

Lecture Note 13

Doing Differences-in-Differences (and Event-Study Designs)

1 Oh No, Not Again!

This spring, we saw a garden-variety bank run, something that seemed unlikely in modern times. Worried about the falling value of long-dated low-interest US Treasury securities, [Silicon Valley Bank's \(SVB\) depositors began to get nervous](#). When interest rates rise, low-yield bonds are worth less. That's ok for bondholders who plan to hold 'em to maturity. But SVB had to sell a large chunk of its vast bond holdings to meet its obligations – and fast! With its bond assets substantially devalued and its investors spooked, SVB failed to raise the cash needed to cover customers' increasingly panicked withdrawals. And so, SVB went bust.

Bank runs have long been a regrettable feature of financial life. On the eve of the Great Depression, for instance, Caldwell and Company was the largest Southern banking chain. Alas, in November 1930, mismanagement and the October 1929 stock market crash brought the Caldwell empire down. Within days, Caldwell's collapse felled closely tied banking networks in Tennessee, Arkansas, Illinois, and North Carolina, precipitating a run on Mississippi banks in December 1930.

Policymakers facing a bank run have a choice: open the flow of credit or turn off the tap. On one hand, lending to troubled banks may allow them to meet increasingly urgent withdrawal demands and stave off depositor panic. On the other hand, support for bad banks raises the specter of moral hazard. If bankers know that the central bank will lend cheaply during a crisis, they needn't take care to avoid crises in the first place. Which strategy is better?

[Richardson and Troost \(2009\)](#) tackle this question by exploiting the differential response of regional Federal Reserve Banks to the Caldwell collapse:

- The 6th (Atlanta) District favored lending to troubled banks and increased bank lending by about 40% within 4 weeks of Caldwell's collapse
- The 8th (St. Louis) District decreased bank lending by 10% in the same period
- As it happens, the 6th/8th District border runs smack through the middle of Miss., so we get a well-controlled within-state contrast in lending policy
- Think of the 8th as a passive control group, and the 6th as the treatment group, experimenting with crisis lending
- 8 months into the crisis:
 - 132 banks were open in the 8th
 - 121 banks were open in the 6th
 - A deficit of 11 banks in the 6th: *The easy-money treatment effect is negative!*
- Look again - On July 1, 1930 (*before* the Caldwell crisis):
 - 165 banks were open in the 8th
 - 135 banks were open in the 6th

1.1 Parallel worlds

DD uses parallelism to adjust for differences across districts in the pre-treatment period:

- Let $Y_{d,t}$ denote the number of banks observed operating in District d in year t
 - The DD effect of loose money in the 6th (treatment) District during the Caldwell crisis compares post-crisis differences with the baseline difference between the two districts:

$$\begin{aligned}\delta_{DD} &= (Y_{6,1931} - Y_{8,1931}) - (Y_{6,1930} - Y_{8,1930}) \\ &= (121 - 132) - (135 - 165) \\ &= -11 - (-30) = 19.\end{aligned}$$

- Equivalently, DD contrasts the *change* in the number of banks operating in the two districts:

$$\begin{aligned}\delta_{DD} &= (Y_{6,1931} - Y_{6,1930}) - (Y_{8,1931} - Y_{8,1930}) \\ &= (121 - 135) - (132 - 165) \\ &= -14 - (-33) = 19.\end{aligned}$$

- Either way you look at it, DD controls confounding from fixed (and therefore pre-treatment) differences in levels that arise even in the absence of treatment

Parallelism Meets Potentials

- The heart of the DD setup is an *additive model for potential outcomes* in the no-treatment state:

$$Y_{d,t}(0) = \beta_d + \gamma_t. \tag{1}$$

$Y_{d,t}(0)$ is notation for the *potential outcome* describing what happens in district d and period t in the absence of an intervention (defined for all d and t)

- Parameter β_d is called a “district effect”; parameter γ_t is called a “year effect”; as we’ll soon see, these parameters are coefficients on dummy variables for state and year
- Assuming that the causal effect $Y_{d,t}(1) - Y_{d,t}(0)$ is constant:

$$Y_{d,t}(1) = \beta_d + \gamma_t + \delta_{DD}.$$

So,

$$Y_{6,1931} - Y_{6,1930} = (\gamma_{1931} + \delta_{DD}) - \gamma_{1930}$$

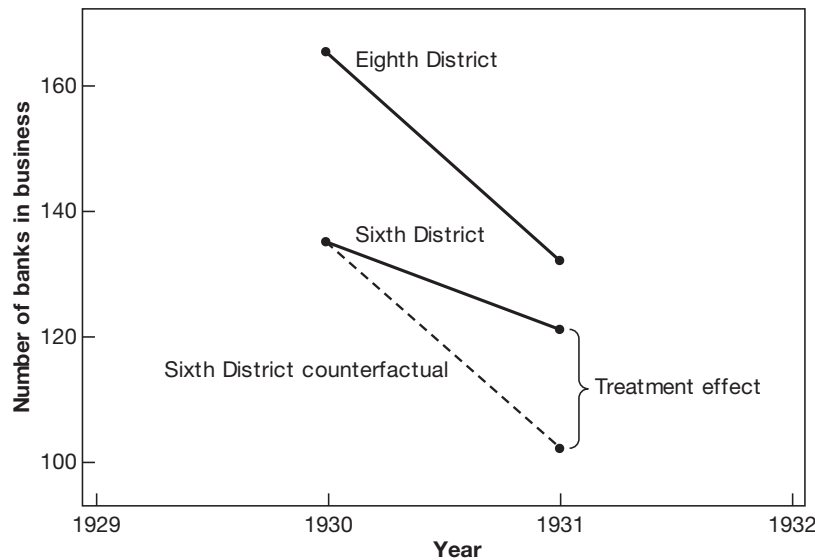
$$Y_{8,1931} - Y_{8,1930} = \gamma_{1931} - \gamma_{1930}$$

