Lab Notes

2017-6-2

Difference-in-Differences

Card and Kreuger (1994) study the effects of a minimum wage law change in New Jersey on unemployment. On April 1 1992, New Jersey changed its minimum wage from $4.25 to $5.05. They, along with many other policymakers, were interested in whether this rise in the minimum wage would raise unemployment.

Why would it potentially do this?

What would be the issues with comparing unemployment in towns and cities in New Jersey before the law was implemented and after the law was implemented?

What would be a field experiment that could be designed in New Jersey to evaluate the impact of the minimum wage? Would it be feasible?

While the minimum wage law was passed in New Jersey, it remained unchanged in Pennsylvania. The formed the basis of a quasi-experiment, whereby unemployment in New Jersey (treatment group) could be compared to unemployment in Pennsylvania (control group) to estimate the impact of the minimum wage on unemployment.

Card and Kreuger survey fast food employers in February 1992 (before law) and November 1992 (after law).

```r
rm(list=ls())

DiD <- matrix(c("A","C","B","D"), ncol=2)
rownames(DiD) <- c("Before Law", "After Law")
colnames(DiD) <- c("New Jersey (Treatment)", "Pennsylvania (Control)")

library(knitr)
kable(DiD)
```

	New Jersey (Treatment)	Pennsylvania (Control)
Before Law	A	B
After Law	C	D

A and **C** are the mean levels of unemployment in New Jersey before and after the implementation of the minimum wage law, respectively. **B** and **D** are the mean levels of unemployment in Pennsylvania before and after the implementation of the minimum wage law, respectively.

The difference-in-differences is (**C**-**A**)-(**D**-**B**). In other words, it is the difference in the change in unemployment in New Jersey compared to Pennsylvania.

If the minimum wage law does indeed raise unemployment, then the change in unemployment in New Jersey should be greater than the change in unemployment in Pennsylvania, and the difference-in-differences estimate should be positive. Or there should be evidence that the minimum wage law increases unemployment.

Is this identification strategy valid? What does it critically depend on?

The key assumption of this research design, also sometimes called the **nonequivalent group design**, is that Pennsylvania is an appropriate control group for New Jersey. Under what conditions is this plausible or implausible?

The control group, Pennsylvania, must be a counterfactual for the treatment group, New Jersey. It must describe what would have happened in New Jersey if it had not received the minimum wage law. This means that whatever factors change with time in New Jersey must affect Pennsylvania in the same way.

This is also called the **parallel trends assumption** because the change in unemployment in New Jersey and Pennsylvania must be parallel to one another if New Jersey had not passed the minimum wage law, for the research design to be valid.

