In its simplest sense, fixed effects can be included into a regression model by simply including a set of dummy variables for the desired level of the effect.
- For example, if using a panel framework, i.e. a regression model including multiple time periods, one might want to include Time Fixed Effects in an attempt to control for unobserved differences that vary across time.

Fixed Effects Estimation

- Similarly, one might be concerned of unobservables that do not vary across time, but do vary across geography. Then, the inclusion of state or county fixed effects may be appropriate (depending on the level of the data or the policy change of interest).
- Even finer of a measure, however somewhat more difficult to interpret intuitively, are fixed effects that control for unobservables that vary within individual people. Examples of these include Mother fixed effects, Child fixed effects, or even Sibling fixed effects.
- Though difficult because of the need of very rich data, sometimes it may be appropriate to include multiple types of fixed effects. For example, we might include Mother, State, and Time fixed effects.

Fixed Effects Estimation

Let's go back to the regression we ran last week, with ADHD diagnosis as the dependent variable. What happens when we include various different fixed effects into the specification?

Recall that our previous specification was:

$$ADHD = \alpha + \beta X + \varepsilon$$

where X was a matrix of control variables including total family income, age, dummy variables for male, black, white, female-headed households, college educated mothers, birth order, and low birth weight.

Let's replace the continuous age variable with categories of age for Junior High and High School aged children and let's remove birth order and low birth weight which were insignificant from last time and instead include a dummy for having two or more siblings.

Fixed Effects Estimation

- Also, let's move away from simple descriptive analysis and move in the direction of inferring causality. ADHD diagnosis is quite subjective, and perhaps diagnosis can be influenced by exogeneous policy change. Specifically, let's consider a state-level policy that provides financial rewards to teachers and administrators at high-achieving schools.
- How would an education policy providing rewards to teachers influence ADHD diagnosis? A study by Sax and Kautz (2003) showed that about half of childhood ADHD diagnoses are first suggested by the child's teacher at school. If teachers have a new incentive to get students to perform better, then perhaps they will become more likely to recommend ADHD diagnosis to a child on the margin of diagnosis.
- Hence, policy change offering financial rewards to teachers based on student performance creates the grounds for a natural experiment. We can model this using a FE framework. Is the causal mechanism clear?

Fixed Effects Estimation

Let's add some fixed effects to that model. First, since we have data from three panel years: 1997, 2002, and 2007, it may be appropriate to include a year fixed effect. Moreover, to control for unobservables that vary within states but not across time, lets throw in a state fixed effect. Can we include a mother fixed effect as well, control for unobservables that may be mother-specific? Our new, all inclusive model is:

$$ADHD = \alpha + \gamma Rewards + \beta X + T + S + M + \varepsilon$$

where T denotes a dummy variable for each panel year, S denotes a dummy for each state, and M denotes a dummy for each mother. The coefficient γ tells us the effect of a child living in a "rewards" state.

Fixed Effects Estimation

Table 11: Inclusion of Fixed Effects in ADHD Model

Inclusion of Fixed Effects in ADHD Model					
	(1)	(2)	(3)	(4)	(5)
	adhd	adhd	adhd	adhd	adhd
Rewards	0.031***	0.030***	-0.002	-0.013	-0.018
	(0.009)	(0.009)	(0.015)	(0.015)	(0.016)
TFI	-0.000	-0.001	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
jrhigh	0.035***	0.029***	0.034***	0.031***	0.017
	(0.010)	(0.011)	(0.010)	(0.011)	(0.013)
highschool	0.016	0.002	0.014	0.002	-0.025
	(0.012)	(0.014)	(0.013)	(0.015)	(0.019)
male	0.067***	0.067***	0.066***	0.079***	0.078***
	(0.009)	(0.009)	(0.009)	(0.014)	(0.014)
black	0.012	0.013	0.013	0.015	0.024
	(0.016)	(0.016)	(0.018)	(0.310)	(0.311)
white	0.018	0.019	0.027	0.015	0.024
	(0.016)	(0.016)	(0.017)	(0.083)	(0.084)
femalehead	-0.001	-0.002	0.001	-0.008	-0.012
	(0.011)	(0.011)	(0.011)	(0.019)	(0.019)
mothercollege	-0.023**	-0.022**	-0.017*	-0.004	-0.005
	(0.009)	(0.009)	(0.010)	(0.025)	(0.026)
twosiblings	0.021**	0.019*	0.028***	0.047***	0.038**
	(0.009)	(0.010)	(0.009)	(0.013)	(0.017)
Year FE	No	Yes	No	No	Yes
State FE	No	No	Yes	No	Yes
Mother FE	No	No	No	Yes	Yes
Observations	3321	3321	3321	3321	3321

[a] Results are for years 1997, 2002, and 2007 within the PSID sample.
[b] Standard errors (clustered by state) are in parentheses.
[c] *, **, and *** correspond to significance at the 10%, 5%, and 1% levels, respectively.
[d] States that do not have monotonicity within treatment are dropped from analysis.