- The double-diff (DD) captures causal effect δ :

$$\begin{aligned}&\{Y_{6,1931} - Y_{6,1930}\} - \{Y_{8,1931} - Y_{8,1930}\} \\ &= (Y_{6,1931} - Y_{8,1931}) - (Y_{6,1930} - Y_{8,1930}) = \delta_{DD}\end{aligned}$$

- Every DD picture tells a story

FIGURE 5.1
Bank failures in the Sixth and Eighth Federal Reserve Districts



Notes: This figure shows the number of banks in operation in Mississippi in the Sixth and Eighth Federal Reserve Districts in 1930 and 1931. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district in this period as did in the Eighth.

TABLE 5.1
Wholesale firm failures and sales in 1929 and 1933

	1929	1933	Difference (1933–1929)
Panel A. Number of wholesale firms			
Sixth Federal Reserve District (Atlanta)	783	641	–142
Eighth Federal Reserve District (St. Louis)	930	607	–323
Difference (Sixth–Eighth)	–147	34	181
Panel B. Net wholesale sales (\$ million)			
Sixth District Federal Reserve (Atlanta)	141	60	–81
Eighth District Federal Reserve (St. Louis)	245	83	–162
Difference (Sixth–Eighth)	–104	–23	81

Notes: This table presents a DD analysis of Federal Reserve liquidity effects on the number of wholesale firms and the dollar value of their sales, paralleling the DD analysis of liquidity effects on bank activity in Figure 5.1.

1.2 Common Trends

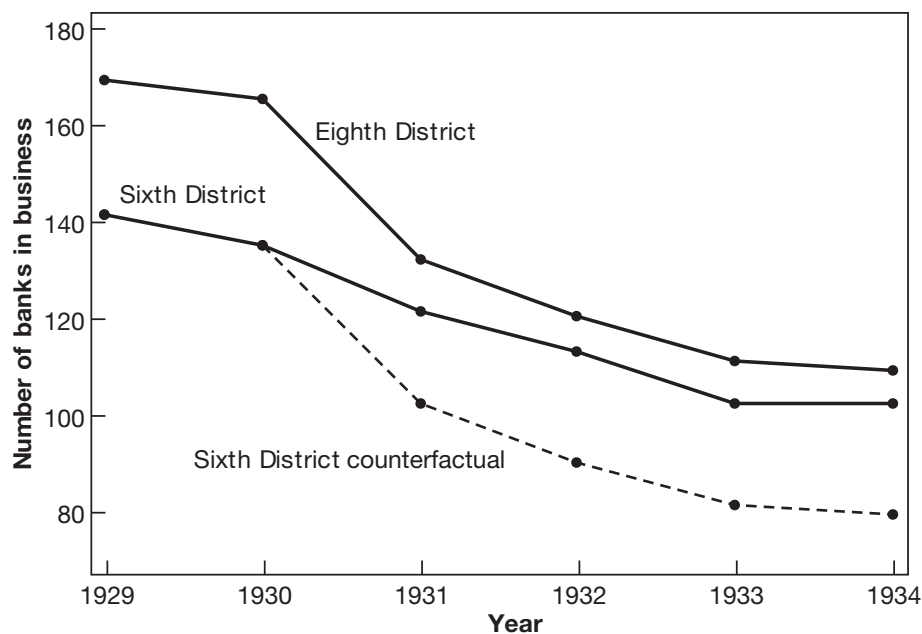
- The DD model for counterfactual no-treatment outcomes in both districts allows for:
 - Time-invariant district effects, β_d
 - Common period (year) effects, γ_t
- The key DD assumption is parallel outcome trends in treatment and control units (districts)
- Common trends can be applied to transformed data, e.g.,

$$\log Y_{d,t}(0) = \beta_d + \gamma_t$$

But common trends in logs does not imply (indeed, contradicts) common trends in levels

- DD identification is fickle
- Luckily, Mis-sis-sippi is DD heaven:

FIGURE 5.3
Trends in bank failures in the Sixth and Eighth Federal Reserve Districts, and the Sixth District's DD counterfactual



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

1.3 Regression DD Heads South

- Stack data on districts and years in a sample of size 12 (with 6 years for each district); call this Y_{dt} for the number of banks operating in district d in year t
- Let $TREAT_d$ indicate data from the 6th District and let $POST_t$ indicate post-treatment years
- The regression DD estimator, δ_{rDD} , comes from fitting:

$$Y_{dt} = \alpha + \beta TREAT_d + \gamma POST_t + \delta_{rDD}(TREAT_d \times POST_t) + e_{dt}$$

The DD treatment effect is the coefficient on an interaction term!

- With two periods, $\alpha = \beta_8 + \gamma_{1930}$; $\beta = \beta_6 - \beta_8$; $\gamma = \gamma_{1931} - \gamma_{1930}$. Regression DD controls for additive year and district effects.
- With two periods, estimates of δ_{DD} and δ_{rDD} coincide (show this). With more, δ_{rDD} is more precise than the simple four-number DD recipe (The data in Fig 5.3 generate an estimate of 21 banks saved, with a standard error of about 11)
- With more than two periods, the regression-DD residual allows for the fact that the additive model fits imperfectly
- Regression DD
 - generates SEs (but beware of serial correlation)
 - facilitates specification testing, as we'll soon see
- Compare DD policy analysis with the fixed effects panel-data estimator we used to estimate the returns to schooling in LN12:
 - Same econometric idea: differencing eliminates unobserved individual effects (here, for districts; in LN12, for twins)
 - Earlier, we analyzed a large sample of microdata; policy DD typically uses aggregate data such as for states and regions (this may complicate statistical inference)

2 Does the Min Matter?

- Last January, Pres. Biden hiked the min to \$15/hour for federal employees/contractors; Dems hope to make this universal (though Amazon beat 'em to it in 2018)
 - Nice work if you can get it! But can you indeed get it – that's the \$30,000 question.
 - What does history teach us?
- On April 1, 1992, New Jersey imposed a state minimum wage of \$5.05. The Federal minimum wage was then \$4.25
 - Card and Krueger (1994) surveyed fast food restaurants in NJ and Eastern PA before this change (February 1992) and after (November 1992).
 - A DD classic!

TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in:		<i>t</i> ^a
	NJ	PA	
1. <i>Distribution of Store Types (percentages):</i>			
a. Burger King	41.1	44.3	−0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	−1.1
e. Company-owned	34.1	35.4	−0.2
2. <i>Means in Wave 1:</i>			
a. FTE employment	20.4 (0.51)	23.3 (1.35)	−2.0
b. Percentage full-time employees	32.8 (1.3)	35.0 (2.7)	−0.7
c. Starting wage	4.61 (0.02)	4.63 (0.04)	−0.4
d. Wage = \$4.25 (percentage)	30.5 (2.5)	32.9 (5.3)	−0.4
e. Price of full meal	3.35 (0.04)	3.04 (0.07)	4.0
f. Hours open (weekday)	14.4 (0.2)	14.5 (0.3)	−0.3
g. Recruiting bonus	23.6 (2.3)	29.1 (5.1)	−1.0
3. <i>Means in Wave 2:</i>			
a. FTE employment	21.0 (0.52)	21.2 (0.94)	−0.2
b. Percentage full-time employees	35.9 (1.4)	30.4 (2.8)	1.8
c. Starting wage	5.08 (0.01)	4.62 (0.04)	10.8
d. Wage = \$4.25 (percentage)	0.0	25.3 (4.9)	—
e. Wage = \$5.05 (percentage)	85.2 (2.0)	1.3 (1.3)	36.1
f. Price of full meal	3.41 (0.04)	3.03 (0.07)	5.0
g. Hours open (weekday)	14.4 (0.2)	14.7 (0.3)	−0.8
h. Recruiting bonus	20.3 (2.3)	23.4 (4.9)	−0.6

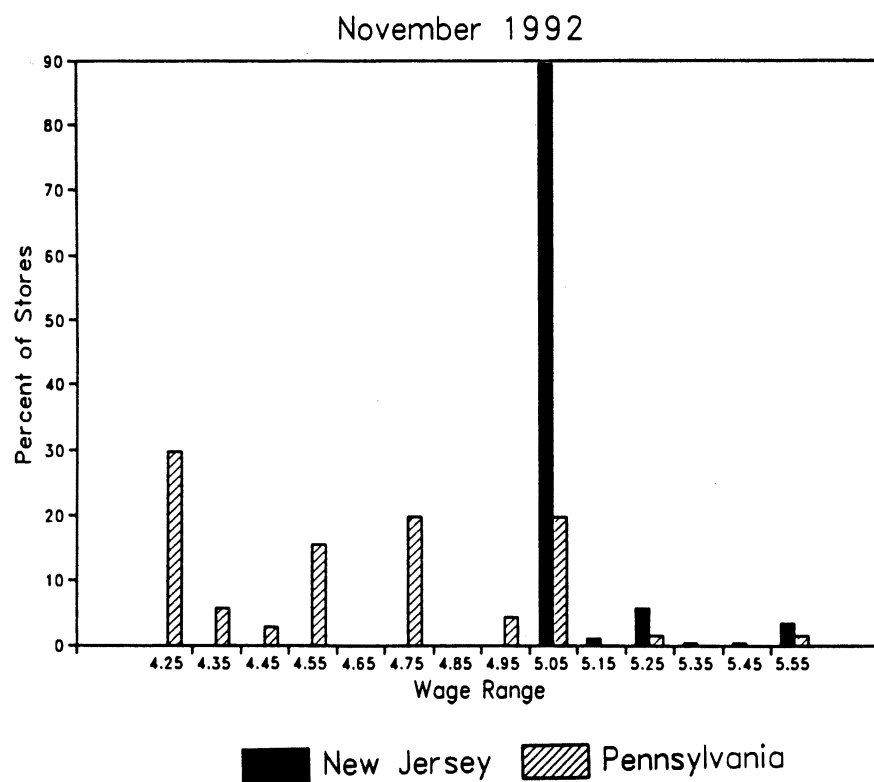
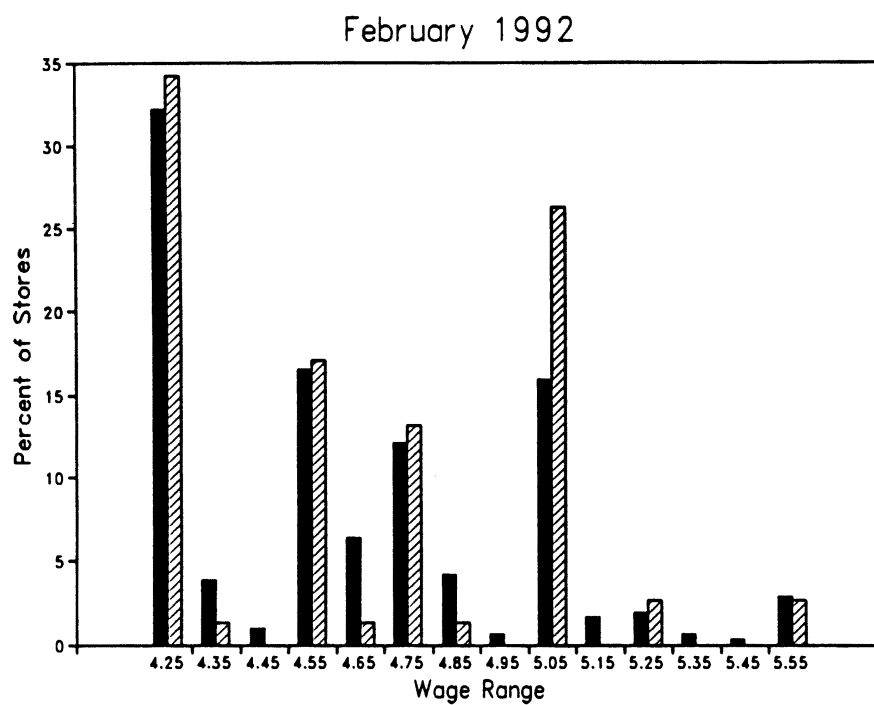