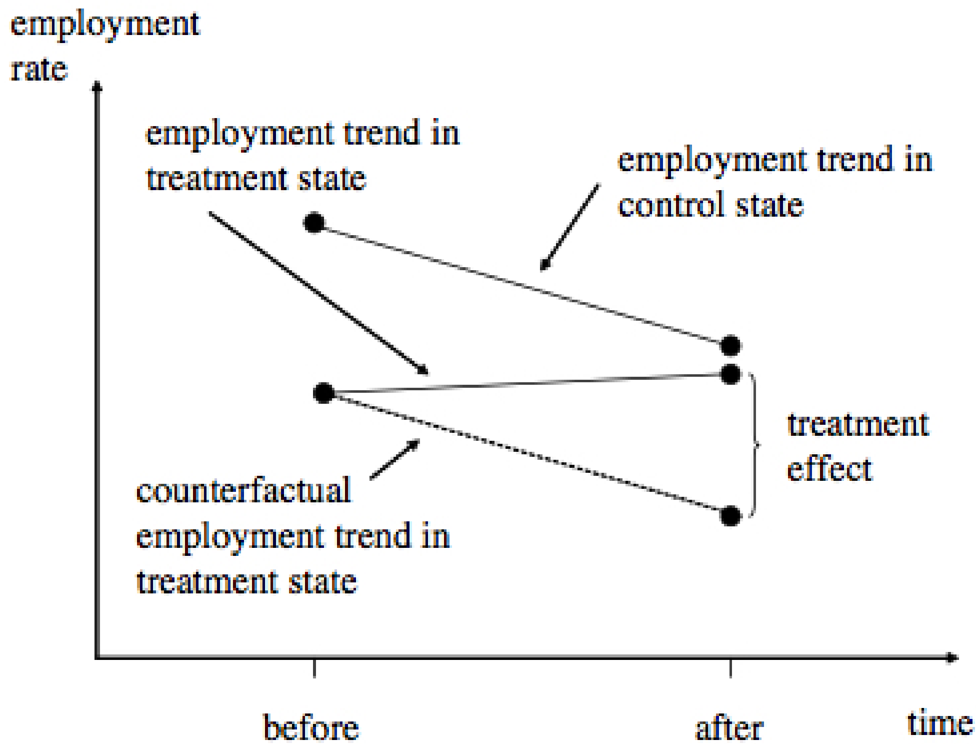

Figure 1:

Do you believe this a valid research design for evaluating the impact of minimum wage?

Regression and Difference-in-differences

The difference-in-differences can be expressed as follows:

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \beta(NJ_s \times d_t) + \epsilon_{ist}$$

y_{ist} is unemployment in restaurant i in state s in year t. NJ_s and d_t are dummy variables indicating New Jersey (versus Pennsylvania) and the post-treatment year (versus the pre-treatment year), respectively. There are four parameters in the model.

$$\alpha = E(y_{ist}|s = PA, t = Feb)$$

This is mean unemployment among restaurants in Pennsylvania before the minimum wage law.

$$\gamma = E(y_{ist}|s = NJ, t = Feb) - E(y_{ist}|s = PA, t = Feb)$$

This is the difference in mean unemployment among restaurants in New Jersey bersus Pennsylvania before the minimum wage law.

$$\lambda = E(y_{ist}|s = PA, t = Nov) - E(y_{ist}|s = PA, t = Feb)$$

This is the difference in mean unemployment among restaurants in Pennsylvania before and after the minimum wage law.

$$\beta = \big(E(y_{ist}|s = NJ, t = Nov) - E(y_{ist}|s = NJ, t = Feb)\big) - \big(E(y_{ist}|s = PA, t = Nov) - E(y_{ist}|s = PA, t = Feb)\big)$$

This is the difference in the differences of mean unemployment among restaurants before and after the minimum wage law in New Jersey compared to Pennsylvania. Hence, the name difference-in-differences. β is the estimate of the causal effect or more specifically the effect of the treatment on the treated.

