cz363_final_dd

Chuyuan Zhong

2023-05-09

Consider a state government in the United States that administers a child support program. This program requires non-custodial parents (NCPs) to pay monthly child support payments to custodial parents of their children. If NCPs fail to make payments, they can fall into owing child support debt which can lead to incarceration.

Suppose that a state child support agency introduced a program to provide intensive case management services to NCPs so that they could help them find jobs to make child support payments and have better relationships with their non-custodial children and the custodial parents. The program was administered in selected local child support sites across the state, and in these sites, the program was rolled out at varying dates.

Please download final_data_dd.csv. This dataset contains the following site-level panel data (each row in the dataset corresponds to an observaRon for a given site at a given calendar quarter):

- **site_id** = identification number for a local child support office site
- **year_qtr** = calendar quarter of observation (eg 2011.25 is the second quarter in 2011)
- **treatment** = indicator equal to 1 if the site was selected for treatment
- **treatment_year_qtr** = calendar quarter the site began the program if selected for treatment
- **ncp_emp_rate** = employment rate for NCPs served by the site in the calendar quarter
- **smom_emp_rate** = employment rate for single mothers served by the site in the calendar quarter
- **ncp_wdebt** = fraction of NCPs served by the site in the calendar quarter that owe child support debt
- **Nncp** = number of NCPs served by the site in the calendar quarter

Explain how a Difference-in-Difference research design can be conducted in this setting to examine the impacts of the intensive case management services on the likelihood of owing child support debt and employment of NCPs and single mothers.

How can selected sites be compared with non-selected sites? What assumptions are necessary to estimate causal impacts of the program on outcomes, and is there any evidence to support these assumptions? What are the impacts of the intensive case management services on sites' outcomes? Do the intensive case management services increase employment? Describe your analysis, assumptions, results, and conclusions.

```
# read the dataset
df <- read.csv('/Users/Chuyuan/Downloads/final_data_dd.csv')
head(df)
```

```
##   site_id year_qtr treatment treatment_year_qtr ncp_wdebt ncp_emp_rate
## 1  186000  2011.00         0                  0 0.4881410    0.3865531
## 2  186000  2011.25         0                  0 0.4250312    0.2347802
## 3  186000  2011.50         0                  0 0.3876731    0.3146987
## 4  186000  2011.75         0                  0 0.4888856    0.3694253
```

```
## 5  186000  2012.00         0                    0 0.3735682    0.2986981
## 6  186000  2012.25         0                    0 0.4233907    0.4062268
##   smom_emp_rate Nncp
## 1     0.5244962   68
## 2     0.5858592   61
## 3     0.6279745   51
## 4     0.6571974   27
## 5     0.7097299   33
## 6     0.6034863   63
```

```
## Check the dimensionality
cat("The dataset has", nrow(df), "observations and", ncol(df), "variables.")
```