Fixed Effects Estimation

- With no fixed effects and with only the inclusion of year fixed effects, the policy variable of interest, Rewards, is positively and significantly related to ADHD diagnosis at the highest possible level.

- The Rewards coefficient indicates that if a child lives in a state issuing rewards to teachers of high achieving schools, the child is about 3% more likely to be diagnosed with ADHD than the reference group.

- With Fixed Effects, there is a trade-off between controlling for observables and allowing the data to speak. In some cases, adding fixed effects will mask a lot of the signal of the data.

- Note that out of all the controls, the controls for male and two or more siblings are the only that remain robust under the inclusion of all three levels of fixed effects. These are said to be highly robust.

Difference-in-Differences

Easily the most commonly used identification strategy in health economics is known as Difference-in-Differences, also known as DD or diff-in-diff. A DD methodology attempts to mimic the setup up a natural experiment by separating observations into two groups: Treatment and Control, and comparing average outcomes across time within these two groups. An example of a DD model within a panel econometrics frameworks may be:

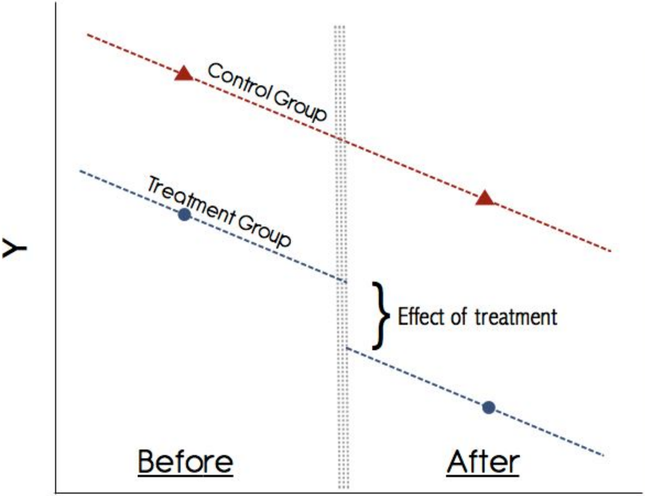
$$y_{it} = \alpha_i + \tau_t + \beta X_{it} + \rho D_{it} + \varepsilon_{it}$$

where y is some outcome variable for individual i in year t, τ is a time fixed effect, X is a matrix of time varying covariates such as total family income, age, etc., D_{it} is the effect of some treatment. Here, the coefficient of interest is ρ. If ρ is significant, then given that we set up the model correctly, we can claim that the treatment has a causal influence on the outcome of interest.

Difference-in-Differences

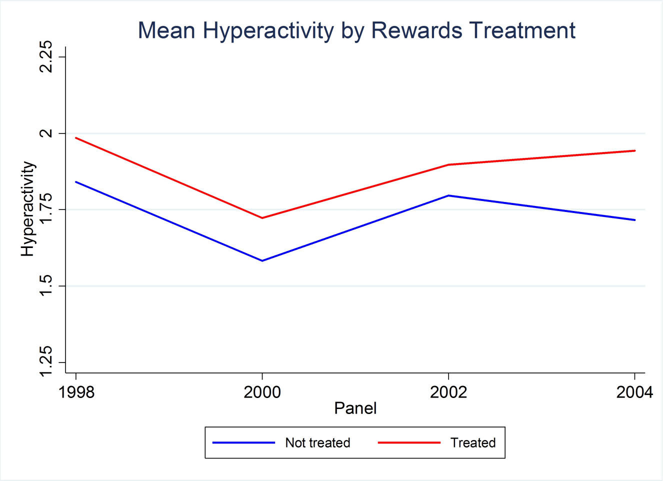
The DD estimator relies on what is called the "Common Trend Assumption." Ideally, we would like to know the difference between a person in a world in which she received the treatment and one in which she does not. Of course, only one of these is observable in practice. Therefore, we look for people that are comparable and that have the same "pre-treatment trends" in the outcome.

Difference-in-Differences



Note that the two groups were trending identically in the pre-treatment period.