Card and Krueger's (1994) analysis can be replicated as follows.

```
rm(list=ls())
library(plyr)

data <- read.csv("public.csv", header=T)
colnames(data)
```

```
##  [1] "SHEET"    "CHAINr"   "CO_OWNED" "STATEr"   "SOUTHJ"   "CENTRALJ"
##  [7] "NORTHJ"   "PA1"      "PA2"      "SHORE"    "NCALLS"   "EMPFT"
## [13] "EMPPT"    "NMGRS"    "WAGE_ST"  "INCTIME"  "FIRSTINC" "BONUS"
## [19] "PCTAFF"   "MEAL"     "OPEN"     "HRSOPEN"  "PSODA"    "PFRY"
## [25] "PENTREE"  "NREGS"    "NREGS11"  "TYPE2"    "STATUS2"  "DATE2"
## [31] "NCALLS2"  "EMPFT2"   "EMPPT2"   "NMGRS2"   "WAGE_ST2" "INCTIME2"
## [37] "FIRSTIN2" "SPECIAL2" "MEALS2"   "OPEN2R"   "HRSOPEN2" "PSODA2"
## [43] "PFRY2"    "PENTREE2" "NREGS2"   "NREGS112"
```

```
#Variable construction

data$EMPTOT <- data$EMPPT*0.5 + data$EMPFT + data$NMGRS  #FTE employment, Wave 1
data$EMPTOT2 <- data$EMPPT2*0.5 + data$EMPFT2 + data$NMGRS2 #FTE employment, Wave 2

data$DEMP <- data$EMPTOT2 - data$EMPTOT # Difference in total employment
data$PCHEMPC <- 2*(data$EMPTOT2-data$EMPTOT) / (data$EMPTOT2+data$EMPTOT) # Pct change in employment
data$PCHEMPC[data$EMPTOT2==0] <- -1

data$DWAGE <- data$WAGE_ST2 - data$WAGE_ST # Difference in wage
data$PCHWAGE <- (data$WAGE_ST2 - data$WAGE_ST) / data$WAGE_ST # Pct change in wage

data$GAP <- (5.05-data$WAGE_ST)/(data$WAGE_ST) # Difference between new minimum wage and old wage as a
data$GAP[data$STATEr==0] <- 0
data$GAP[data$WAGE_ST>=5.05] <- 0

data$NJ <- data$STATEr # New Jersey indicator

data$BK <- 0
data$BK[data$CHAINr==1] <- 1 # Burger King indicator
data$KFC <- 0
data$KFC[data$CHAINr==2] <- 1 # KCF indicator
data$ROYS <- 0
data$ROYS[data$CHAINr==3] <- 1 # Roy Rogers indicator
```

```r
data$WENDYS <- 0
data$WENDYS[data$CHAINr==4] <- 1 # Wendy's indicator

data$PMEAL <- data$PSODA+data$PFRY+data$PENTREE # Price of meal, Wave 1
data$PMEAL2 <- data$PSODA2+data$PFRY2+data$PENTREE2 # Price of meal, Wave 2
data$DPMEAL <- data$PMEAL2-data$PMEAL # Price of meal, Difference

data$CLOSED <- 0 # Closed indicator
data$CLOSED[data$STATUS2==3] <- 1

data$FRACFT <- (data$EMPFT/data$EMPTOT) # Fraction of full time employment, Wave 1
data$FRACFT2 <- data$EMPFT2/data$EMPTOT2 # Fraction of full time employment, Wave 2

data$ATMIN <- 0
data$ATMIN[data$WAGE_ST==4.25] <- 1 # At min wage indicator, Wave 1

data$NEWMIN <- 0
data$NEWMIN[data$WAGE_ST2==5.05] <- 1 #At min wave indicator, Wave 2

data$ICODE[data$NJ==0] <- "PA Store"
data$ICODE[data$NJ==1 & data$WAGE_ST==4.25] <- "NJ Store, Low-Wage"
data$ICODE[data$NJ==1 & data$WAGE_ST>=5.00] <- "NJ Store, Hi-Wage"
data$ICODE[data$NJ==1 & data$WAGE_ST>4.25&data$WAGE_ST<5] <- "NJ Store, Med-Wage"
data$ICODE[data$NJ==1 & data$WAGE_ST<4.25] <- "NJ Store, Bad Wage"
```

Similar to the balance checks from last week's discussion of regression discontinuity designs, Card and Krueger (1994) take the time to compare the means of several key variables between New Jersey and Pennsylvania, in Wave 1 and Wave 2. This is displayed in Table 2.

```r
a1 <- mean(na.omit(data$BK[data$NJ==1]))
a2 <- mean(na.omit(data$BK[data$NJ==0]))

b1 <- mean(na.omit(data$KFC[data$NJ==1]))
b2 <- mean(na.omit(data$KFC[data$NJ==0]))

c1 <- mean(na.omit(data$ROYS[data$NJ==1]))
c2 <- mean(na.omit(data$ROYS[data$NJ==0]))

d1 <- mean(na.omit(data$WENDYS[data$NJ==1]))
d2 <- mean(na.omit(data$WENDYS[data$NJ==0]))

e1 <- mean(na.omit(data$CO_OWNED[data$NJ==1]))
e2 <- mean(na.omit(data$CO_OWNED[data$NJ==0]))

t1 <- summary(lm(BK ~ NJ, data=data))$coefficients[6]
t2 <- summary(lm(KFC ~ NJ, data=data))$coefficients[6]
t3 <- summary(lm(ROYS ~ NJ, data=data))$coefficients[6]
t4 <- summary(lm(WENDYS ~ NJ, data=data))$coefficients[6]
t5 <- summary(lm(CO_OWNED ~ NJ, data=data))$coefficients[6]


Tab21 <- matrix(c(a1, a2, b1, b2, c1, c2, d1, d2, e1, e2), byrow=T, nrow=5)
Tab21 <- cbind(Tab21, c(t1, t2, t3, t4, t5))
colnames(Tab21) <- c("NJ", "PA", "t")
```

TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in: NJ	Stores in: PA	t[a]
1. *Distribution of Store Types (percentages):*			
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. *Means in Wave 1:*			
a. FTE employment	20.4	23.3	−2.0
	(0.51)	(1.35)	
b. Percentage full-time employees	32.8	35.0	−0.7
	(1.3)	(2.7)	
c. Starting wage	4.61	4.63	−0.4
	(0.02)	(0.04)	
d. Wage = $4.25 (percentage)	30.5	32.9	−0.4
	(2.5)	(5.3)	
e. Price of full meal	3.35	3.04	4.0
	(0.04)	(0.07)	
f. Hours open (weekday)	14.4	14.5	−0.3
	(0.2)	(0.3)	
g. Recruiting bonus	23.6	29.1	−1.0
	(2.3)	(5.1)	
3. *Means in Wave 2:*			
a. FTE employment	21.0	21.2	−0.2
	(0.52)	(0.94)	
b. Percentage full-time employees	35.9	30.4	1.8
	(1.4)	(2.8)	
c. Starting wage	5.08	4.62	10.8
	(0.01)	(0.04)	
d. Wage = $4.25 (percentage)	0.0	25.3	—
		(4.9)	
e. Wage = $5.05 (percentage)	85.2	1.3	36.1
	(2.0)	(1.3)	
f. Price of full meal	3.41	3.03	5.0
	(0.04)	(0.07)	
g. Hours open (weekday)	14.4	14.7	−0.8
	(0.2)	(0.3)	
h. Recruiting bonus	20.3	23.4	−0.6
	(2.3)	(4.9)	

Notes: See text for definitions. Standard errors are given in parentheses.
[a]Test of equality of means in New Jersey and Pennsylvania.

Figure 2:

```r
rownames(Tab21) <- c("Burger King", "KFC", "Roy Rogers", "Wendy's", "Company-Owned")
kable(Tab21, caption="Table 2.1: Distribution of Store Types (percentages)")
```

Table 2: Table 2.1: Distribution of Store Types (percentages)

	NJ	PA	t
Burger King	0.4108761	0.4430380	-0.5198108
KFC	0.2054381	0.1518987	1.0778194
Roy Rogers	0.2477341	0.2151899	0.6060797
Wendy's	0.1359517	0.1898734	-1.2175859
Company-Owned	0.3413897	0.3544304	-0.2187241