FIGURE 1. DISTRIBUTION OF STARTING WAGE RATES

- As for the within-Miss. liquidity experiment, the simplest DD analysis involves just 4 numbers, repeated in MHE Table 5.2.1

TABLE 5.2.1
Average employment in fast food restaurants before and after the
New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (.51)	–2.89 (1.44)
2. FTE employment after, all available observations	21.17 (.94)	21.03 (.52)	–.14 (1.07)
3. Change in mean FTE employment	–2.16 (1.25)	.59 (.54)	2.76 (1.36)

Notes: Adapted from Card and Krueger (1994), table 3. The table reports average full-time-equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all restaurants with data on employment. Employment at six closed restaurants is set to zero. Employment at four temporarily closed restaurants is treated as missing. Standard errors are reported in parentheses.

Simple DD lives or dies on the common trends assumption

- Alas, reality is uncommonly messy

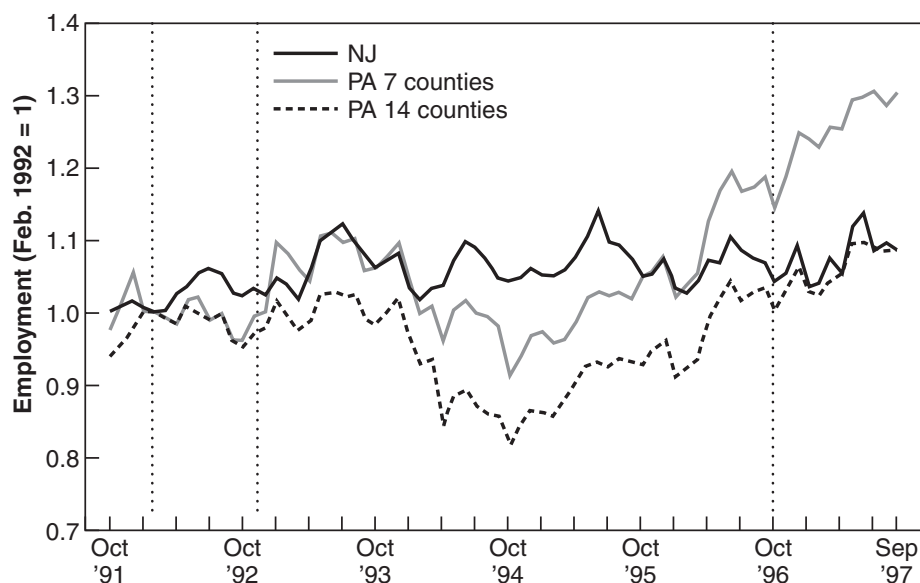


Figure 5.2.2 Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum

- But we can look within states too

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey ^a			Differences within NJ ^b	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low– high (vii)	Midrange– high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores ^c	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 ^d	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

^aStores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

^bDifference in employment between low-wage (\$4.25 per hour) and high-wage ($\geq \$5.00$ per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

3 Drink, Drank, . . . DD With Many States and Years

Alcohol is the most widely abused intoxicant. How should it be regulated? [MADD](#) lobbies to limit access for youth, while college presidents [call to liberalize](#). Who's right?

- State MLDA running from age 18-21 generate up to three treatment effects relative to age 21: the effects of legal drinking at age 18, 19, and 20.
 - We capture this with a single variable called $LEGAL_{st}$, the fraction of 18-20 year olds allowed to drink in state s and year t . In some states, no one under 21 is allowed to drink, while in states with an age-19 MLDA, roughly two thirds of under-21 year olds can drink, and in states with an age-18 MLDA, all 18-21 year olds can drink.
- Because $LEGAL_{st}$ varies by both state and year, we can use a generalized DD model to capture its effects
 - The generalized DD model exploits *within-state variation* in MLDA policy
 - Sketch a state-year panel data structure ...
- Using data on death rates from 1970-83 in the 50 states plus DC (denoted M_{st}), a multi-state regression DD model looks like this:

$$Y_{st} = \alpha + \delta_{rDD}LEGAL_{st} + \sum_{k=AL}^{WY} \beta_k STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j YEAR_{jt} + e_{st}$$

Dummy variables $STATE_{ks}$ switch on when state k in the summation equals state s on the left hand side. State effects, β_k , are the coefficients on these dummies (DC is the reference state). Similarly, the year effects, γ_j , are coefficients on dummies $YEAR_{jt}$ that switch on when year j in the summation equals year t on the left hand side (1970 is the reference year)

- This can be written more compactly as

$$Y_{st} = \beta_s + \gamma_t + \delta_{rDD}LEGAL_{st} + e_{st}, \quad (2)$$

where β_s is again a state effect and γ_t is again a year effect

- Nowadays, equation (2) is said to describe a *two-way fixed effects* (TWFE) model
- The table below reports regression DD estimates of MLDA-induced deaths among 18-20 Year Olds, from 1970 - 1983, a period when stata MLDAs were changing