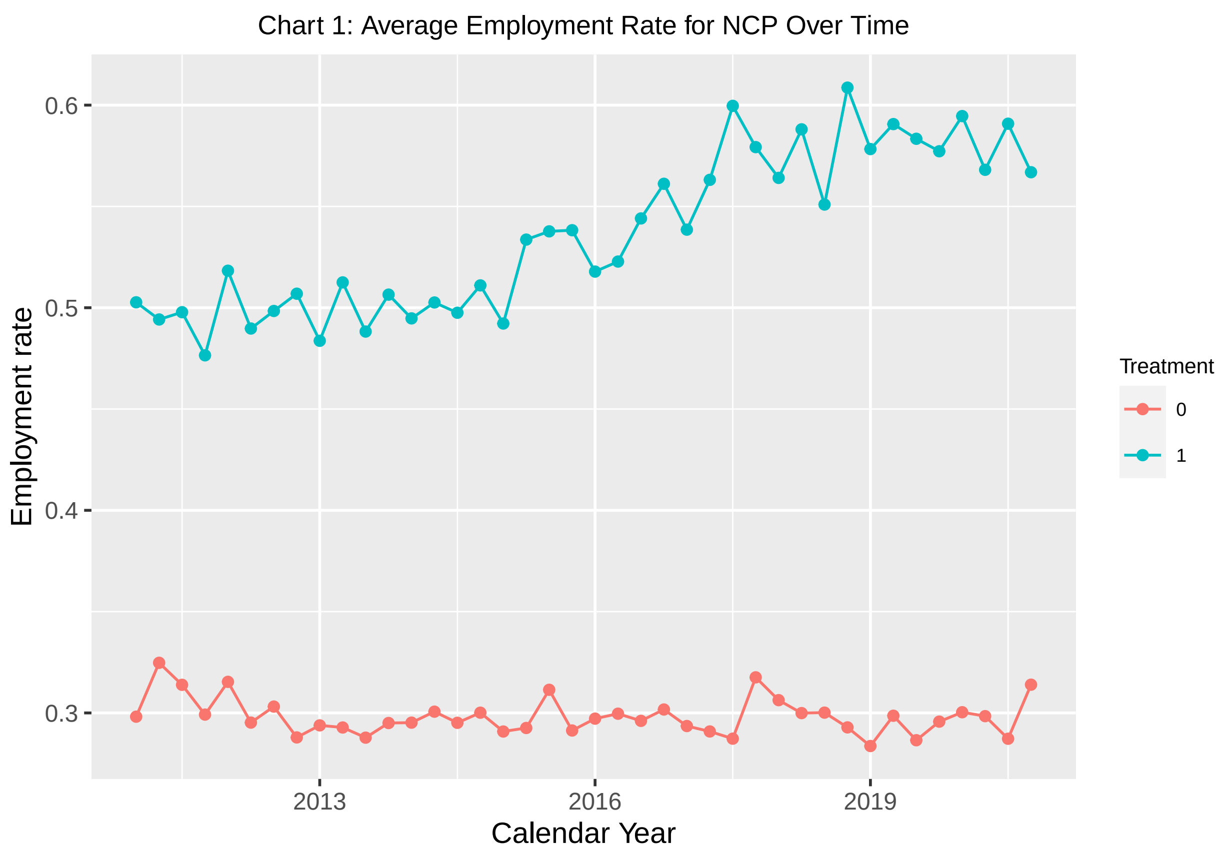
```
## The dataset has 7440 observations and 8 variables.
```

```
## Find unique number of year quarters
unique_quarters <- unique(df$treatment_year_qtr)
unique_quarters
```