Difference-in-Differences



Difference-in-Differences

Consider a DD model in the context of ADHD diagnosis and state-level school accountability policy may be of the form:

$$ADHD_{ist} = \alpha_i + \gamma PreTreat + \rho D_{ist} + \beta X_{ist} + \varepsilon_{ist}$$

where the dependent variable is ADHD diagnosis of individual i, living in state s, in year t, X is a matrix of time varying covariates for individual i living in state s, and D is the treatment status of the state in which individual i resides in time period t.

In other words, our treatment group consists of individuals that reside in states that offer rewards to teachers, and our control group consists of those that live in states that do not offer financial rewards.

Difference-in-Differences

adhd	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
PreTreat	.0172454	.0136389	1.26	0.206	-.0095007	.0439915
PostTreat	.0264713	.0134076	1.97	0.048	.0001786	.0527639
TFI	-.0002253	.0010147	-0.22	0.824	-.0022151	.0017644
jrhigh	.0318943	.011019	2.89	0.004	.0102859	.0535027
highschool	.0026403	.016679	0.16	0.874	-.0300675	.0353481
male	.0440579	.0095028	4.64	0.000	.0254227	.062693
black	.007532	.0177401	0.42	0.671	-.0272567	.0423208
white	.0109952	.0174267	0.63	0.528	-.0231788	.0451693
femalehead	.0063167	.0118	0.54	0.592	-.0168235	.0294568
mothercollege	-.0163686	.0102004	-1.60	0.109	-.0363719	.0036347
twosiblings	.0209086	.0099128	2.11	0.035	.0014693	.0403479
_cons	-.0060182	.0218137	-0.28	0.783	-.0487952	.0367588

Note that the pre-treatment dummy is insignificant, while the post-treatment control is positive and significant at the 95% level.

Here, the Treatment Effect of living in a rewards state is that children are 2.6% more likely to be diagnosed with ADHD than those living in a non-rewards state.

Instrumental Variables

Another common identification strategy in health economics is known as Instrumental Variables (IV) Estimation. Intuitively, IV is used when the correlation between the explanatory variable of interest and the dependent variable does not plausibly reflect the causal relationship between the two.

An IV is used when the explanatory variable is suspected to be correlated with the error term (i.e. the explanatory variable of interest is endogenous). when an explanatory variable is endogenous, OLS produces biased estimates of the relationship (recall the assumptions of OLS). If an instrument is available, however, consistent estimates of the causal relationship can still be obtained. An instrument is a variable that does not itself belong in the explanatory equation but is correlated with the endogenous explanatory variable.

Instrumental Variables

There are two main requirements for using an IV:

1. The instrument must be correlated with the endogenous explanatory variable. If the relationship is not strong enough, the instrument is said to be "weak."
2. The instrument cannot be correlated with unobservables within the error term. In other words, the instrument cannot suffer from the same endogeneity as the endogeneous explanatory variable. If this condition is met, the IV is said to meet the "exclusion restriction"

Instrumental Variables

Though IV may seem confusing, it is actually pretty intuitive. Suppose we want to estimate the effect of economic growth on the formation of democratic nations:

$$Democratic = \alpha + \beta Income_{pc} + \varepsilon$$

where the dependent variable is whether a country is a democracy, and the explanatory variable is a proxy for economic growth, per-capita Income.

A potential problem with this estimation is that yes, perhaps as people become more well-off in a nation, they may become more likely to rebel against a controlling government and force democracy. Alternatively, however, it may be the case that a country being democratic actually increases per capita income. This is textbook simultaneity, i.e. reverse causality. The $Income_{pc}$ explanatory variable is potentially endogeneous.

Instrumental Variables

To control for this endogeneity, we need an IV, such as the one used by (Lundberg et al, 2016) -forthcoming in *Journal of Applied Econometrics.*

$$Democratic = \alpha + \beta Income_{pc} + \varepsilon$$

To estimate the above relationship in a set of Sub-Saharan African countries, (Lundberg et al., 2016) use rainfall as an instrument for income. Does rainfall meet the two requirements for an IV?:

1. Is rainfall highly correlated with the endogeneous explanatory variable? Yes it is, in Sub-Saharan Africa. In these nations, agriculture is the number one source of income. Hence, if they have larger amounts of rainfall, it has been shown that they obtain greater income.
2. Is rainfall uncorrelated with unobservables, or in this case with the probability of a nation being a democracy? Yes, likelihood of democracy should not be influenced by rainfall.

Instrumental Variables

Once an instrument is established to be valid, estimation via Two-Stage Least Squares (2SLS) may occur. In 2SLS, two stages of estimation are performed. Suppose we have the model:

$$y = \alpha + \beta X + \varepsilon$$

where the explanatory variable X is endogeneous.

Further suppose that we find a valid instrumental variable Z. 2SLS can be performed by doing the following:

1. First stage: Regress X on Z, i.e. run the regression $x = \alpha + \beta Z + \varepsilon$, and obtain the fitted value $\hat{X}$
2. Second stage: Plug the fitted values $\hat{X}$ into the original regression equation in place of X, i.e. run the regression $y = \alpha + \beta \hat{X} + \varepsilon$

Instrumental Variables

What if we wanted to estimate the effect of a family having a child with ADHD, and we looked at dependent variables including labor market outcomes of parents and relationship dissolution outcomes of parents. (Kvist et al., 2013) do exactly this using a sample of Danish families. Their model of labor in its simplest form is:

$$\textit{Hours worked} = \alpha + \beta \textit{Child ADHD} + \varepsilon$$

where here the dependent variable is number of hours worked per week, and the explanatory variable is whether the family has a child diagnosed with ADHD. Other outcome variables studied are whether or not the parent is employed, and whether the mother and father dissolved their relationship or divorced.

Instrumental Variables

$$Hours\ worked = \alpha + \beta Child\ ADHD + \varepsilon$$

- There is a potential endogeneity problem with the above analysis. Perhaps the type of mother that will "game the system" and get her child diagnosed with ADHD to potentially gain benefits through Supplemental Security Income is the same type of mother that is lazy and will shirk at the labor market. If this is the case, then Child ADHD is endogeneous, and this will overstate the effects of having a child diagnosed with ADHD on the parents labor market outcomes.

- If we were to perform the above analysis on a sample of American children, perhaps we could use the Rewards policy discussed earlier as an instrument for ADHD diagnosis.

Instrumental Variables

Is whether a state offers financial rewards to teachers of high-achieving schools a valid instrument for childhood ADHD diagnosis? Is it:

1. Highly correlated with the endogenous explanatory variable? According to Bokhari and Schneider (2011) and Chen and Hampton (2016), children that live in states offering rewards to teachers are more likely to be diagnosed with ADHD.

2. Uncorrelated with unobservables affecting the parents labor market outcomes? Yes, it seems plausible that a policy issued by policymakers within state governments should be completely unrelated to unobservables that may affect parents' labor market decisions.

If both of the above conditions hold, then Rewards is a valid IV in this situation, and estimation can be performed by using 2SLS.

Regression Discontinuity Design

- Regression Discontinuity is a pretest-posttest design that elicits causal effects by assigning a cutoff or threshold above or below which a an intervention is assigned.
- In short, RDD is used when there is a distinct cutoff of some kind. One may believe that observations on one side of this cutoff will have differing effects from those on the other side.
- A good example of RDD in an ADHD application is in (Elder, 2010) and (Evans et al., 2010).
- Each study the effect of School Start Date (SSD) on probability of ADHD diagnosis. They hypothesize that children born immediately before the SSD, hence entering kindergarten the least mentally developed of their cohort, are more likely to be diagnosed with ADHD than those born immediately after the cutoff.

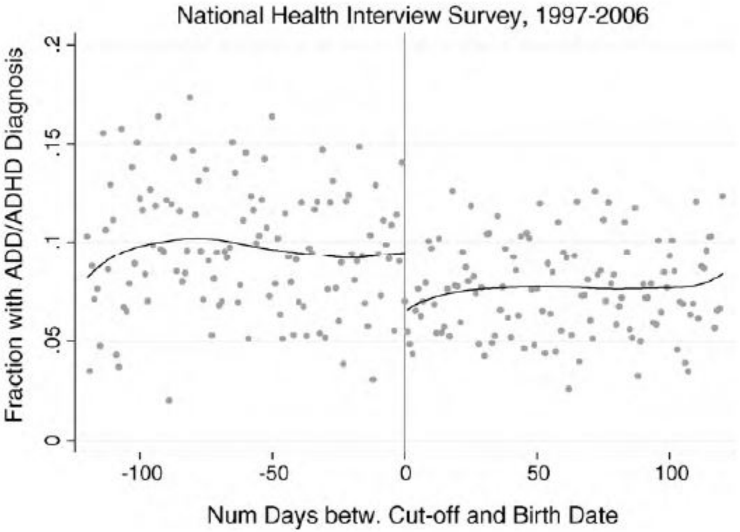
Regression Discontinuity Design

An example of a SSD RDD model may be

$$y = \alpha + \beta_1 X + \beta_2 Young + h(z) + \varepsilon$$

where Young is a dummy variable for whether a child is young relative to his cohort, and h(z) is some smooth function of z, a variable measuring the number of days between a child's birthday and the SSD cutoff.

Regression Discontinuity Design



from (Evans et al., 2010)

Next Class

Insurance Theory (Ch. 8 FGS)