Applied Microeconometrics II , Lecture 4

Ciprian Domnisoru

Aalto University

Triple differences (DDD)

Even more controls: difference in differences in differences

- Consider the 1993 New Jersey minimum wage example
- Assume we also had data on employment in sectors not affected by minimum wage legislation.
- We have used employment in the fast food sector in a different (control) state as the control group.
- We could use employment in the non affected sector in the treatment state as an additional control group.

Difference in differences in differences

- DDD approach combines both strategies and computes 2 DD estimators:
- In order to control for different time trends in the affected versus the non affected sector, for each state, compute :

$$DD_{NJ} = [E(Y_{st}|s = NJ, t = 1, sector = aff) - E(Y_{st}|s = NJ, t = 0, aff)]$$
$$-[E(Y_{st} \mid s = NJ, t = 1, unaff) - E(Y_{st} \mid s = NJ, t = 0, unaff)]$$

- DDD estimator is given by the difference between the two DD estimators: $DDD = DD_{NJ} - DD_{Penn}$

DDD in regression format

DDD can be estimated as β , the coefficient on the triple interaction, in the following regression:

- $y_{it} = \beta_0 + \alpha NJ + \mu AFF + \lambda POST + \tau(NJ * AFF) + \delta(POST * NJ) + \pi(POST * AFF) + \beta POST * NJ * AFF$

Difference in differences in differences example

- Gruber (1994) studied state mandates for employer-provided health insurance to cover pregnancy costs. The research question is whether these mandates had an impact on wages (whether firms can pass the costs through to employees).
- State law changes to mandate coverage of pregnancy costs in 1976. Gruber has 3 treatment (IL, NJ and NY) and 5 nearby control states - concerned about different wage time trends in the treated and the control states. He employs a DDD approach, taking employees over 40 and single males as additional (non affected) control group.

Difference in differences in differences example

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

Notes: Cells contain mean log hourly wage for the group identified. Standard errors are given in parentheses; sample sizes are given in square brackets. Years before/after law change, and experimental/nonexperimental states, are defined in the text. Difference-in-difference-in-difference (DDD) is the difference-in-difference from the upper panel minus that in the lower panel.

Difference in differences in differences example

- Affected women in treatment states experienced a 3.4% decline in wages while the wages of eligible women in control states increased by 2.8%. The DD estimate : there was a signicant 6.2% relative fall in wages for women in treatment states.
- However, if the labor markets in treatment states experienced a distinct time trend in the observation period, the 2.8% increase in wages in the control group does not form a valid counterfactual.
- Gruber finds a 0.8% relative wage decline of non affected individuals in the treatment vs control states.
- The overall effect of the mandate is therefore a decline in affected women's wages by 5.4%.

DDD as a robustness check

- if the DD of the placebo (unaffected group, older men , in our example) is 0, DDD=DD, but with larger standard errors
- if the DD of the placebo is large and significant, DD estimation is compromised.
- in Gruber's paper, DDD serves as a robustness check, slightly decreasing the estimate.

DD with spatial variation as control

DD with spatial variation as control: Impact of Crime Risk on Property Values

- A number of papers have documented an inverse relationship between property values and local crime rates. Geographic amenities and other local environmental factors correlated with both.
- Linden and Rockoff (2008) combine data from the housing market with data from sex offender registrations to estimate individuals' valuation of living in close proximity to a convicted criminal
- the timing of a sex offender's arrival allows them to confirm the absence of substantive preexisting differences in property values and to control for the remaining minor differences
- this estimation strategy hinges on the relative similarity of homes sold within 0.1 miles of an offender to homes sold between 0.1 and 0.3 miles of an offender
- the average price of homes sold closest to the offender declines by roughly 4 percent (about $5,500)

Application: Impact of Crime Risk on Property Values

- Difference-in-differences specification

$$log(P_{ijt}) = \alpha_{jt} + \beta X_i + (\omega_0 D_{ijt}^{\frac{3}{10}} + \pi_0 D_{ijt}^{\frac{1}{10}})$$
$$+(\omega_1 D_{ijt}^{\frac{3}{10}} + \pi_1 D_{ijt}^{\frac{1}{10}}) \times Post_{it} + \epsilon_{ijt}$$

  - $log(P_{ijt})$ is the log of the deflated sale price P of the house i in neighborhood j in year t
  - $D_{ijt}^{\frac{1}{10}}$ is an indicator variable set to one if a property sale occurs within 0.1 miles of an offender's address
  - $D_{ijt}^{\frac{3}{10}}$ is an indicator variable set to one if a property sale occurs within 0.3 miles of an offender's address
  - $Post_{it}$ is an indicator for whether the sale takes place after the offender's arrival
  - α_{jt} is a set of neighborhood-year fixed effects
  - X_i is a vector of observable property characteristics
  - *impact of a sex offender on home values* is given by π_1

Application: Impact of Crime Risk on Property Values

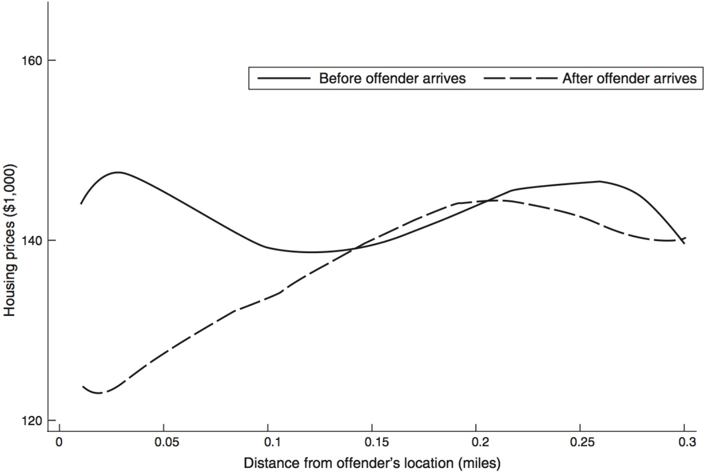

FIGURE 2B. PRICE GRADIENT OF DISTANCE FROM OFFENDER
(Sales during year before and after arrival)

Note: Results from local polynomial regressions (bandwidth = 0.075 miles) of sale price on distance from offender's future/current location.

Application: Impact of Crime Risk on Property Values

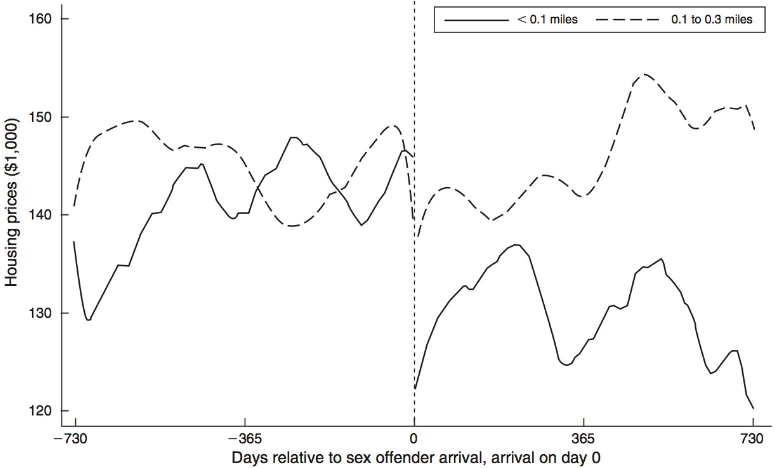

FIGURE 3B. PRICE TRENDS BEFORE AND AFTER OFFENDERS' ARRIVALS
(Parcels within three-tenths of a mile of offender location)

Note: Results from local polynomial regressions (bandwidth = 90 days) of sale price on days before/after offender arrival.

Application: Impact of Crime Risk on Property Values

TABLE 3—IMPACT OF SEX OFFENDERS' LOCATIONS ON PROPERTY VALUE AND SALE PROBABILITY

	Log (sale price) pre-arrival		Log (sale price), pre- and post-arrival				Probability of sale†
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Within 0.1 miles of offender	−0.340 (0.052)*	−0.007 (0.013)	−0.007 (0.012)	<0.001 (0.013)	−0.006 (0.012)	−0.006 (0.012)	−0.029 (0.035)
Within 0.1 miles × post-arrival			−0.033 (0.019)+	−0.041 (0.020)*	−0.036 (0.021)+	−0.116 (0.059)+	0.126 (0.059)*
Dist* ≤ 0.1 miles × post-arrival (0.1 Miles = 1)						0.107 (0.064)+	
Within 1/3 miles of offender				−0.010 (0.007)			
Within 1/3 miles × post-arrival				0.010 (0.010)	0.003 (0.016)	0.004 (0.016)	−0.055 (0.040)
H_0: within 0.1 miles × post-arrival = 0			*p-value* = 0.079	*p-value* = 0.0443	*p-value* = 0.0828	*p-value* = 0.0502	*p-value* = 0.0361
Housing characteristics		✓	✓	✓	✓	✓	✓
Year fixed effects	✓						
Neighborhood-year fixed effects		✓	✓	✓			
Offender area-year fixed effects					✓	✓	✓
Restricted to offender areas					✓	✓	✓
2 years pre- and post-arrival							
Standard errors clustered by…	*Neighborhood*	*Neighborhood*	*Neighborhood*	*Neighborhood*	*Offender area*	*Offender area*	*Offender area*
Sample size	164,993	164,968	169,557	169,557	9,086	9,086	1,519,364
R^2	0.01	0.84	0.83	0.83	0.75	0.75	0.01

Note: Pre-arrival (post-arrival) refers to the two-year period before (after) the date upon which offenders registered their current address. Standard errors in parentheses.

Application: Impact of Crime Risk on Property Values

areg logprice closeoffender postmove closepostmove distpost HEATED AGE NEW AIRCOND BEDROOMS BATHS CNTLH* CNTLW* CNTLBQM1* BQM2, absorb(srnyear) cluster(srn)

```
Linear regression, absorbing indicators         Number of obs     =      9,086
Absorbed variable: srn_year                     No. of categories =        717
                                                F(  30,    173)   =      59.22
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7544
                                                Adj R-squared     =     0.7324
                                                Root MSE          =     0.2912

                                    (Std. Err. adjusted for 174 clusters in srn)
```

log_price	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
close_offender	-.0061233	.0122997	-0.50	0.619	-.0304001	.0181536
post_move	.0040452	.0164753	0.25	0.806	-.0284732	.0365637
close_post_move	-.1156934	.0586683	-1.97	0.050	-.2314911	.0001044
dist_post	.1074868	.0640508	1.68	0.095	-.0189348	.2339083
HEATED	.0002667	.0000168	15.87	0.000	.0002335	.0002999
AGE	-.0076055	.0010155	-7.49	0.000	-.0096097	-.0056012
NEW	-.1023447	.0253246	-4.04	0.000	-.1523296	-.0523597

Standard errors for regression DD

Standard Errors for Regression DD

- Regression DD is a special case of estimation with *panel data*
- economic data of this sort typically exhibit a property called *serial correlation*- the values of variables for nearby periods are likely to be similar
- When the dependent variable in a regression is serially correlated, the residuals from any regression model explaining this variable are often serially correlated as well
- if we ignore this problem and use the simple standard error formula, we exaggerate the precision of regression estimates – standard errors are too small
- clustering – option *vce(cluster clustvar)* in Stata – allows for correlated data within researcher-defined clusters (*clustvar*)

How many clusters? Multi-way clustering

- Ideally, the number of **clusters** should be very large (approach infinity)
- Ideally, the number of **treated clusters** should be very high (approach infinity)
- Be especially wary of few treated units (less than 10) and few clusters (less than 50).
- Conley, T. G., Taber, C. R. 2011. Inference with "difference in differences" with a small number of policy changes. Review of Economics and Statistics 93: 113–125.
- Cameron,A. Colin and Douglas L. Miller. 2015. A practitioner's guide to cluster-robust inference, Journal of Human Resources, 50(2): 317-72
- Varying sizes among clusters. Use **wild bootstrap**. Mackinnon, James G. 2019. How cluster-robust inference is changing applied econometrics, Canadian Journal of Economics
- **boottest** after estimation
- Not enough clusters, not enough control groups even for bootstrap? Consider regional variation, synthetic control method or regression discontinuity methods.
- Cluster along two dimensions? Do you have enough clusters on both dimensions? Check out the Stata **reghdfe** command

The boottest command - an example from my work

Table A7: Effects of the Berthoin Reform on Educational Attainment and Earnings, FQP, Men, Global Polynomial Approach

	All	Fathers employed in:		Father's education:	
		Lower education occupations	Higher education occupations	Lower education	Higher education
First stage effect of CSL	0.451***	0.593**	0.076	0.452***	0.164
	(0.108)	(0.214)	(0.227)	(.150)	(.493)
F (first stage instrument)	17.46	7.64	0.11	9.09	0.11
Reduced form effect:	0.048***	0.045**	0.041	0.044	0.030
	(0.014)	(0.017)	(0.026)	(0.026)	(0.049)
OLS estimate	0.058***	0.053***	0.054***	0.055***	0.058***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
2SLS point estimate	0.107***	0.075***	0.536	0.099**	0.183
	(0.028)	(0.024)	(0.132)	(.045)	(.531)
AR confidence intervals	[.057,.179]	[.031, 166]	...	[-.021, .206]	...
Wild bootstrap p-value	0.010	0.031	0.267	0.414	0.689
Observations	13,207	6,702	6,263	8,758	4,455

Notes: The sample is restricted to French men born in France between 1944 and 1962, who are employed in France, and for whom I observe earnings information. Earnings are monthly earnings, obtained by dividing yearly net earnings by the number of months worked the previous year. Regressions include survey year fixed effects, quadratic polynomials in year of birth on either side of the policy discontinuity, and controls for part-time work status and place of birth. Standard errors are clustered at the year of birth level.*significant at 10 %; **significant at 5%; ***significant at 1%.

Domnisoru, Ciprian. Heterogeneity across Families in the Impact of Compulsory Schooling Laws, 2021. *Economica*

Permutation or randomization tests

- Suppose only one state (out of 50) is treated.
- Re-estimate the equation of interest 50 additional times, replacing the indicator for the treated state with a placebo state every time.
- With fifty placebo estimates, achieving 10% significance from a two-tailed test requires that the state originally treated be ranked second from the top or bottom of the **placebo distribution**, while 5% significance requires that it should be ranked at the top or the bottom.

Permutation or randomization tests: example

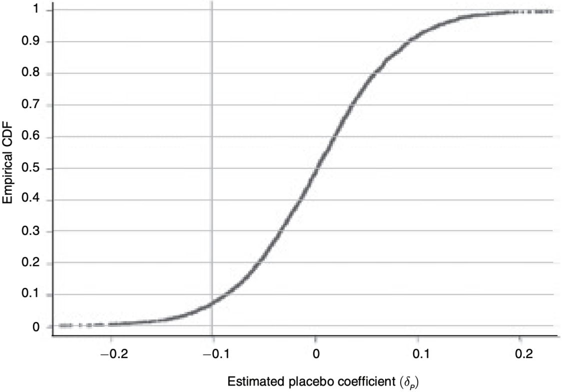

FIGURE 1. DISTRIBUTION OF PLACEBO ESTIMATES: LOG QUANTITY

Notes: This figure plots the empirical distribution of placebo effects (G) for log quantity. The CDF is constructed from 4,725 estimates of δ_p using the specification in column 3 of Table 4. No parametric smoothing is applied: the CDF appears smooth because of the large number of points used to construct it. The vertical line shows the treatment effect estimate reported in Table 4.

The Event Study Design

Leads and lags

$$Y_{it} = \alpha_i + \gamma_t + \delta_{DD} TREAT_{it} + e_{it}$$

- $y_{it} = \alpha_i + \gamma_t +$
  $\beta_{T_{-k}} + ...\beta_{T_{-3}} + \beta_{T_{-2}} + \beta_{T_0} + \beta_{T_{+1}} + \beta_{T_{+2}} + ...\beta_{T_{+k}} + \rho X_{it} + u_{it}$
- Plus terms are called "lags", negative terms "leads"
- There may or may not be never treated units.
- Units may be treated at different times- we are applying a normalization.

“Event-Study” Research Design: Example

- Juvenile curfews are local ordinances proscribing minors, generally within a specified age range, from occupying public areas and streets during particular times
- a study by the U.S. Conference of Mayors (1997) found that 80 percent of the 347 cities with population over 30,000 had youth curfews
- curfew laws – enacted at different times in different cities – are usually referred to as the “event” in the “event-study” design
- Previous studies relied on variation in the date of adoption of city curfew laws to identify treatment effects on criminal behavior
- these studies may easily generate biased results if curfew laws are enacted in response to city-specific trends in arrests

Application: The Impact of Juvenile Curfew Laws

- Consider the following econometric model of arrests

$$R_{ct} = \sum_{\tau=-q}^{m} \beta_\tau D_{ct}^{\tau} + \theta_c + \psi_t + \epsilon_{ct}$$

- R_{ct} is the log of the number of arrests of individuals in some age group of interest in city c in calendar year t
- θ_c is a set of city-fixed effects
- ψ_t is a set of year-fixed effects
- ϵ_{ct} is an error term that may exhibit arbitrary dependence within city but is uncorrelated with the other right-hand side variables
- The D_{ct}^{τ} are a series of "event-time" dummies that equal one when curfew enactment is τ periods away in city c. Formally, we may write:

$$D_{ct}^{\tau} \equiv 1[t - e_c = \tau]$$

- $1[.]$ is an indicator for the expression in brackets being true
- e_c is the year a curfew is enacted in city c

"Event-Study" Research Design

- Recall the econometric model of arrests

$$R_{ct} = \sum_{\tau=-q}^{m} \beta_\tau D_{ct}^{\tau} + \theta_c + \psi_t + \epsilon_{ct}$$

- Thus, the β_τ coefficients represent the time path of arrests relative to the date of curfew enactment for cities subject to the curfew, conditional on city- and year-fixed effects
- If curfews are randomly assigned, this restriction should hold:

$$\beta_\tau = 0 \quad \forall \tau < 0$$

  - in words, this condition states that curfew enactment is not, on average, preceded by trends in city-specific arrests

"Event-Study" Research Design

- Notice that not all of the β_τ's can be identified as the D_{ct}^{τ}'s are perfectly collinear in the presence of the city effects
- Thus, we should normalize $\beta_{-1} = 0$, so that all postenactment coefficients can be thought of as treatment effects
- We should also impose the following endpoint restrictions:

$$\beta_\tau = \begin{cases} \overline{\beta} & \text{if} \quad t \geq 6 \\ \underline{\beta} & \text{if} \quad t \leq -6 \end{cases}$$

  - which simply state that any **dynamics** wear off after six years
  - these restrictions help to reduce some of the collinearity between the year and event-time dummies

"Event-Study" Research Design

- Recall the endpoint restrictions:

$$\beta_\tau = \begin{cases} \overline{\beta} & \textit{if} \quad t \geq 6 \\ \underline{\beta} & \textit{if} \quad t \leq 6 \end{cases}$$

- Because the sample is unbalanced in event time, these endpoint coefficients give unequal weight to cities enacting curfews early or late in the sample

- For this reason, the analysis should be focused on the event-time coefficients falling within a five-year window that are identified off of a nearly balanced panel of cities

- Hypothesis testing is conducted using robust standard errors clustered at the city level

Application: The Impact of Juvenile Curfew Laws

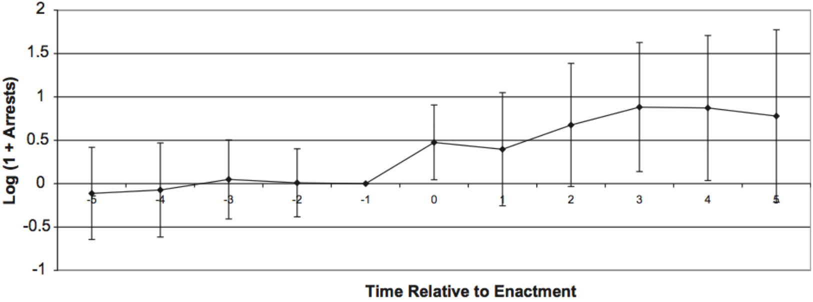

Figure 1. Arrests of Youth below Curfew Age for Curfew and Loitering Violations.

Application: The Impact of Juvenile Curfew Laws

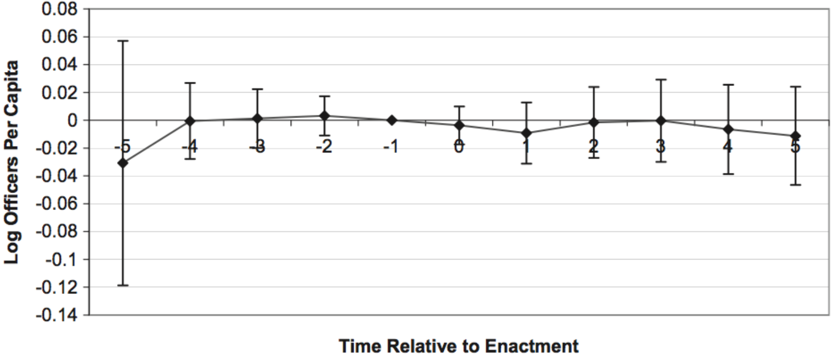

Figure 4. Log Officers per Capita.

Application: The Impact of Juvenile Curfew Laws

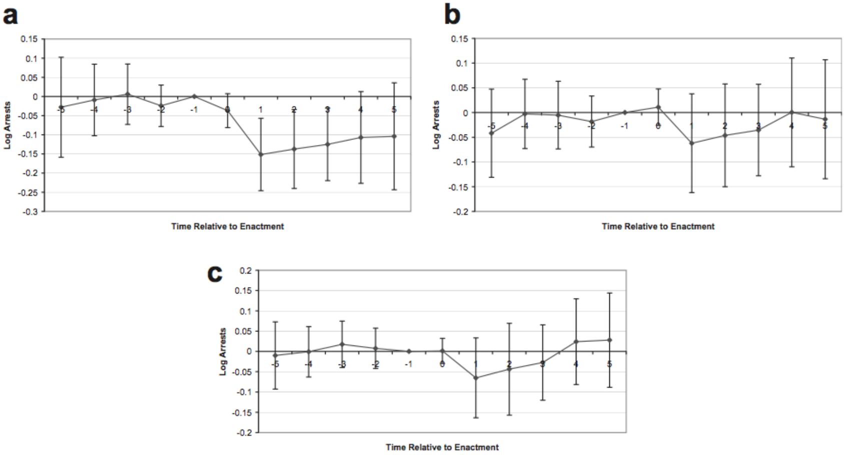

Figure 2. (a) Youth below Curfew Age. (b) Young Adults above Curfew Age. (c) Adults Age 25+.

Autor(2003): Question 2 in your Assignment 2

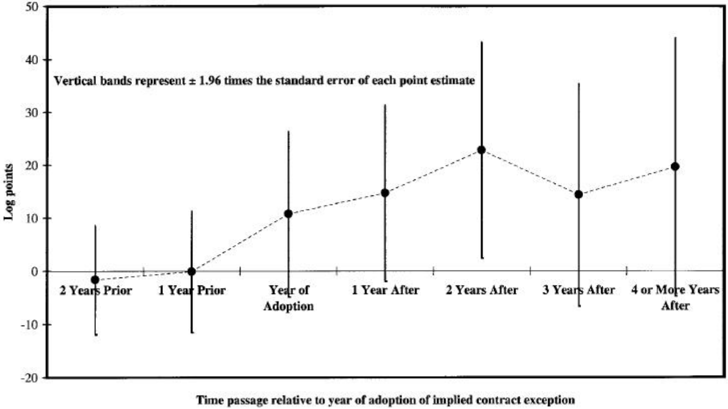

FIG. 3.—Estimated impact of implied contract exception on log state temporary help supply industry employment for years before, during, and after adoption, 1979–95.

Corrections for heterogeneity

- Abraham and Sun extension of the basic event study model:

$$Y_{it} = \alpha_i + \gamma_t + \sum_{\ell \neq \{-1, <-K\}} \beta_\ell D^\ell_{i,t} + \epsilon_{it}$$

$$Y_{it} = \alpha_i + \gamma_t + \sum_c \sum_{\ell \neq \{-1, <-K\}} \beta_{c\ell} \mathbb{1}\{C_i = c\} \times D^\ell_{i,t} + \epsilon_{it}$$

  - Interact the relative time indicators with indicators for the treatment initiation year group (c)
  - Control units are all units not treated. If most of the units are treated by the end, run into issues of nonrepresentative controls.