Table 2.1 is meant to show that the stores in their sample have similar average food prices, store hours, and employment levels.

```r
a1 <- mean(na.omit(data$EMPTOT[data$NJ==1]))
a2 <- mean(na.omit(data$EMPTOT[data$NJ==0]))

b1 <- mean(na.omit(data$FRACFT[data$NJ==1]))
b2 <- mean(na.omit(data$FRACFT[data$NJ==0]))

c1 <- mean(na.omit(data$WAGE_ST[data$NJ==1]))
c2 <- mean(na.omit(data$WAGE_ST[data$NJ==0]))

d1 <- mean(na.omit(data$ATMIN[data$NJ==1]))
d2 <- mean(na.omit(data$ATMIN[data$NJ==0]))

e1 <- mean(na.omit(data$PMEAL[data$NJ==1]))
e2 <- mean(na.omit(data$PMEAL[data$NJ==0]))

f1 <- mean(na.omit(data$HRSOPEN[data$NJ==1]))
f2 <- mean(na.omit(data$HRSOPEN[data$NJ==0]))

g1 <- mean(na.omit(data$BONUS[data$NJ==1]))
g2 <- mean(na.omit(data$BONUS[data$NJ==0]))

t1 <- summary(lm(EMPTOT ~ NJ, data=data))$coefficients[6]
t2 <- summary(lm(FRACFT ~ NJ, data=data))$coefficients[6]
t3 <- summary(lm(WAGE_ST ~ NJ, data=data))$coefficients[6]
t4 <- summary(lm(ATMIN ~ NJ, data=data))$coefficients[6]
t5 <- summary(lm(PMEAL ~ NJ, data=data))$coefficients[6]
t6 <- summary(lm(HRSOPEN ~ NJ, data=data))$coefficients[6]
t7 <- summary(lm(BONUS ~ NJ, data=data))$coefficients[6]

Tab22 <- matrix(c(a1, a2, b1, b2, c1, c2, d1, d2, e1, e2, f1, f2, g1, g2), byrow=T, nrow=7)
Tab22 <- cbind(Tab22, c(t1, t2, t3, t4, t5, t6, t7))
colnames(Tab22) <- c("NJ", "PA", "t")
rownames(Tab22) <- c("FTE employment", "Percentage full-time employees", "Starting wage", "Wage=$4.25 (
kable(Tab22, caption="Table 2.2: Means in Wave 1")
```

Table 3: Table 2.2: Means in Wave 1

	NJ	PA	t
FTE employment	20.4394081	23.3311688	-2.3506289
Percentage full-time employees	0.3284887	0.3503780	-0.7252299
Starting wage	4.6121338	4.6301316	-0.4052694
Wage=$4.25 (percentage)	0.3051360	0.3291139	-0.4132053
Price of full meal	3.3510611	3.0423684	3.7916935
Hours open (weekday)	14.4184290	14.5253165	-0.3034418
Recruiting bonus	0.2356495	0.2911392	-1.0272850

Table 2.2 shows that average employment was similar per store in New Jersey compared to Pennsylvania. Starting wages were also similar, and there were no significant cross-state differences in average hours of operation, the fraction of full-time workers, and the prevalance of bonus programs to recruit new workers, although the average price of a full meal (medium soda, small fries, and entree) were significantly higher in New Jersey.

```
a1 <- mean(na.omit(data$EMPTOT2[data$NJ==1]))
a2 <- mean(na.omit(data$EMPTOT2[data$NJ==0]))

b1 <- mean(na.omit(data$FRACFT2[data$NJ==1]))
b2 <- mean(na.omit(data$FRACFT2[data$NJ==0]))

c1 <- mean(na.omit(data$WAGE_ST2[data$NJ==1]))
c2 <- mean(na.omit(data$WAGE_ST2[data$NJ==0]))

d1 <- mean(na.omit(data$NEWMIN[data$NJ==1]))
d2 <- mean(na.omit(data$NEWMIN[data$NJ==0]))

e1 <- mean(na.omit(data$PMEAL2[data$NJ==1]))
e2 <- mean(na.omit(data$PMEAL2[data$NJ==0]))

f1 <- mean(na.omit(data$HRSOPEN2[data$NJ==1]))
f2 <- mean(na.omit(data$HRSOPEN2[data$NJ==0]))

g1 <- mean(na.omit(data$SPECIAL2[data$NJ==1]))
g2 <- mean(na.omit(data$SPECIAL2[data$NJ==0]))

t1 <- summary(lm(EMPTOT2 ~ NJ, data=data))$coefficients[6]
t2 <- summary(lm(FRACFT2 ~ NJ, data=data))$coefficients[6]
t3 <- summary(lm(WAGE_ST2 ~ NJ, data=data))$coefficients[6]
t4 <- summary(lm(NEWMIN ~ NJ, data=data))$coefficients[6]
t5 <- summary(lm(PMEAL2 ~ NJ, data=data))$coefficients[6]
t6 <- summary(lm(HRSOPEN2 ~ NJ, data=data))$coefficients[6]
t7 <- summary(lm(SPECIAL2 ~ NJ, data=data))$coefficients[6]

Tab22 <- matrix(c(a1, a2, b1, b2, c1, c2, d1, d2, e1, e2, f1, f2, g1, g2), byrow=T, nrow=7)
Tab22 <- cbind(Tab22, c(t1, t2, t3, t4, t5, t6, t7))
colnames(Tab22) <- c("NJ", "PA", "t")
rownames(Tab22) <- c("FTE employment", "Percentage full-time employees", "Starting wage", "Wage=$4.25 (
kable(Tab22, caption="Table 2.3: Means in Wave 2")
```