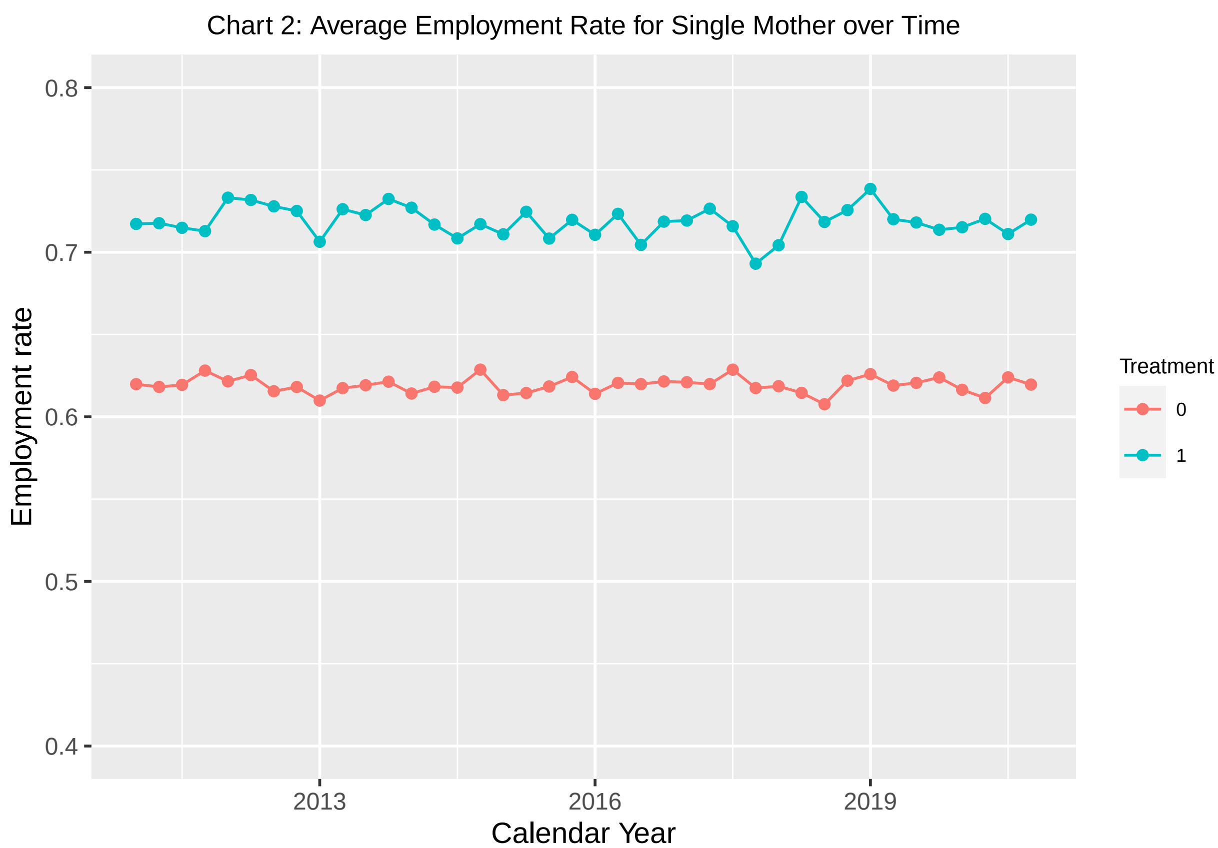
```
##  [1]    0.00 2015.50 2014.00 2016.00 2014.25 2014.75 2016.50 2015.00 2016.25
## [10] 2015.25 2016.75 2015.75 2014.50
```

```
# Find the average of employment rate and debt fraction
dft1 <- df %>%
  group_by(year_qtr, treatment) %>%
  summarise(ncp_emp_rate_mean = mean(ncp_emp_rate),
            smom_emp_rate_mean = mean(smom_emp_rate),
            ncp_wdebt_mean = mean(ncp_wdebt),
            Nncp_mean = mean(Nncp))
```

```
## Plot Average Employment rate for NCP
dft1 <- na.omit(dft1)
ggplot(dft1, aes(x = year_qtr, y = ncp_emp_rate_mean, color = factor(treatment))) +
  geom_line() +
  geom_point() +
  labs(title = "Chart 1: Average Employment Rate for NCP Over Time",
       x = "Calendar Year",
       y = "Employment rate",
       color = "Treatment") +
  theme(plot.title = element_text(size = 10, hjust=0.5),
        legend.title = element_text(size = 8),
        legend.text = element_text(size = 7))
```



