ECONOMETRIA APLICADA AVANZADA

Diferencias en Diferencias

Junio 10-11, 2023

Cristina Tello-Trillo, Census Bureau & University of Maryland

Example: Meyer et al. (1995)

- Meyer, Viscusi, and Durbin evaluate effect of increase in disability payments on time out of work.
- In Kentucky and Michigan the benefit amount after an injury for high-earnings individuals was raised by approximately 50 percent, while the low-earnings individuals were unaffected by the benefit increase.
- Outcome:
  - Time out of work
- Treatment:
  - Increase in disability payment
- Treatment Group:
  - High earners
- Control Group:
  - Low earners

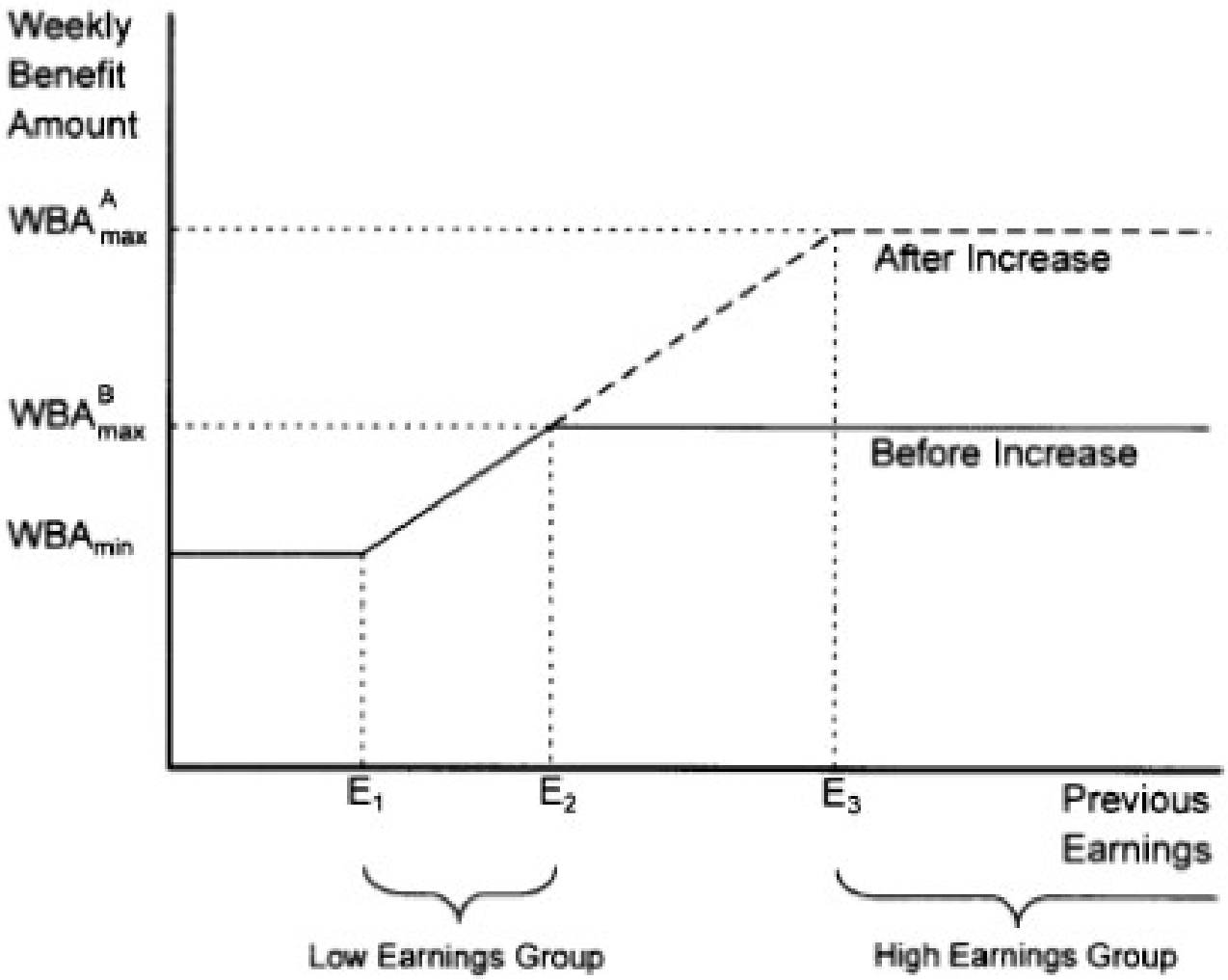


FIGURE 1. TEMPORARY TOTAL BENEFIT SCHEDULE BEFORE AND AFTER AN INCREASE IN THE MAXIMUM WEEKLY BENEFIT

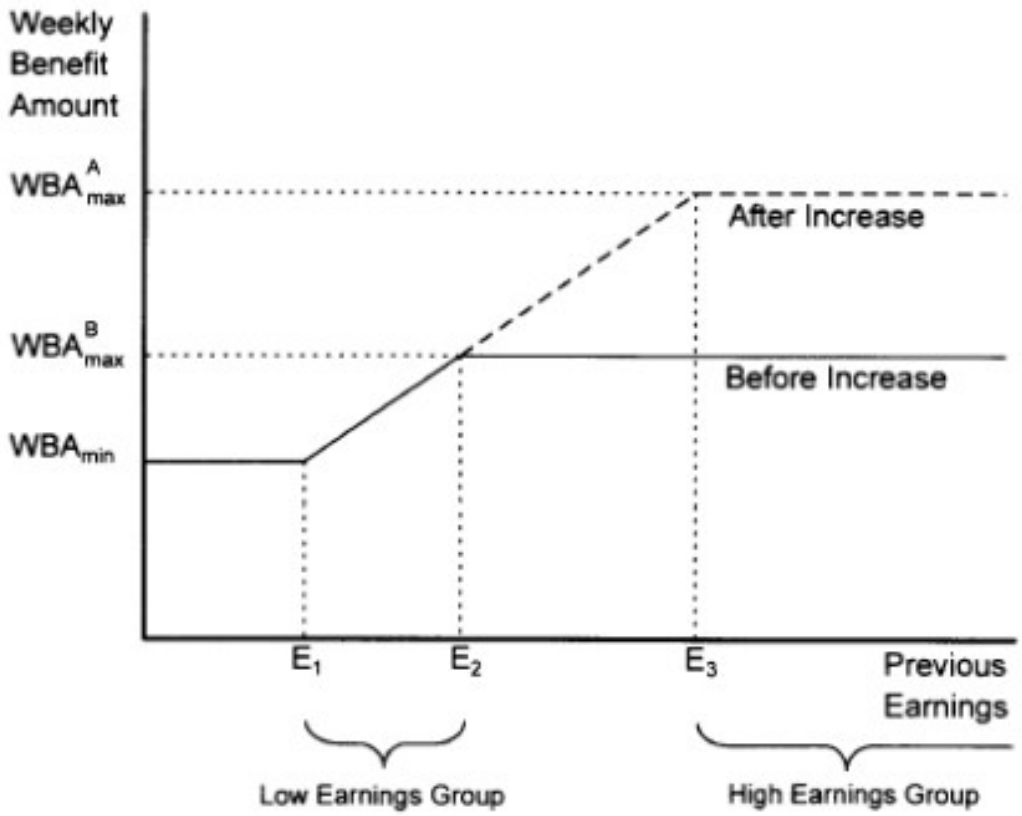


FIGURE 1. TEMPORARY TOTAL BENEFIT SCHEDULE BEFORE AND AFTER AN INCREASE IN THE MAXIMUM WEEKLY BENEFIT

- The group of workers with earnings of at least E3 experience the full effect of the benefit increase (treatment group)
- The comparison group those with earnings between E1 and E2 (the low-earnings group). The benefits these individuals receive are unaffected by the increase in the maximum weekly benefit.

- Workers' compensation programs are run by the individual states and differ widely in their coverage, types of benefits, levels of benefits, and available methods of insurance underwriting.
- Workers' compensation provides both payments for medical care and indemnity (cash) benefits for work-related injuries.
- There were only three large increases in the temporary total maximum benefit levels in the states and time periods included in the NCCI data base used in this study.
- These increases occurred in Florida, Kentucky, and Michigan. The Florida increase coincided with a major overhaul of the workers' compensation law, so that the before versus after comparisons reflect multiple aspects of the change in benefit structure.
  - The Kentucky increase of July 15, 1980, raised the maximum benefit from $131 to $217 per week, a 66-percent increase.
  - The Michigan increase on January 1,1982, raised the maximum benefit from $181 to $307 per week, a 70-percent increase

Data

- The data source for this study is the Detailed Claim Information (DCI) data base collected by the National Council on Compensation Insurance (NCCI).
- The key variables in the data set that we use are: date injured, duration of temporary total benefits, total medical costs, previous earnings, weekly benefit amount, benefit type (i.e., temporary total or permanent partial), type of injury (body part affected and the type of damage).

TABLE 1—REPLACEMENT RATES, EARNINGS, AND DEMOGRAPHIC CHARACTERISTICS DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

	Kentucky			Michigan		
Variable	Before increase (1)	After increase (2)	Percentage change (3)	Before increase (4)	After increase (5)	Percentage change (6)
Maximum benefit ($)	131.00	217.00	65.65	181.00	307.00	69.61
Replacement rate, high earnings (percent)	32.70 (0.25)	51.02 (0.37)	56.02 (1.65)	30.01 (0.35)	44.15 (0.48)	47.14 (2.33)
Replacement rate, low earnings (percent)	66.42 (0.20)	66.66 (0.22)	0.36 (0.44)	66.64 (0.24)	66.35 (0.30)	−0.45 (0.58)
Average benefit (1983 $), high earnings	151.08 (0.96)	239.09 (1.32)	58.25 (1.33)	220.66 (1.78)	320.48 (2.27)	45.24 (1.56)
Average benefit (1983 $), low earnings	118.58 (0.64)	118.26 (0.74)	−0.27 (0.82)	183.66 (0.78)	182.77 (0.93)	−0.45 (0.58)

- Replacement rate: fraction of previous earnings replaced by workers' disability compensation.
- The replacement rate rises dramatically for the high-earnings group which received the benefit increase but remains constant for the low earnings group.

Example: Meyer et al. (1995)

- Open INJURY.DTA
- Outcome= Mean Duration (semanas fuera del trabajo)
- sort highearn afchnge

	Before Change	After Change	Difference
Treatment Group = High earn			
Control Group = Low earn			
	Difference in Differences=		

```
. by highearn afchnge: sum durat
```

-> highearn = 0, afchnge = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
durat	**2294**	**7.475044**	**17.25335**	**.25**	**182**

-> highearn = 0, afchnge = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
durat	**2004**	**8.611527**	**21.9445**	**.25**	**182**

-> highearn = 1, afchnge = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
durat	**1472**	**11.76155**	**30.01366**	**.25**	**182**

-> highearn = 1, afchnge = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
durat	**1380**	**13.93152**	**30.39186**	**.25**	**182**

tab highearn afchng, sum(durat)

```
. tab highearn afchng, sum(durat)

    Means, Standard Deviations and Frequencies of duration of benefits
```

=1 if high earner	=1 if after change in benefits		
	0	1	Total
0	7.4750436	8.6115269	8.0049442
	17.253352	21.944498	19.586868
	2294	2004	4298
1	11.761549	13.931522	12.811536
	30.013656	30.391862	30.211425
	1472	1380	2852
Total	9.1504912	10.781028	9.9222028
	23.186733	25.854768	24.497542
	3766	3384	7150

TABLE 4—KENTUCKY AND MICHIGAN: DURATION AND MEDICAL COSTS OF TEMPORARY TOTAL DISABILITIES DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