Table 4: Table 2.3: Means in Wave 2

	NJ	PA	t
FTE employment	21.0274295	21.1655844	-0.1194924
Percentage full-time employees	0.3586555	0.3038485	1.7530818
Starting wage	5.0808491	4.6174648	19.7141728
Wage=$4.25 (percentage)	0.8549849	0.0126582	20.9596502
Price of full meal	3.4147541	3.0266197	4.7261223
Hours open (weekday)	14.4197819	14.6538462	-0.6731678
Recruiting bonus	0.2031746	0.2337662	-0.5903938

Table 2.3 shows that the average starting wage in New Jersey increased by 10 percent following the rise of the minimum wage. By wave 2, virtually all restaurants in New Jersey that had been paying less than $5.05 per hour reported a starting wage equal to the new rate. Only two other variables showed a relative change between waves, the fraction of full-time employees and the price of a meal.

The authors also find that stores with missing data on employment, wages, and prices are similar in other respects to stores with complete data.

These checks are meant to show that fast food restaurants in Pennsylvania, just across the Delaware River, are an appropriate control group for fast food restaurants in New Jersey.

The difference-in-differences estimator is implemented below.

```
EMPTOTFULL <- c(data$EMPTOT, data$EMPTOT2)
NJFULL <- c(data$NJ, data$NJ)
WAVE2FULL <- c(rep(0,410), rep(1, 410))

DDdata <- as.data.frame(cbind(EMPTOTFULL, NJFULL, WAVE2FULL))

DDdata[1:15,]
```

```
##    EMPTOTFULL NJFULL WAVE2FULL
## 1       24.75      1         0
## 2       26.75      1         0
## 3       24.00      1         0
## 4          NA      1         0
## 5       18.00      1         0
## 6       28.00      1         0
## 7       11.50      1         0
## 8       18.50      0         0
## 9       10.00      1         0
## 10      11.25      1         0
## 11      12.00      1         0
## 12      11.00      1         0
## 13      14.00      1         0
## 14       9.25      1         0
## 15      29.50      1         0
```

```
MODiD <- lm(EMPTOTFULL ~ NJFULL + WAVE2FULL + NJFULL*WAVE2FULL, data=DDdata)

summary(MODiD)
```

```
##
## Call:
## lm(formula = EMPTOTFULL ~ NJFULL + WAVE2FULL + NJFULL * WAVE2FULL,
```

```
##     data = DDdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.166  -6.439  -1.027   4.473  64.561
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        23.331      1.072  21.767   <2e-16 ***
## NJFULL             -2.892      1.194  -2.423   0.0156 *
## WAVE2FULL          -2.166      1.516  -1.429   0.1535
## NJFULL:WAVE2FULL    2.754      1.688   1.631   0.1033
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.406 on 790 degrees of freedom
##   (26 observations deleted due to missingness)
## Multiple R-squared:  0.007401,   Adjusted R-squared:  0.003632
## F-statistic: 1.964 on 3 and 790 DF,  p-value: 0.118
```

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

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	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.

[a]Stores 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 ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Figure 3:

The results show that the change in employment in New Jersey was actually larger than the change in employment in Pennsylvania. However, the estimates are statistically insignificant, and Card and Krueger conclude that there is no evidence of an employment decline after the minimum wage increase.

Difference-in-differences and Fixed Effects

Difference-in-differences estimation is a version of fixed effects estimation that uses aggregate data. Fixed effects regressions require panel data or repeated measurements on the same unit over time. But the regressor of interest may only vary at an aggregated level.