TABLE 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80 (4.59)	8.47 (5.10)	12.41 (4.60)	9.65 (4.64)
Motor vehicle accidents	7.59 (2.50)	6.64 (2.66)	7.50 (2.27)	6.46 (2.24)
Suicide	.59 (.59)	.47 (.79)	1.49 (.88)	1.26 (.89)
All internal causes	1.33 (1.59)	.08 (1.93)	1.89 (1.78)	1.28 (1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds. The table shows coefficients on the proportion of legal drinkers by state and year from models controlling for state and year effects. The models used to construct the estimates in columns (2) and (4) include state-specific linear time trends. Columns (3) and (4) show weighted least squares estimates, weighting by state population. The sample size is 714. Standard errors are reported in parentheses.

- Because data over time are likely serially correlated; standard errors here cluster on state (see MM Chapter 5 appendix)
- Estimates of δ_{rDD} suggest that legal alcohol access causes about 10 additional deaths among 18-20 year olds, of which about 7 are the result of motor vehicle accidents.
 - The estimated MVA effect is reasonably precise, with a standard error of about 2.5.
 - Regression DD generates little evidence of an impact of legal drinking on deaths from internal causes.
- Results here are close to those from an MLDA regression discontinuity design (as we'll soon see)

3.1 Worried About Uncommon Trends? Relax, Have a Drink!

- With many states and years, we can relax the common trends assumption and allow separate linear trends for each state. Regression DD with state-specific trends looks like this:

$$Y_{st} = \alpha + \delta_{rDD}LEGAL_{st} + t \left[\sum_{k=AL}^{WY} \theta_k STATE_{ks} \right] + \sum_{k=AL}^{WY} \beta_k STATE_{ks} + \sum_{j=1971}^{1983} \gamma_j YEAR_{jt} + e_{st} \quad (3)$$

Coefficient θ_k captures the linear trend for state k

- We can also write

$$Y_{st} = \beta_s + \gamma_t + \theta_s t + \delta_{rDD} LEGAL_{st} + e_{st},$$

where state-specific trends are denoted θ_s

- Figs 5.4-5.6 shows how this works for the state of Allatsee, which reduced its MLDA to 18 in 1975 and neighboring Alabaster, which held the line at 21.
 - Fig. 5.4 is the ideal parallel trends scenario:

FIGURE 5.4
An MLDA effect in states with parallel trends

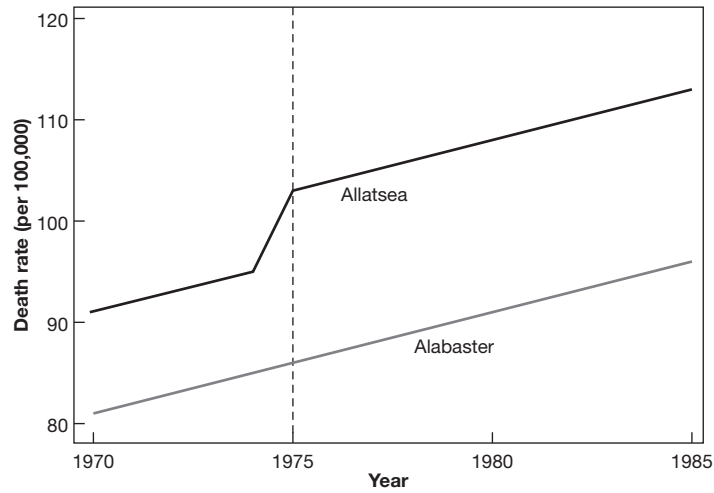
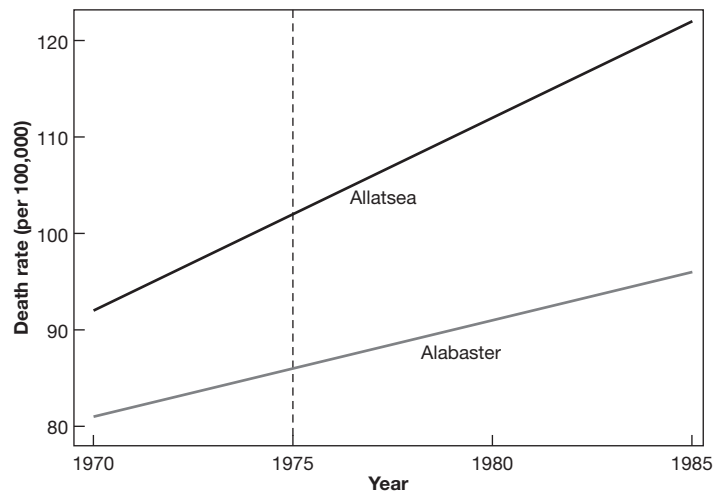
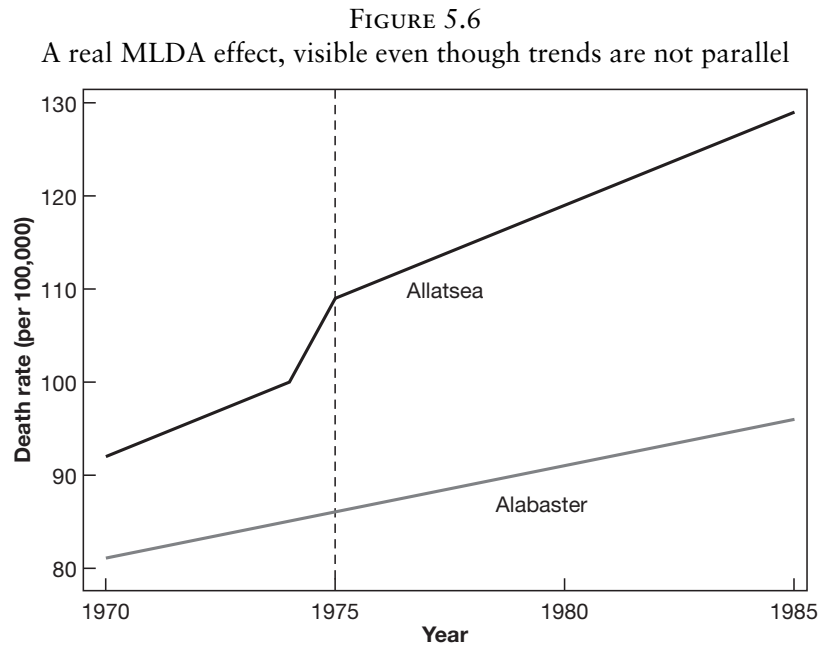


FIGURE 5.5
A spurious MLDA effect in states where trends are not parallel



- Fig. 5.5 is worrying: trends are unchanged in 1975, yet regression DD is likely to generate estimates suggesting an MLDA effect!

- Luckily, the trends here are linear. The model in (3) allows for this, recovering the correct treatment effect in the presence of state-specific trends
- This picture tells the story:



- As it turns out, the MLDA estimates shown in Table 5.2 are not highly sensitive to control for trends.
 - That’s DD heaven!
 - The event-study framework offers an alternative approach to this sort of dynamic DD
- But weight! (Happily this doesn’t matter much, either)
 - Read all about it MM Section 5.2

3.2 OVBeer Taxes

- State policy is a messy business, with frequent changes on many fronts
- An important consideration in research on drinking is the price of a drink; this price, it turns out, is mostly state taxes
- Data on many states and years allow us to explore the effects of similarly-timed policy innovations such as coincident tax changes and MLDA changes
- Regression DD results from models including state beer taxes appear in Table 5.3

TABLE 5.3
Regression DD estimates of MLDA effects controlling for beer taxes

Dependent variable	Without trends		With trends	
	Fraction legal (1)	Beer tax (2)	Fraction legal (3)	Beer tax (4)
All deaths	10.98 (4.69)	1.51 (9.07)	10.03 (4.92)	−5.52 (32.24)
Motor vehicle accidents	7.59 (2.56)	3.82 (5.40)	6.89 (2.66)	26.88 (20.12)
Suicide	.45 (.60)	−3.05 (1.63)	.38 (.77)	−12.13 (8.82)
Internal causes	1.46 (1.61)	−1.36 (3.07)	.88 (1.81)	−10.31 (11.64)

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds, controlling for state beer taxes. The table shows coefficients on the proportion of legal drinkers by state and year and the beer tax by state and year, from models controlling for state and year effects. The fraction legal and beer tax variables are included in a single regression model, estimated without trends to produce the estimates in columns (1) and (2) and estimated with state-specific linear trends to produce the estimates in columns (3) and (4). The sample size is 700. Standard errors are reported in parentheses.

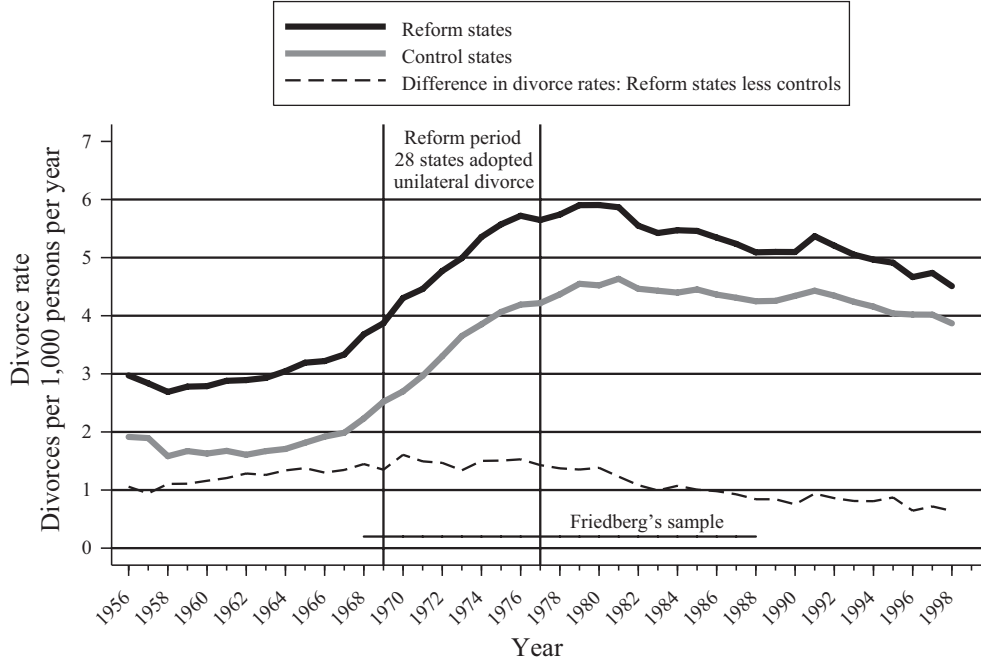
From *Mastering Metrics: The Path from Cause to Effect*. © 2015 Princeton University Press. Used by permission. All rights reserved.

4 Eventful Studies

As can be seen in the figure below, divorce rates (measured as annual number of new divorces per thousand persons in each state) shot up in the 1970s. Could be good for unhappy couples, but perhaps not for everyone: divorce makes women and children poorer.

- Changes in family law might be to blame for increased divorce. Starting in the late 1960s, many states introduced unilateral divorce laws, allowing husbands and wives to divorce without spousal consent.
- Luckily (for econometricians like [Wolfers \(2006\)](#) interested in divorce), states changed their divorce laws at different times. This facilitates a many-state DD analysis like that used to study the MLDA above. In this context, however, we deploy a new kind of DD analysis called an *event-study design*.

WOLFERS: DID UNILATERAL DIVORCE LAWS RAISE DIVORCE RATES?



- Modern event study models allow for time-varying treatment effects
- The payoff to this extension is a more nuanced picture of policy effects, and a framework that can be used to validate the key parallel trends assumption. The cost is the elaborate notation needed to explain the setup, and the risk of implementation mess-ups

4.1 Staggering Notation

- Let D_{st} indicate states and years allowing unilateral divorce. In a *staggered adoption design*, as in the unilateral divorce setting, treatment stays on once switched on, so $\Delta D_{st} \equiv D_{st} - D_{st-1}$ equals one in the year that state s implements unilateral divorce and is zero otherwise.
 - This makes ΔD_{st} a *treatment switch*
- Lagged treatment switches (*lags* for short), denoted ΔD_{st-j} , equal one in year t when state s switched to unilateral divorce j years ago.
 - California, for instance, adopted unilateral divorce in 1970. $D_{CA,t}$ therefore equals 1 for $t = 1970$ and later; $\Delta D_{CA,t}$ equals 1 only in 1970; $\Delta D_{CA,t-2}$ equals 1 only in 1972.
- Leading treatment switches (*leads* for short), denoted ΔD_{st+j} , equal one in year t when state s switches to unilateral divorce j years ahead of t .
- Only one among the set of treatment switches $\{\dots, \Delta D_{st+2}, \Delta D_{st+1}, \Delta D_{st}, \Delta D_{st-1}, \Delta D_{st-2}, \dots\}$ is switched on in a given state and year.
 - In California in 1972, for instance, $\Delta D_{CA,t-2} = 1$, while $\Delta D_{CA,1972-j}$ for all other $\pm j$ equals 0
 - This feature helps us model time-varying treatment effects, called *dynamic* effects

The Event-Study Equation

An event study model with q lags and m leads can be written:

$$Y_{st} = \sum_{j=-m}^{-2} \tau_j \Delta D_{st-j} + \sum_{j=0}^q \tau_j \Delta D_{st-j} + \gamma_s + \lambda_t + \eta_{st}, \quad (4)$$

where parameters τ_j are the treatment effects of interest and Y_{st} is the divorce rate in percent (a rate of 3/1000 is .3 percent). These are indexed by years since or until the year of adoption, sometimes said to mark *event time*, indexed by j in the summations above. Like basic regression DD, event study models include TWFEs, denoted here by γ_s and λ_t .

- $\Delta D_{st} = 1$ in the year in which unilateral divorce is adopted in state s . The treatment effect associated with this year is τ_0 , a term that appears in the sum $\sum_{j=0}^q \tau_j \Delta D_{st-j}$ when $j = 0$. Treatment effects one and two years after adoption are τ_1 (when $\Delta D_{st-1} = 1$) and τ_2 (when $\Delta D_{st-2} = 1$)
 - In general, coefficients on D_{st-j} (called *lagged effects*) capture treatment-effect dynamics after a policy change, revealing whether effects increase, stabilize, or fade. Unilateral divorce might matter little, for instance, in the year it's first adopted, with a growing impact thereafter.
- *Leading effects* appear in the first sum in (4), the term $\sum_{j=-m}^{-2} \tau_j \Delta D_{st-j}$. For California, the term inside this sum that looks two years ahead switches on in 1968 (this is when $\Delta D_{CA,t+2} = 1$; the effect of this is τ_{-2}).
 - Leads capture anticipatory effects of treatment and, especially, shed light on the key parallel trends assumption
- An event study model omits at least one lead or lag (to see this, note that reform states must reform sooner or later. A full set of switches for such states therefore sums to the state's state dummy).
 - Equation (4) omits $\tau_{-1} \Delta D_{st+1}$ from the sum $\sum_{j=-m}^{-2} \tau_j \Delta D_{st-j}$
 - Dynamic effects are therefore measured relative to divorce rates in the year *ahead* of the reform year. Estimates of τ_0 , for instance, measure the extent to which divorce rose in the year unilateral reform was introduced, relative to the year before, while τ_2 contrasts divorce rates two years after reform with the same pre-reform benchmark.
 - Think of event-study models as doing DD separately for each event-time horizon, both pre and post, with the baseline set to be the year before adoption. To see this, consider a data set with data on two states (CA and NY) and four years (1968, 1969, 1970, 1971). CA reforms in 1970; NY never reforms in this period. Equation (4) implies:

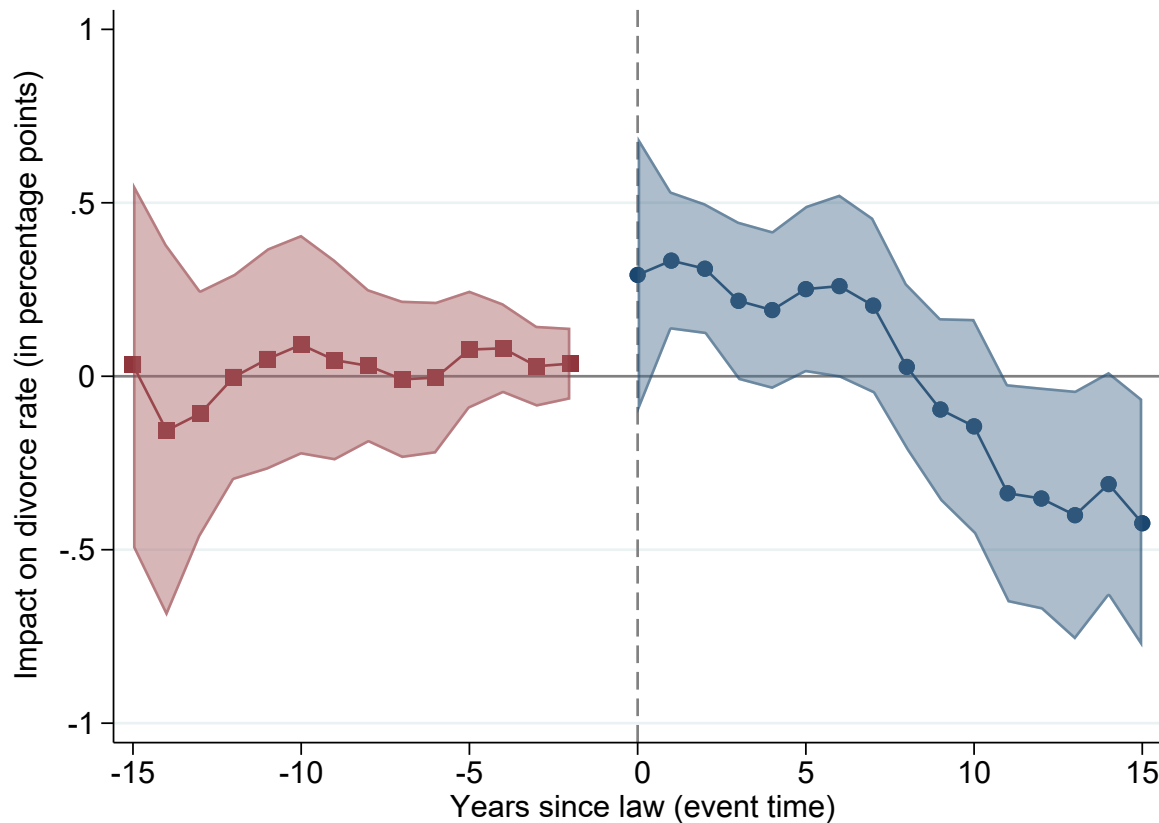
$$\begin{aligned} \tau_1 &= (Y_{CA,71} - Y_{CA,69}) - (Y_{NY,71} - Y_{NY,69}) \\ \tau_0 &= (Y_{CA,70} - Y_{CA,69}) - (Y_{NY,70} - Y_{NY,69}) \\ \tau_{-1} &= 0 \\ \tau_{-2} &= (Y_{CA,68} - Y_{CA,69}) - (Y_{NY,68} - Y_{NY,69}) \end{aligned}$$

In a data set with 8 obs, a regression model with a constant, one state effect, 3 year effects, and 3 treatment effects fits perfectly.

- * Parameter τ_{-2} (a lead) compares divorce changes in CA and NY in pre-treatment years. If this is zero, trends in the two states are parallel and there's a good case for seeing τ_0 and τ_1 as causal effects of adoption.

Estimates

The figure below plots event-study estimates of unilateral divorce effects, with the year before adoption set as reference year. Treatment is coded as staggered-adoption: once reformed, states remain unilateral thereafter. Estimates are computed using data from 1958-1998.



- Unilateral divorce appears to boost divorce rates by around 0.3 percentage points in the first seven years after adoption.
- Estimates then plummet, and, nine years out, turn negative, settling roughly 0.4 percentage points lower. After a period in which couples held together by the old regime separated, divorce rates fell to a new, lower level.
- Evidence of causal validity comes from the pre-treatment coefficients plotted in the figure. Divorce trends in reforming states and years do not appear to have been diverging ahead of the advent of a unilateral regime. In other words, the estimated leads are consistent with the presumption of parallel trends.
- This analysis ignores the possibility that divorce law can affect marriage rates: the denominator is state *population*. You might want to divide by the number married instead.

Binning Business

The event study model generates many estimated effects: 27 leads and 29 lags, to be precise (the figure above shows only 15 leads and lags to avoid clutter).¹

- Standard errors are higher for treatment effects at longer leads and lags because fewer observations contribute to the more distant estimates.
- Suppose we *bin* leads and lags for event time values $j \geq 15$ and $j \leq -15$, with the associated pooled treatment effects labeled τ_{15} and τ_{-15} , respectively. An event study model with binned leads and lags looks like this:

$$\begin{aligned}
 Y_{st} = & \sum_{j=-14}^{-2} \tau_j \Delta D_{st-j} + \sum_{j=0}^{14} \tau_j \Delta D_{st-j} \\
 & + \tau_{-15} \left(\sum_{j \leq -15} \Delta D_{st-j} \right) + \tau_{15} \left(\sum_{j \geq 15} \Delta D_{st-j} \right) \\
 & + \gamma_s + \lambda_t + \eta_{st}.
 \end{aligned} \tag{5}$$

The terms in parentheses collapse treatment indicators for distant event times into single dummies (recall that for any given state and year only one year's treatment switch is ever switched on).

- Wolfers (2006) also pairs years, reporting pooled estimates for $j \in \{0, 1\}$; $j \in \{2, 3\}$, and so on
- Estimates of the binned & paired model align with those in the figure (while the associated precision gain is disappointingly modest).

¹Let $c(s)$ denote a reforming state's reform year, that is, the year state s is first treated. A lag of length q requires at least one treated state with data up to q years after adoption. The longest possible lag is therefore $q = T - \min_s(c(s))$, where T is panel length. Similarly, a lead of length m requires at least one treated state with observations up to m years before adoption. The longest lead is therefore $m = \max(c(s)) - 1$.