	High earnings		Low earnings		Differences		Difference in differences
Variable	Before increase (1)	After increase (2)	Before increase (3)	After increase (4)	[(2)−(1)] (5)	[(4)−(3)] (6)	[(5)−(6)] (7)
Mean duration (weeks)							
Kentucky	11.16	12.89	6.25	7.01	1.72	0.76	0.96
	(0.83)	(0.83)	(0.30)	(0.41)	(1.17)	(0.51)	(1.28)
Michigan	14.76	19.42	10.94	13.64	4.66	2.70	1.96
	(2.25)	(2.67)	(1.09)	(1.56)	(3.49)	(1.90)	(3.97)
Median duration (weeks)							
Kentucky	4.00	5.00	3.00	3.00	1.00	0.00	1.00
	(0.14)	(0.20)	(0.11)	(0.12)	(0.25)	(0.16)	(0.29)
Michigan	5.00	7.00	4.00	4.00	2.00	0.00	2.00
	(0.45)	(0.67)	(0.22)	(0.28)	(0.81)	(0.35)	(0.89)
Mean of log duration							
Kentucky	1.38	1.58	1.13	1.13	0.20	0.01	0.19
	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.07)
Michigan	1.58	1.87	1.41	1.51	0.29	0.10	0.19
	(0.09)	(0.10)	(0.06)	(0.06)	(0.13)	(0.08)	(0.16)
Mean medical cost (dollars)							
Kentucky	3,298.96	1,688.67	878.20	1,155.98	−1,610.29	277.78	−1,888.07
	(1,885.93)	(116.59)	(78.07)	(157.26)	(1,889.53)	(175.57)	(1,897.67)
Michigan	2,229.41	2,585.23	1,538.22	2,017.65	355.82	479.43	−123.61
	(293.93)	(364.76)	(188.49)	(304.83)	(468.45)	(358.40)	(589.83)

- We emphasize the mean of the log of duration because this statistic is likely to be more precisely measured and less susceptible to the influence of a few large obs.
- This issue of robustness (log) is important because the distribution of claims lengths has some large values, but most values are small.

- Statistically, if your variables are right-skew (that is, they have a long tail at the high end) then a measure such as correlation or regression can be influenced a lot by one or a few cases at the high end (outliers, leverage points, influential points).
- Taking the log can aid by reducing or eliminating skew.

Regression estimates of DID

- Y_{it} =Duration (weeks) of the spell of time-out-of-work (i.e. time that you receive workers' compensation disability benefits)
- A_{it} = period after the reform
- H_{it} = treated or high earnings group dummy (Income > E3).

$$Y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 H_{it} + \delta A_{it} H_{it} + \gamma X_{it} + e_{it}$$

- Diff-in-Diff = δ

Regression estimates of DID

IV. Regression Estimates of the Changes in Duration

To account for possible changes in the composition of the sample after the benefit increases, we estimate a series of regression equations that control for all of the available characteristics of the worker, the job, and the injury. Specifically, the estimates in Table 6 control for worker age, marital status, sex, industry, and the severity of the injury as measured by medical costs, hospital days, and type of injury. The specifications that we try are the regression analogue of the differences and difference in differences of Table 4. In all of the equations, the dependent variable is the natural logarithm of duration, measured in weeks. In the absence of censoring and time-

```
. reg durat afchnge highearn treated
```

Source	SS	df	MS
Model	44343.4979	3	14781.166
Residual	4245982.66	7146	594.176136
Total	4290326.16	7149	600.129551

Number of obs =	7150
F(3, 7146) =	24.88
Prob > F =	0.0000
R-squared =	0.0103
Adj R-squared =	0.0099
Root MSE =	24.376

durat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
afchnge	1.136483	.7453242	1.52	0.127	-.3245728	2.597539
highearn	4.286505	.8140426	5.27	0.000	2.690741	5.88227
treated	1.033489	1.178865	0.88	0.381	-1.277435	3.344414
_cons	7.475044	.5089333	14.69	0.000	6.477384	8.472704

Kentucky and Michigan	Before Change	After Change	Difference
Treatment Group = High earn	11.76	13.93	=2.17
Control Group = Low earn	7.47	8.61	=1.14
	4.29		**=1.03**

Difference in Differences

```
 //  logs
global controls "manuf construc head neck upextr trunk lowback lowextr occdis"

reg ldurat afchnge highearn treated male  age prewage $controls, robust

reg ldurat afchnge highearn treated male  age prewage $controls ///
if ky==1, robust
est store regky

reg ldurat afchnge highearn treated male  age prewage $controls ///
if mi==1, robust
est store regmi

esttab regky regmi, se mtitle("Kentucky" "Michigan")  star(* 0.1 ** 0.05 *** 0.01)
```

```
. reg ldurat afchnge highearn treated $controls_disab, robust

Linear regression                               Number of obs     =      6,822
                                                F(18, 6803)       =      93.02
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3095
                                                Root MSE          =     1.0835
```

ldurat	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
afchnge	-.0179322	.0341032	-0.53	0.599	-.0847852	.0489207
highearn	-.2909592	.0622225	-4.68	0.000	-.4129347	-.1689837
treated	.1838059	.0534165	3.44	0.001	.0790929	.2885189
male	-.1185737	.0357572	-3.32	0.001	-.188669	-.0484783
age	-.0005714	.0062689	-0.09	0.927	-.0128605	.0117176
lprewage	.2709347	.0542644	4.99	0.000	.1645595	.37731
manuf	-.1584977	.0307653	-5.15	0.000	-.2188074	-.098188
construc	.1115193	.0390247	2.86	0.004	.0350186	.1880199
head	-.5535192	.1182521	-4.68	0.000	-.7853302	-.3217081
neck	.1736072	.1301383	1.33	0.182	-.0815045	.4287189
upextr	-.0183259	.0913665	-0.20	0.841	-.1974328	.160781
trunk	.0755728	.0952756	0.79	0.428	-.1111971	.2623427
lowback	.0216769	.0923483	0.23	0.814	-.1593546	.2027085
lowextr	.0217744	.0919901	0.24	0.813	-.158555	.2021038
occdis	.4780795	.1969559	2.43	0.015	.0919844	.8641746
married	.0281487	.0310386	0.91	0.364	-.0326968	.0889941
lage	.3076808	.2251274	1.37	0.172	-.1336394	.749001
ltotmed	.3861212	.0115741	33.36	0.000	.3634323	.40881
_cons	-3.352645	.5793962	-5.79	0.000	-4.488443	-2.216847

The reform increase around 17-22% the duration in weeks of time out of work.

	(1) Kentucky	(2) Michigan
afchnge	-0.0149	-0.00752
	(0.0386)	(0.0723)
highearn	-0.223***	0.0480
	(0.0800)	(0.356)
post_treat~t	0.168***	0.224*
	(0.0584)	(0.131)
male	-0.0675*	-0.328***
	(0.0393)	(0.0850)
age	-0.00595	0.0234
	(0.00685)	(0.0149)
lprewage	0.220***	-0.210
	(0.0734)	(0.348)
manuf	-0.158***	-0.129*
	(0.0343)	(0.0678)
construc	0.0554	0.347***
	(0.0436)	(0.0869)
head	-0.443***	-0.808***
	(0.129)	(0.277)

Level-level Regression $y = \beta_0 + \beta_1 x + \epsilon$	y	x	$\Delta y = \beta_1 \Delta x$ "If you change x by one, we'd expect y to change by β_1"
Log-Level Regression $ln(y) = \beta_0 + \beta_1 x + \epsilon$	$ln(y)$	x	$\%\Delta y = 100 \cdot \beta_1 \cdot \Delta x$ "if we change x by 1 (unit), we'd expect our y variable to change by $100 \cdot \beta_1$ percent" Technically, the interpretation is the following: $\%\Delta y = 100 \cdot (e^{\beta_1} - 1)$ but the quoted interpretation is approximately true for values $-0.1 < \beta_1 < 0.1$ (and it's much easier to remember.)
Level-Log Regression $y = \beta_0 + \beta_1 \cdot ln(x) + \epsilon$	y	$ln(x)$	$\Delta y = (\beta_1/100)\%\Delta x$ "If we increase x by one percent, we expect y to increase by $(\beta_1/100)$ units of y." Note, you cannot include obs. for which x<=0 if x is then logged. You either can't calculate the regression coefficients, or may introduce bias.

Log-Log Regression $ln(y) = \beta_0 + \beta_1 \cdot ln(x) + \epsilon$	$ln(y)$	$ln(x)$	%Δy=β1%Δx "if we change x by one percent, we'd expect y to change by β1 percent" Note, you cannot include obs. for which x<=0 if x is logged. You either can't calculate the regression coefficients, or may introduce bias.

TABLE 6—REGRESSION EQUATIONS FOR NATURAL LOGARITHM OF DURATION, HIGH- AND LOW-EARNINGS GROUPS POOLED, AND HIGH-EARNINGS GROUP SEPARATELY

	Specification							
	High- and low-earnings groups pooled				High-earnings group only			
	Kentucky		Michigan		Kentucky		Michigan	
Explanatory variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
After-increase indicator variable	0.016 (0.045)	−0.004 (0.038)	0.082 (0.084)	0.003 (0.073)	0.228 (0.054)	0.149 (0.044)	0.244 (0.136)	0.260 (0.113)
High-earnings-group indicator variable	−1.522 (1.099)	−0.594 (0.930)	5.577 (4.811)	3.607 (4.162)				
After-increase × high-earnings-group indicator variable	0.215 (0.069)	0.162 (0.059)	0.157 (0.153)	0.203 (0.132)				
ln(Previous earnings)	0.258 (0.104)	0.207 (0.088)	0.901 (0.648)	0.139 (0.562)	0.492 (0.163)	0.229 (0.133)	0.067 (0.496)	−0.335 (0.414)
ln(Previous earnings) × high-earnings group	0.232 (0.187)	0.065 (0.158)	−0.973 (0.803)	−0.587 (0.695)				
Male indicator variable	−0.072 (0.046)	−0.070 (0.039)	−0.303 (0.099)	−0.332 (0.086)	−0.088 (0.133)	0.004 (0.108)	−1.053 (0.631)	−0.489 (0.527)
Married indicator variable	0.051 (0.041)	0.055 (0.035)	−0.024 (0.081)	−0.065 (0.070)	0.179 (0.080)	0.112 (0.065)	−0.097 (0.185)	−0.287 (0.154)
ln(Age)	0.252 (0.052)	0.244 (0.044)	0.464 (0.114)	0.481 (0.098)	0.071 (0.092)	0.056 (0.075)	0.796 (0.234)	0.850 (0.195)
ln(Total medical costs)		0.361 (0.011)		0.316 (0.018)		0.421 (0.018)		0.475 (0.037)

- The outcome variable (duration in weeks of Disability compensation) is in logs.
- The reform increase around 15-22% the duration in weeks of Disability Compensation.
- Michigan estimates are not significantly different from zero: in Michigan benefits depended partly on tax filing status and the number of dependents, besides pre-earnings.

- The results of this study suggest a substantial effect of the level of temporary total benefits on the duration of workers' compensation claims.
- The KY estimates are generally significantly different from zero, while the MI estimates are similar in magnitude to the KY estimates but are less precisely measured.
- Should we increase workers' compensation disability benefits?
  - The longer durations that we find after benefit increases may not indicate a loss in social welfare, as longer recovery times may improve subsequent health.
  - Higher benefits may enable injured workers to complete their recovery before returning to work.
  - To examine this question, we would like to be able to examine health status after an individual returns to work.

Regression DID Including Leads and Lags

- Including leads into the DD model is an easy way to analyze pre-trends.
- Lags can be included to analyze whether the treatment effect changes over time after treatment.
- The estimated regression would be:

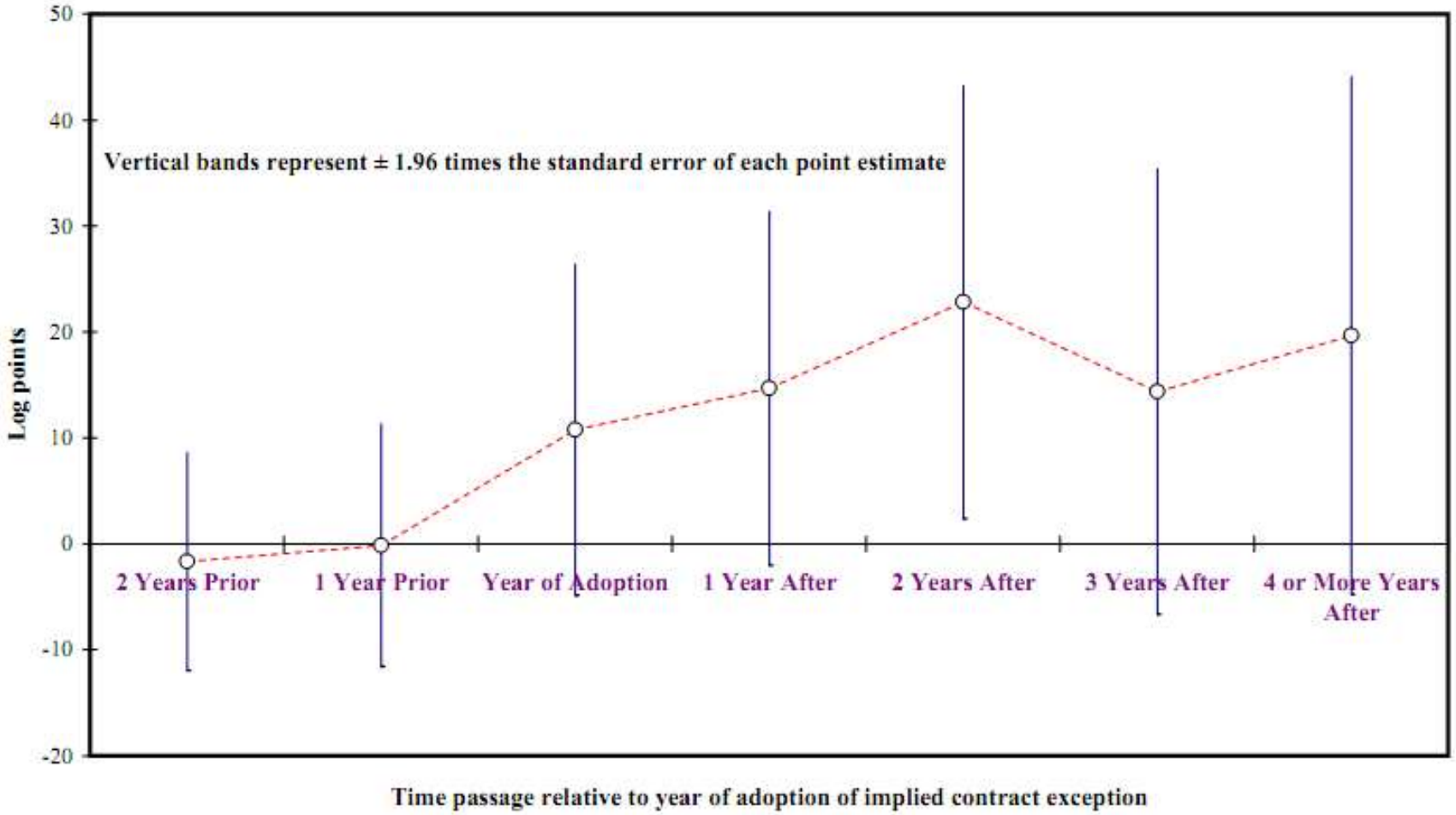
$$Y_{ist} = \alpha + \sum_{t=t-1}^{t+2} I(t) + \beta_{t-1} I(t-1)T_s + \beta_t I(t)T_s + \beta_{t+1} I(t+1)T_s + \beta_{t+2} I(t+2)T_s + X_{ist} + \varepsilon_{ist}$$

Includes q lags (pre-treatment effects) and m leads (post-treatment effects). Treatment occurs in year t.

Study Including Leads and Lags - Autor (2003)

- Autor (2003) includes both leads and lags in a DD model analyzing the effect of increased employment protection on the firm's use of temporary help workers.
- In the US employers can usually hire and fire workers at will.
- Some states courts have made some exceptions to this employment at will rule and have thus increased employment protection.
- Different states have passed these exceptions at different points in time.
- The standard thing to do is to normalize the adoption year to 0.
- Autor then analyzes the effect of these exceptions on the use of temporary help workers.

Temporary workers



- The lags are very close to 0. No evidence for anticipatory effects (good news for the common trends assumption).
- The leads show that the effect increases during the first years of the treatment and then remains relatively constant.

DID Frontier

- The last few years, there has been an overwhelming amount of research papers discussing DID and comparable designs from a methodological perspective.

- What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature (Roth, Sant'Anna, Bilinski, Poe 2022)

Class Exercise

- Kiel and McClain (1995) studied the effect that a new garbage incinerator had on housing values in North Andover, MA.
- The rumor that a new incinerator would be built in North Andover began after 1978, and construction began in 1981. The incinerator was expected to be in operation soon after the start of construction; the incinerator actually began operating in 1985.
- We will use data on prices (KIELMC.dta) of houses that sold in 1978 and another sample on those that sold in 1981.
- The hypothesis is that the price of houses located near the incinerator would fall relative to the price of more distant houses.