In the Card and Kreuger example, the minimum wage changes occurs at the the state level, although employment is observed at the restaurant level. Restaurants are nested within states, with the treatment occuring at the state level.

This can be distinguished from a country-year panel data analysis, where the regressor of interest, say institutional quality, varies by country, and the outcome of interest, say economic growth, varies by country also.

When there are more than two units and time periods, the model can be generalized as follows from Card (1992):

$$y_{ist} = \gamma_s + \lambda_t + \beta(FA_s \times d_t) + \epsilon_{ist}$$
$$y_{ist} = \gamma_s + \lambda_t + \beta_1 FA_s + \beta_2 d_t + \beta_3(FA_s \times d_t) + \epsilon_{ist}$$

γ_s are state dummies, λ_t are time dummies, and $(FA_s \times d_t)$ is an interaction term that multiplies the fraction of teenagers likely to be affected by a minimum wage increase in each state, and a dummy for observations after 1990, when the federal minimum wage increased from \$3.35 to \$3.80. y_{ist} means that employment varies by the individual across states over time.

β_1 is therefore the change in employment that is associated on average with a marginal increase in the fraction of teenagers likely to be affected by a minimum wage, holding constant the increase of the minimum wage.

β_2 is the change in employment that is associated on average with an increase in the minimum wage, holding constant the effect of an increase in the fraction of teengers likely to be affected by the law.

The estimate, β_3, is therefore the difference in the differences of employment that is associated on average with a marginal increase in the fraction of teenagers likely to be affected by the minimum wage increase, before and after the minimum wage increase.

This regression equation can be expanded to include both state-level and individual-level covariates.

$$y_{ist} = \gamma_s + \lambda_t + \beta_1 FA_s + \beta_2 d_t + \beta_3(FA_s \times d_t) + \Psi X_{ist} + \Phi W_{st} + \epsilon_{ist}$$

Going back to the Card and Kreuger example, it turns out that Pennsylvania may not be an appropriate control group for New Jersey. Why?

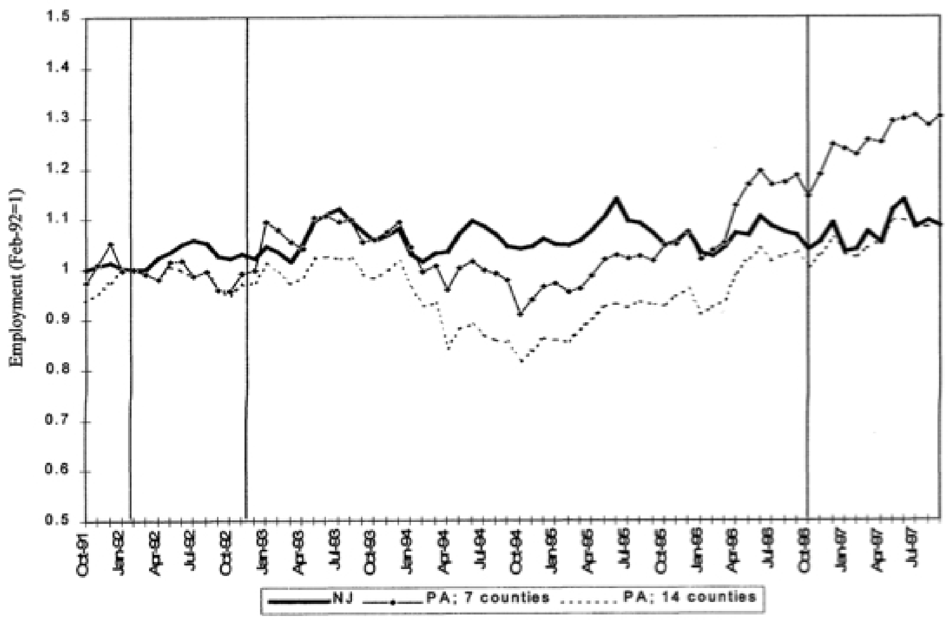

Figure 5.2.2: Employment in New Jersey and Pennsylvania fast-food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum-wage increase.

Figure 4: