

# Difference in Difference

# Outline

- What does a DID design look like?
  - Simple Difference in Difference
  - Generalized Difference in Difference
- What makes the DID work? (Key Assumptions)
- How can things go wrong? (Threats To Validity)
- Variations, Extensions, and Techniques

What is the simple DID?

# Simple DID Data Structure

- Data varies by
  - state ( $s$ )
  - time ( $t$ )
  - Observed outcome is  $Y_{st}$
- Only two periods  $t = 1, 2$
- Intervention will occur in one group of observations and not in the other group.

# Simple DID

- The simple DID is almost a cliché at this point:
  - 2 Groups
  - 2 Time Periods
  - One group is exposed to treatment between periods.
  - Design can avoid bias from special classes of omitted variables

# The DID Estimator

- The classic DID estimator is the difference between two before – after differences.
  - Before after change observed in the treatment group.
  - Before after change observed in the control group.
- The idea is that the simple pre-post design may be biased because of unobserved factors that affect outcomes and that changed along with the treatment.
- If these unobserved factors also affected the control group, then double differencing can remove the bias and isolate the treatment effect.

# Four Ways To See The Simple DID

- The DID table
- The DID regression
- The DID Graph
- Explicit Model of Omitted Variables

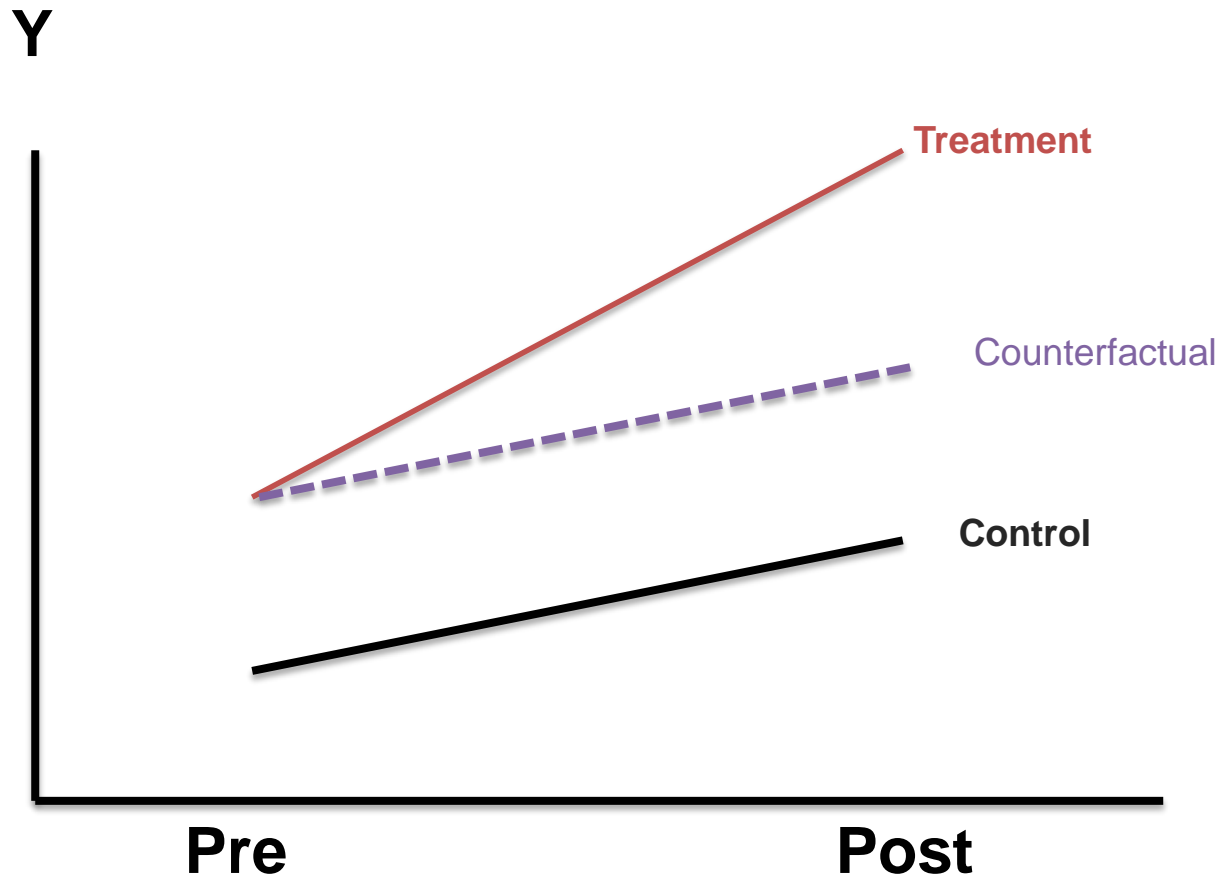
	Pre	Post	Change
Group 1 (Treated)	$Y_{11}$	$Y_{12}$	$\Delta Y_T = (Y_{12} - Y_{11})$
Group 2 (Control)	$Y_{21}$	$Y_{22}$	$\Delta Y_C = (Y_{12} - Y_{11})$
DID			$DID = \Delta Y_T - \Delta Y_C$



# The DID Regression

- $Y_{st} = \beta_0 + \beta_1 \text{Treat}_s + \beta_2 \text{Post}_t + \beta_3 (\text{Treat}_s \times \text{Post}_t) + \varepsilon_{st}$ 
  - $Y_{st}$  is the observed outcome in group  $s$  and period  $t$ .
  - $\text{Treat}_s$  is a dummy variable set to 1 if the observation is from the “treatment” group in either time period.
  - $\text{Post}_t$  is a dummy variable set to 1 if the observation is from the post treatment period in either group.
  - $\beta_3$  is the DID estimate of the treatment effect. It is identical to the double difference in means given in the table.

# The DID Graph



# Examples of Simple DID Designs

# Card and Krueger (1994)

- What is the effect of a higher minimum wage on employment?
  - Data on employment patterns in fast food restaurants.
  - New York and New Jersey are observed before and after New York increased its minimum wage.
  - Famously found that the minimum wage did not reduce employment.

# Card (1990)

- How do immigrant workers affect native wages and employment?
  - Data on wages and employment of low skill native workers.
  - Miami and Comparison Cities are observed before and after the Marial Boatlift in which there was a sudden influx of immigrants that increased Miami labor supply by more than 7%.
  - Found almost no impact of immigration on labor market outcomes in Miami

# Meyer, Viscusi, and Durban (1990)

- What is the effect of Worker's Compensation benefit generosity on return to work times?
- Kentucky and Michigan increased generosity for high income workers. No change for low income workers.
- WC claims data from before and after the change for both types of workers.
- Found that better benefits led to large increases in time off work.

# Some Comments

- Substantive importance of these studies has dwindled over time.
- But these papers from the early 1990s helped launch the quasi-experimental movement in economics.
- DID strategies are very popular for economic policy analysis because:
  - Policies often vary across states.
  - Policies often target specific sub-populations.
  - DID analysis works well with commonly available survey data and also with many types of administrative data.

# Is the DID “just” a version of CITS?

- The regression model and the graph certainly look a lot like the CITS model.
- And – in fact – many people argue that DID is not a design but an estimation strategy used to analyze the CITS.
- I think this is incorrect. DID is not equivalent to CITS.
  - It revolves around different assumptions, employs a different approach to removing bias, and faces different threats to validity.



Time is Not Central To DID

# Madrian (1994)

- Is “Job Lock” a real thing?
  - Job Lock is when you can’t change jobs because you are afraid of losing your health insurance.
- Compare voluntary “job switching/turnover” rates among people with:
  - Employer Provide Health Insurance vs Not
  - Spouse has Insurance vs Spouse Has No Insurance
  - High Health Costs vs Low Health Costs.
- Found that Job Lock reduces job switching by more than 25%.

# Some Comments

- Madrian's study remains substantively important.
  - “guaranteed issue” regulations
  - Need to maintain non-group markets for health insurance.
- But her study also provides a very early example of non-temporal DID study designs.
  - No before vs after comparison.
  - Same logic of grouped/additive omitted variables.

If It's Not About Time...  
Then What Is It About?

# It helps to adopt more formal notation

- Potential Outcomes have the following form:
  - $Y(0)_{st} = \theta_s + \lambda_t + \varepsilon_{st}$
  - $Y(1)_{st} = \beta + \theta_s + \lambda_t + \varepsilon_{st}$
- Treatment Assignment Rule:
  - $T_{st} = 1[t = \text{post}, s = \text{treat}]$
- Observed Outcomes:
  - $Y_{st} = Y(1)_{st} T_{st} + Y(0)_{st} (1 - T_{st})$
  - $Y_{st} = Y(0)_{st} + (Y(1)_{st} - Y(0)_{st}) T_{st}$

# Rewrite in Error Structure Form

- $Y_{st} = Y(0)_{st} + (Y(1)_{st} - Y(0)_{st}) T_{st}$
- $Y_{st} = \{\theta_s + \lambda_t + \varepsilon_{st}\} + [\{\beta + \theta_s + \lambda_t + \varepsilon_{st}\} - \{\theta_s + \lambda_t + \varepsilon_{st}\}] T_{st}$
- $Y_{st} = T_{st}\beta + \theta_s + \lambda_t + \varepsilon_{st}$ 
  - $T_{st}$  is a dummy set to 1 if the treatment is active in group  $s$  at time  $t$ .
  - $\theta_s$  is a group fixed effect; equivalent to the group coefficient in the simple regression.
  - $\lambda_t$  is a time fixed effect; equivalent to the time coefficient in the simple DID regression.

# Equivalency

- In the 2 x 2 Case:
  - $Y_{st} = T_{st}\beta + \theta_s + \lambda_t + \epsilon_{st}$  AND
  - $Y_{st} = \beta_0 + \beta_1 \text{Treat}_s + \beta_2 \text{Post}_t + \beta_3 (\text{Treat}_s \times \text{Post}_t) + \epsilon_{st}$
- Are different ways to write the exact same thing.

	Pre	Post	Change
Group 1 (Treated)	$Y_{11} = \theta_1 + \lambda_1 + \varepsilon_{11}$	$Y_{12} = \beta + \theta_1 + \lambda_2 + \varepsilon_{12}$	$\Delta Y_T = \beta + \lambda_2 - \lambda_1$
Group 2 (Control)	$Y_{21} = \theta_2 + \lambda_1 + \varepsilon_{21}$	$Y_{22} = \theta_2 + \lambda_2 + \varepsilon_{22}$	$\Delta Y_C = \lambda_2 - \lambda_1$
DID			DID = $\beta$



# What Is The Lesson Here?

- DID can eliminate a special class of omitted variables.
  - Omitted Variables that don't change within specified/known sub-groups of the data.
  - The omitted variables have to follow an “additive” structure.
  - Instead of “theta-s” and “Lamda-t”, you could easily have “theta-Medical Procedure” and “Lambda-State”.
  - Any two groupings will do so long as the design suits the research question and accounts for key threats to validity.
  - Still: time variation is very common in applied work.

# Some further points

- There is nothing special about the 2 X 2 structure. You can easily expand to cover cases with:
  - Many groups
  - Many time periods
  - Or other groupings.

# Recap: More Abstract Notation

## How does this help with anything?

1. It helps think about assumptions and identification. (It helps me, at least.)
2. It shows us how to scale up to allow for many groups and periods.
3. It gives us clues for to handle even more exotic version of DID.
4. This is the way that 75% of DID models are presented in papers.

# Generalized DID

# Adding Groups and Time Periods

- How should we describe the DID in a more general setting?
  - Let  $s = 1 \dots S$  and  $t = 1 \dots T$
  - Let  $T_{st} = 1$  if the treatment is in place in state  $s$  and year  $t$ ;  $T_{st} = 0$  otherwise.
- In this case:
  - The table doesn't work well because there are many dimensions.
  - The conventional Treat, Post, Treat x Post model also fails.
  - But the grouped error structure form works fine.

# Generalized DID Formula

$$Y_{st} = T_{st}\beta + \theta_s + \lambda_t + \varepsilon_{st}$$

- Here there are  $S \times T$  observations in theory.
  - Sometimes unbalanced so that  $N < S \times T$ . No big deal.
- The treatment variable is also more flexible:
  - Different start time in each state is allowed.
  - Switching (on/off) is allowed.
  - The treatment can be continuous rather than binary.

# Fitting the Model To Data

- Generalized DID turns out to be a “two-way fixed effects” regression model.
  - Easy to estimate these models in standard statistical packages.
  - Formal packages ask you to specify the  $s$  and  $t$  variables and to identify what variables will be used as fixed effects.
  - When in doubt, use OLS regression and include a dummy variable for each level of  $s$  and each level of  $t$ . (Leave one of each out if you want a constant term to appear in the model.)

# What Does It Mean?

- Intuition is the same as before.
  - Assumption is that outcomes in treated and untreated units would “follow a common path” through time in the absence of a treatment effect.
  - It’s ok for the level of the outcomes to differ across groups. (Group Effects).
  - Treatment effects are measured as a widening or narrowing of the gap between treated and control units, that occurs when treated units are actually treated.



# Time is Only A Typical Example

- Groups are often defined by geography and time period. But other examples occur as well:
  - Product categories and geography.
  - Age/Gender Groups and geography or time.
  - Etc
- What matters is that it is credible to deal with “group non-equivalence” through differencing. (Think of the table in which error terms cancel.)

# Simple Extensions

- Control for ‘time varying covariates’ by adding them to the regression model.
  - More precise Standard Errors
  - Perhaps the DID assumptions hold better conditional on covariates.
- Fit models with ‘placebo laws’ to check that there is no effect in periods that don’t indicate a treatment change.

# More Elaborate Extensions

- Allow for time-varying treatment effects.
- Ashenfelter's Dip
- Test the “common trends” assumption using:
  - Group specific trends
  - An additional untreated comparison group (triple differences).

# Allowing for Time Varying Treatment Effects

- Immediate Effect:

$$Y_{st} = T_{st}\beta + \theta_s + \lambda_t + \varepsilon_{st}$$

- Phased In Effect using Lags:

$$Y_{st} = T_{st}\beta_0 + T_{st-1}\beta_1 + \dots + \beta_k T_{st-k} + \theta_s + \lambda_t + \varepsilon_{st}$$

# Ashenfelter's Dip

- In program evaluation studies, there is often a “dip” in earnings or employment levels in the periods leading up to treatment.
  - People lose their jobs shortly before joining the treatment group.
  - People in the comparison group do not.
- A pre-treatment “dip” or “trend” that is peculiar to treated units would create bias in DID studies.
- Problem is broader than job training programs.

# Testing for Pre-Treatment Trends

$$Y_{st} = T_{st}\beta_0 + T_{st+1}\beta_1 + \dots + \beta_k T_{st+m} + \theta_s + \lambda_t + \varepsilon_{st}$$

- Previously, we used a set of lagged treatment indicators to measure the delayed effect of a treatment.
- This time we want to know if there is a “slope change” for units that are “about to become treated”.
- To do that, use Leads of the Treatment instead of Lags of the Treatment.
- Sometimes called a modified “Granger Causality” test

# Testing for Pre-Treatment Differential Trends

- Under the null of no pre-treatment trends, we expect the coefficients on the leads to be zero.
  - Inspect the coefficients on the leads.
  - Very common to plot the coefficients and confidence intervals leading up to and after the onset of treatment. (Event Study Graph)
  - Compute an F-test that the leading coefficients are jointly equal to zero.

# Examples



# Autor (2003)

- Do “unjust dismissal” regulations lead firms to employ more “outsourced and temporary workers”?
- Generalized DID based on the timing of adoption across states and year.
- Finds that these laws explain about 20% of the growth in outsourced and temporary employment rates.
- Nice application of the event study approach on the next slide. Note that the regression coefficients are being plotted on the Y-axis.

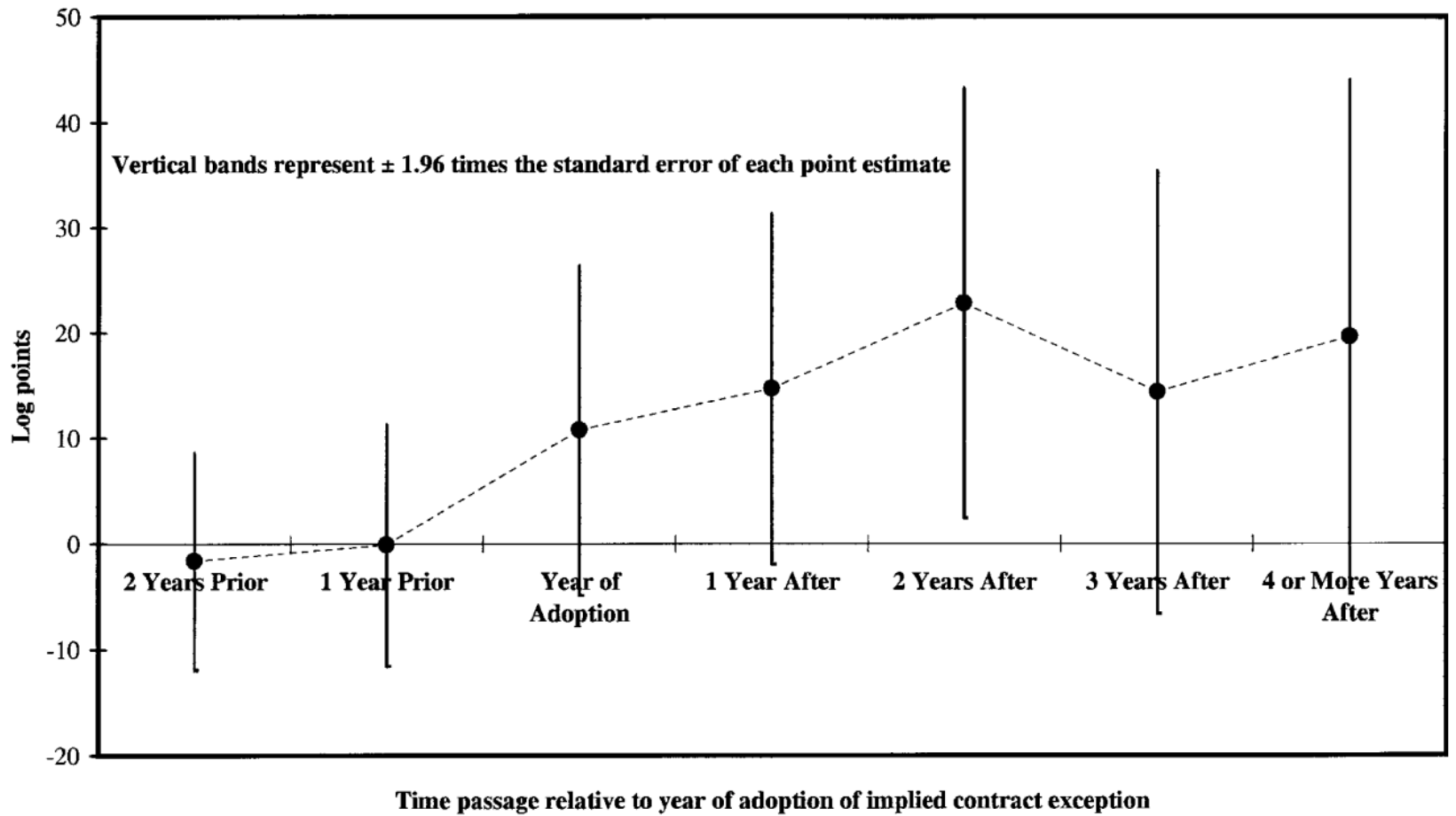


FIG. 3.—Estimated impact of implied contract exception on log state temporary help supply industry employment for years before, during, and after adoption, 1979–95.

# Other Examples of Generalized DID

- EITC and Infant Health (Hoynes et al)
- Hospital Strikes (Gruber and Kleiner)
- Medicaid and Dental Care Utilization (Choi)
- Donor Cycles (Dickert-Conlin, Elder, Moore)
- Occupational Regulation in Dentistry (Wing and Marier)

# Problems With Standard Errors

- The Moulton Problem and how it operates in DID
- The critique by Duflo et al (2002)
- Making SE methods for DID with a small number of clusters is now a sizable academic field of interest.

# Advice about Cluster SE

- Less than 40 clusters means you need to at least include a different approach to statistical inference.
- For 40 – 50 clusters, cluster SE may or may not be sufficient but most people seem to think it is alright as long as data are aggregated to the cell level.
- For 100+ clusters, life is good.

# Alternatives to Cluster SE

- Cluster Bootstrap
  - Block Bootstrap
  - Wild Cluster Bootstrap
- Permutation and Randomization Tests (advanced placebo laws)
  - Rosenbaum
  - Imbens
  - Conley and Taber
- Hierarchical Modeling

# How Do Permutation Tests Work

- The idea is to work out the sampling distribution by fitting the model to “placebo laws”
- Basic Algorithm:
  - Loop over  $P$  permutations
    - Randomly “shuffle” the treatment variable.
    - Fit the DID model using the shuffled treatment
    - Store the DID coefficient
  - End Loop
  - Plot the density function of the placebo DID estimates.
  - Compute p-values and other statistics using the distribution.

# The logic of placebo law tests

- The placebo distribution should be centered on zero because the placebo laws are not real and should not create systematic effects.
- Compare the actual DID estimate to the placebo law distribution.
  - If the real DID estimate is “in the tails” of the placebo distribution, then reject the null of no treatment effect.
  - The idea is that an effect this large would be very unlikely under the null hypothesis.



# Why do people like placebo tests?

- They sound very logical and provide a compelling link between p-values and treatment effects of interest.
- They don't revolve around assumptions about the shape of the error distribution.
- They can usually be adapted to only consider certain types of law generating processes.
  - Only specific states change laws.
  - Law changes always happen in specific years.
  - Etc.

# Wing and Marier (2014)

- Does broadening the legal scope of practice granted to “ancillary” health providers affect health care cost and access?
  - Study what happens to prices and utilization when states allow dental hygienist to legally perform a broader set of dental procedures.
  - Compare prices for different procedures across different states and years.
  - Find that broader scope of practice reduces costs and increases dental visits.

# Simplified Example: You could do it many different ways

Product	Dentist States	Hygienist States	Task Characteristic
Fillings			Riskier
Teeth Cleaning			Low Risk

Product	Before	After	Task Characteristic
Fillings			Riskier
Teeth Cleaning			Low Risk

State	Before	After	Task Characteristic
Dentist State			na
Hygienist State			na

# A full scale panel makes it easier...

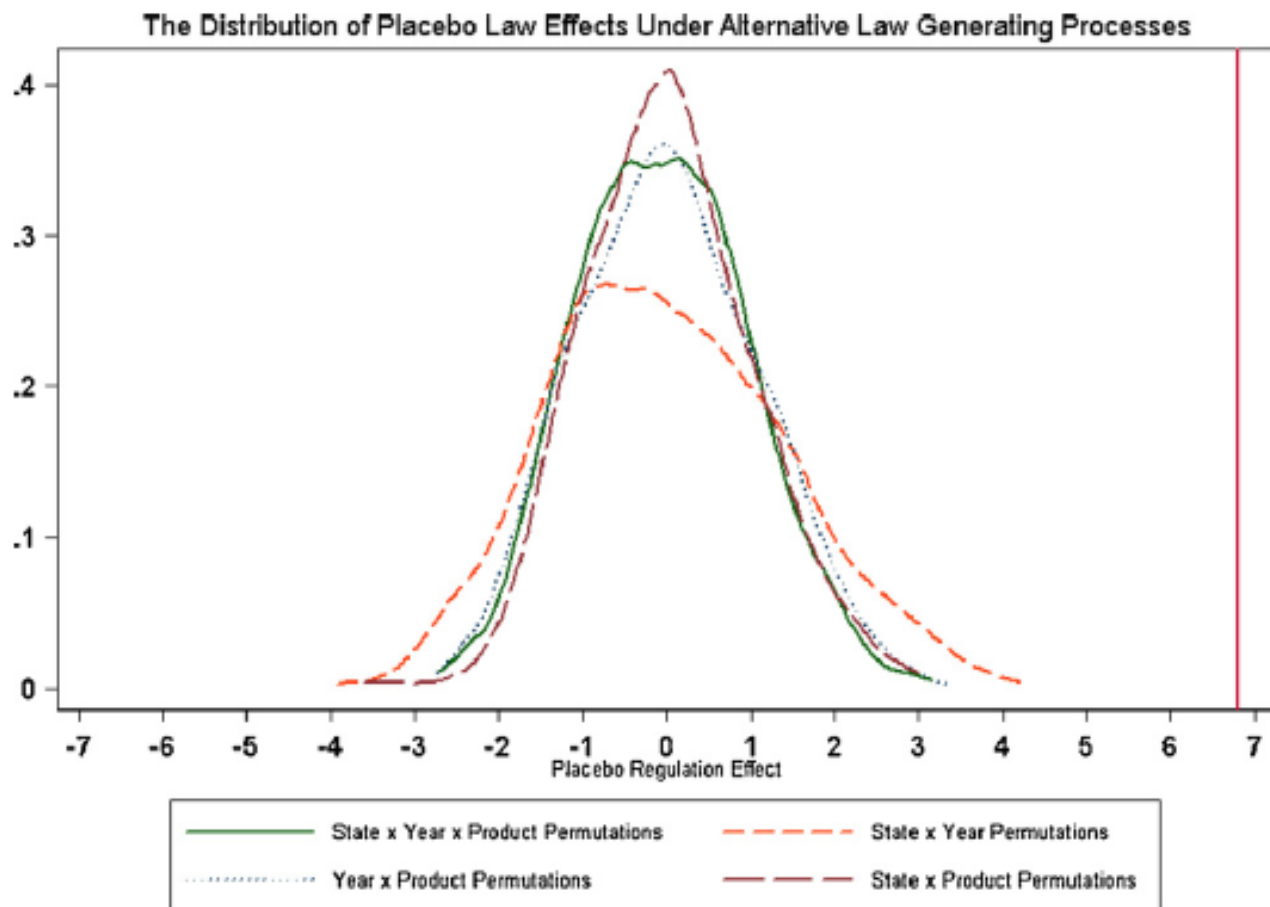
- Include state effects, year effects, procedure effects, state x year effects, procedure x year effects, procedure x state effects
- Triple Differencing with high dimension

**Table 4**

Effects of licensing regulations on the log of cell median prices.

Statistic	1	2	3	4	5
Regulation effect	0.12	0.12	0.12	0.12	0.07
SE	0.03	0.03	0.03	0.03	0.03
<i>p</i>	<i>p</i> < .01	<i>p</i> < .01	<i>p</i> < .01	<i>p</i> < .01	<i>p</i> < .01
Covariates	No	Yes	Yes	Yes	Yes
Product fixed effects	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
State × year fixed effects	No	No	Yes	Yes	Yes
Product × year fixed effects	No	No	No	Yes	Yes
Product × state fixed effects	No	No	No	No	Yes
<i>R</i> squared	0.645	0.646	0.650	0.651	0.695
<i>N</i>	2979	2979	2979	2979	2979

*Note:* Standard errors are calculated using a Huber–White robust variance matrix that allows for clustering at the state level.



**Fig. 4.** The graph shows a kernel density plot of the distribution of 2000 placebo estimates of the effects of the regulation on cell level prices. The different lines show the distributions that arise from different law generating procedures. Each regression allowed for state, year, and product fixed effects. The estimate observed in the actual data is indicated by the vertical red line at about \$6.80. The graph shows that an estimate as large as \$6.80 is extremely unlikely to have arisen by chance. The placebo law distribution provides a way to conduct statistical inference under less restrictive assumptions than standard methods such as cluster-robust standard errors.

# Balancing Tests

- Trick is to assess balance in a way that takes the design into account.
- Compare balance after accounting for the fixed effects structure using a regression based approach.

# Wing and McConeghy (2015)

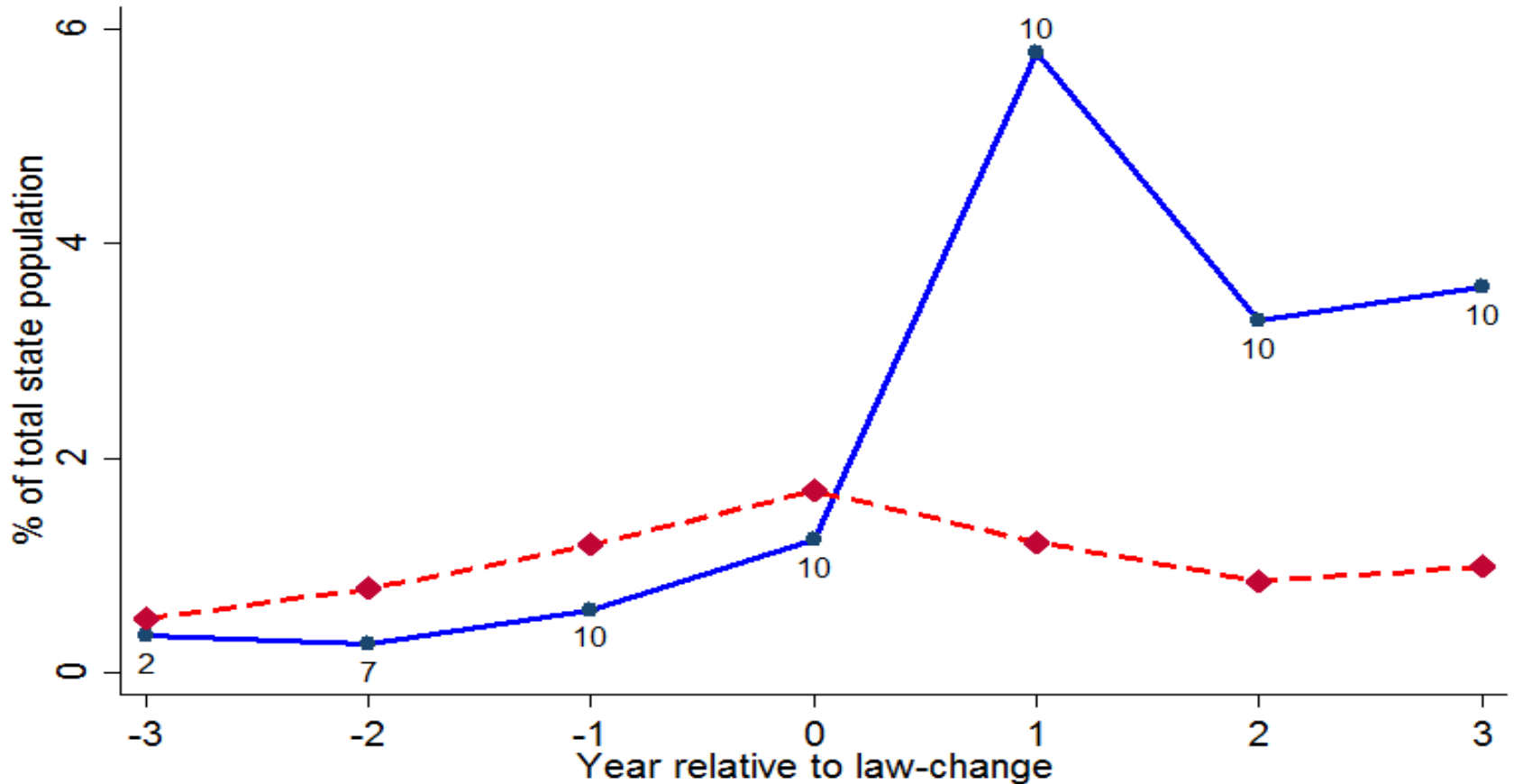
- How do pharmacy based vaccination laws affect influenza vaccination rates?
- DID based on state law changes.
- Were the data balanced?



# Balancing Tests

<b>Demographic characteristics</b>				
<b>Variable</b>	<b>Flexible PBI (Treated)</b>	<b>Strict-PBI (Control)</b>	<b>DID</b>	<b>P-value</b>
Age (years)	53.0	50.6	0.10	0.35
Gender (%)	38.5	38.8	0.00	0.26
Respondent is parent (%)	31.9	33.8	0.00	0.54
Respondent had health insurance plan (%)	88.7	87.7	0.00	0.55
Currently employed (%)	55.1	56.7	0.01	0.05
Income $\geq$ 75,000\$ (%)	24.1	19	0.01	0.16
Currently married (%)	56	54	0.00	0.55
Black, non-hispanic (%)	6.9	9.5	0.00	0.89
Hispanic (%)	6.6	8.8	-0.01	0.20
Education level > high-school (%)	60	55.9	0.01	0.06
Current smoker (%)	18.2	21.2	0.00	0.30
Respondent has diabetes (%)	10.1	9	0.00	0.51

# Event Study Plot



Marker labels designate number of states which contribute to the estimates

# Conclusions and Discussion Points

- DID turns on the idea of group level unobserved factors that are “invariant” at known levels of aggregation.
- This is an abstract way to think about something simple.
- You make progress by attaching specific names to the confounders you need to avoid.

# Naming the Confounders in the Dental Paper

$$Y_{spt} = T_{spt}\beta + \theta_s + \theta_t + \theta_p + \theta_{st} + \theta_{pt} + \theta_{sp} + \varepsilon_{spt}$$

- What do the different thetas mean?
  - Product Effect is a measure of any features of product risk or difficulty that could be correlated with both regulations and prices.
  - What about the others?