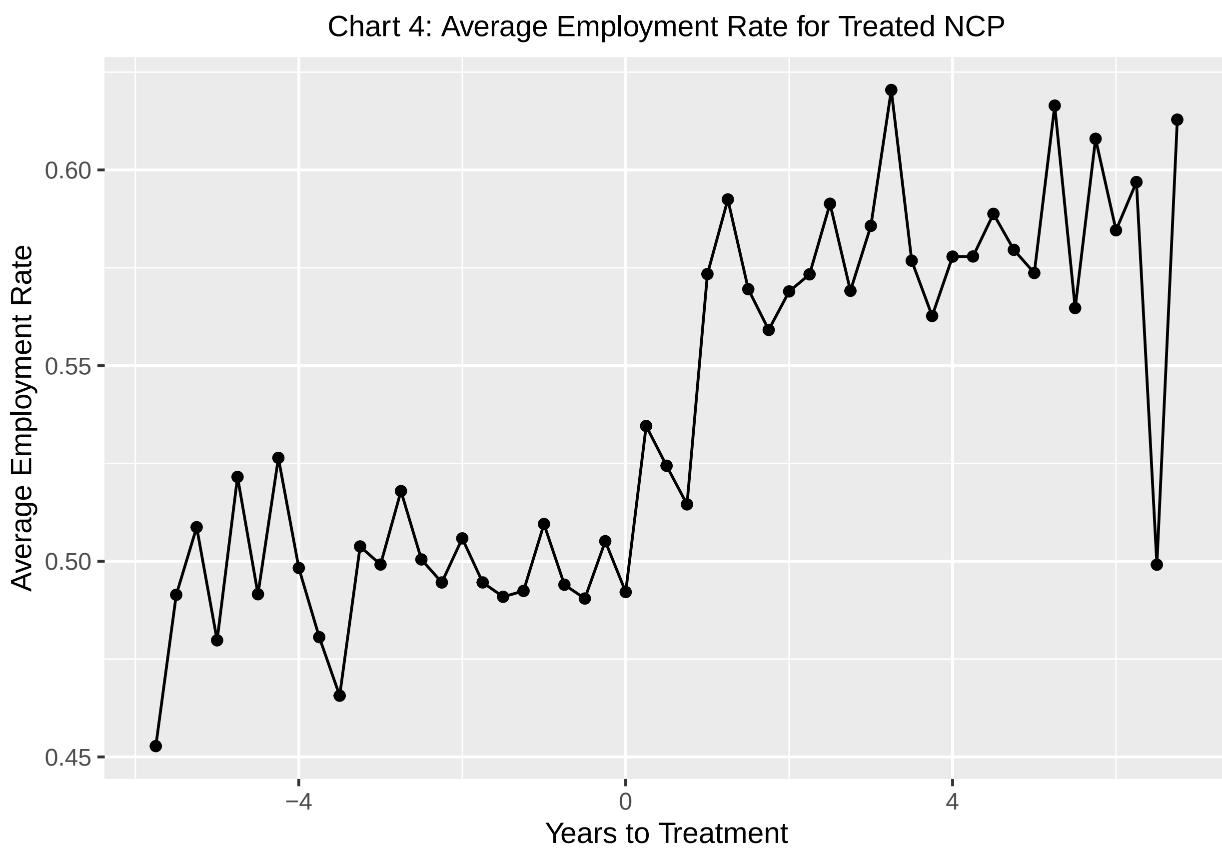

```
ggsave("emp_ncp.png")


## plot Average Employment rate for Single Mother
ggplot(dft1, aes(x = year_qtr, y = smom_emp_rate_mean, color = factor(treatment))) +
  geom_line() +
  geom_point() +
  labs(title = "Chart 2: Average Employment Rate for Single Mother over Time",
       x = "Calendar Year",
       y = "Employment rate",
       color = "Treatment") +
  theme(plot.title = element_text(size = 10, hjust=0.5),
        legend.title = element_text(size = 8),
        legend.text = element_text(size = 7)) +
  ylim(0.4, 0.8)
```



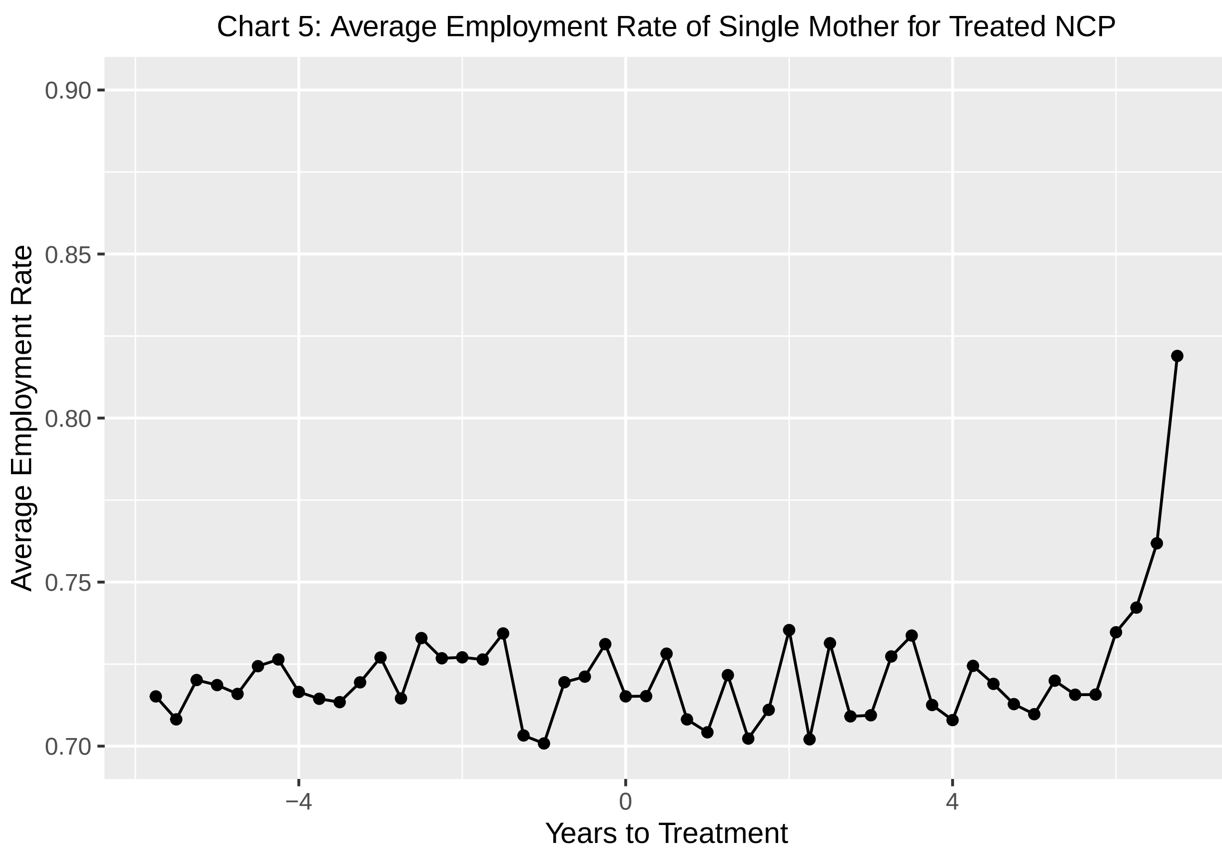

```
ggsave("emp_sm.png")

## Plot average NCP fractions that owe debt
ggplot(dft1, aes(x = year_qtr, y = ncp_wdebt_mean, color = factor(treatment))) +
  geom_line() +
  geom_point() +
  labs(title = "Chart 3: Average Debt Fraction rate over Time",
       x = "Calendar Year",
       y = "Fraction of owing debt",
       color = "Treatment") +
  theme(plot.title = element_text(size = 10, hjust=0.5),
        legend.title = element_text(size = 8),
        legend.text = element_text(size = 7)) +
  ylim(0.1, 0.5)
```


```
ggsave("debt_frac.png")
```

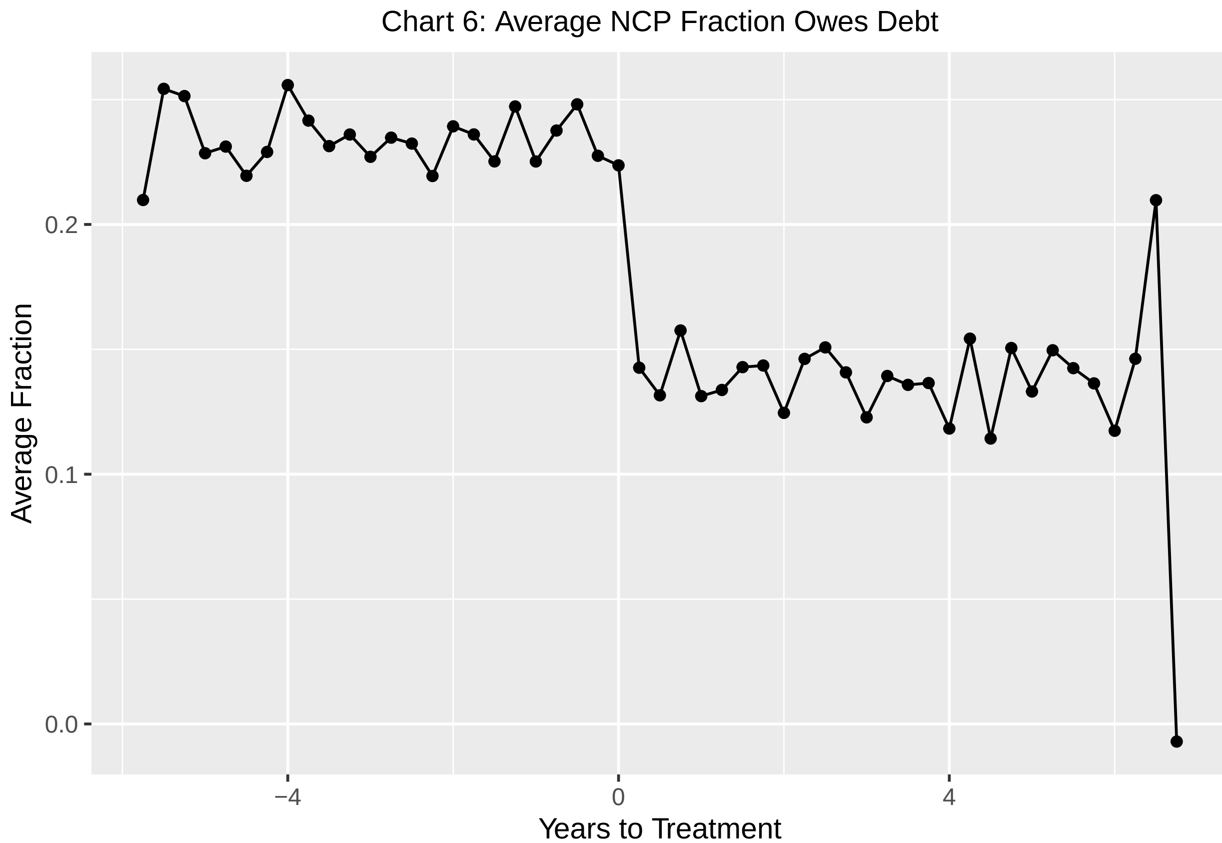
```
## create time_since_treatment variable
df$time_since_treatment <- ifelse(df$treatment==1,
                                  df$year_qtr-df$treatment_year_qtr,
                                  df$treatment)

## plot employment rate for treated NCPs
plotdata <- aggregate(df$ncp_emp_rate, list(df$time_since_treatment, df$treatment), FUN = mean)

ggplot(plotdata[plotdata$Group.2==1,], aes(x = Group.1, y = x)) +
  geom_point() +
  geom_line() +
  xlim(-5.75, 6.75)+
  labs(title = "Chart 4: Average Employment Rate for Treated NCP",
       x = "Years to Treatment",
       y = "Average Employment Rate")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))
```




```
ggsave("emp_ncp_treated.png")

## plot employment rate for treated mothers
plotdata <- aggregate(df$smom_emp_rate, list(df$time_since_treatment, df$treatment), FUN = mean)

ggplot(plotdata[plotdata$Group.2==1,], aes(x = Group.1, y = x)) +
  geom_point() +
  geom_line() +
  xlim(-5.75, 6.75)+
  labs(title = "Chart 5: Average Employment Rate of Single Mother for Treated NCP",
       x = "Years to Treatment",
       y = "Average Employment Rate")+
  theme(plot.title = element_text(size = 11, hjust = 0.5)) +
  ylim(0.7, 0.9)
```




```r
ggsave("emp_sm_treated.png")

## plot treated NCP fraction that owe child support debt
plotdata <- aggregate(df$ncp_wdebt, list(df$time_since_treatment, df$treatment), FUN = mean)

ggplot(plotdata[plotdata$Group.2==1,], aes(x = Group.1, y = x)) +
  geom_point() +
  geom_line() +
  labs(title = "Chart 6: Average NCP Fraction Owes Debt",
       x = "Years to Treatment",
       y = "Average Fraction")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))
```



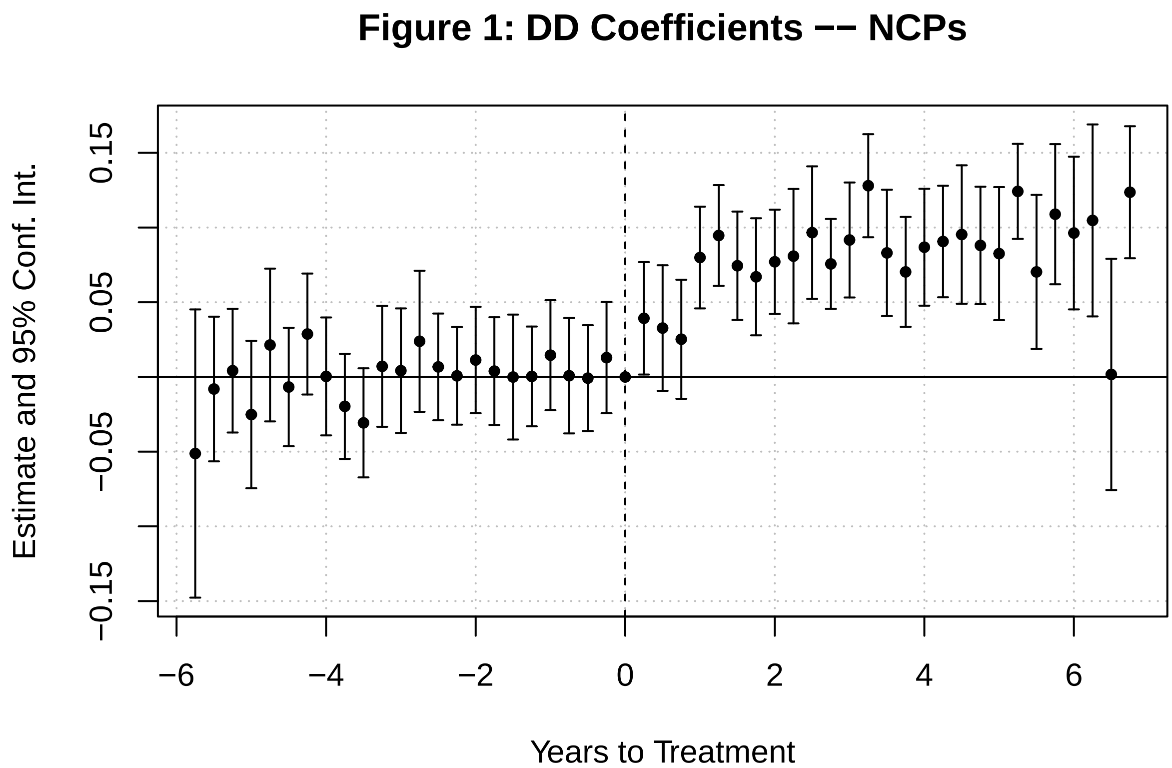

```
ggsave("debt_ncp_treated.png")
```

Now we use the full data (treatment and control teachers) to measure the treatment effect, following regression specification:

$$y_{in} = \alpha_0 + \alpha_1 N_i + \sum_{k=0} [\delta_k D_{ik}] + \gamma_s + \gamma_N + \gamma_t + error_{it}$$

where y_{it} denote average employment rate for NCP i in year t, N_i denote an indicator for being a treated NCP, D_{ik} denotes an indicator variable for being a treated NCP and having time since treatment = k, γ_N denotes numbers of NCPs fixed effects, γ_s denotes site fixed effects, and γ_t denotes calendar year quarter. Plot the δ_k coefficients over time since treatment. Interpret the results illustrated in the plot.

```
## plot coefficients for NCPs
ddreg = feols(ncp_emp_rate ~ i(time_since_treatment, treatment, ref = 0)
               | Nncp + year_qtr + site_id, data = df)
iplot(ddreg, xlab = 'Years to Treatment',
      main = 'Figure 1: DD Coefficients -- NCPs')
```



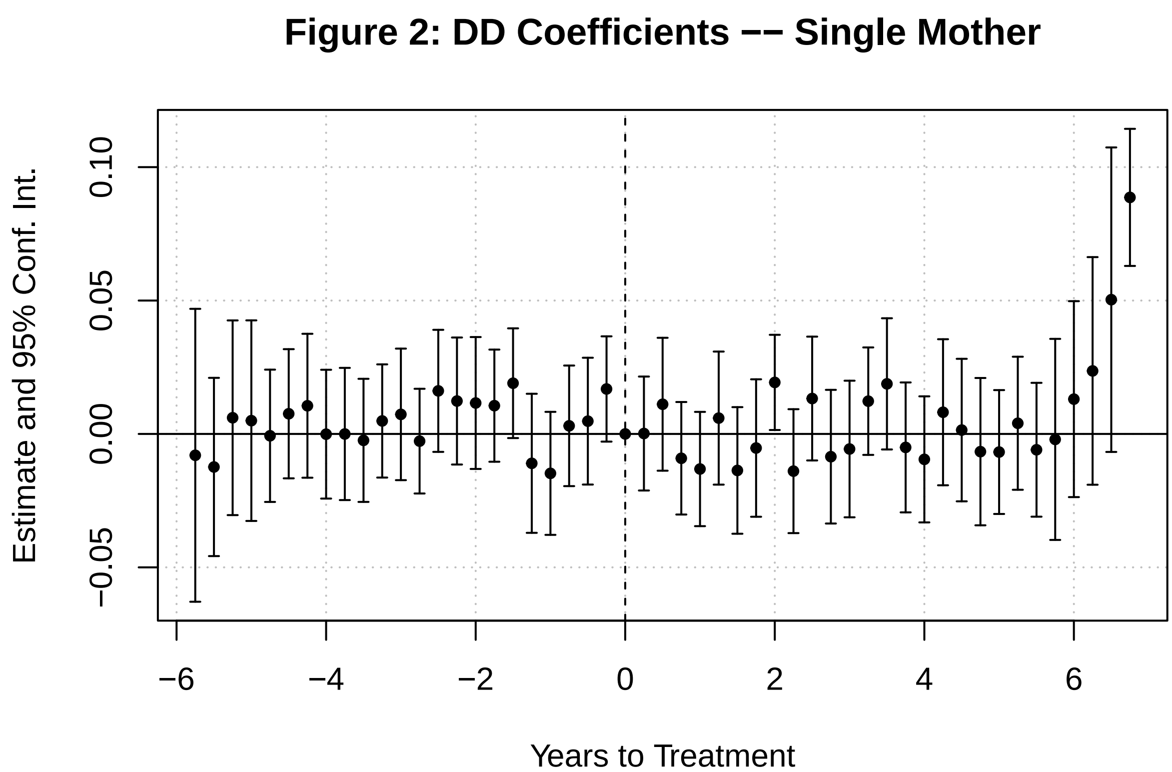

```r
dev.copy(png,"myfile.png")
```

```
## png
##   3
```

```r
dev.off()
```

```
## pdf
##   2
```

```r
## plot coeficients for single mother (control group)
ddreg = feols(smom_emp_rate ~ i(time_since_treatment, treatment, ref = 0)
              | Nncp + year_qtr + site_id, data = df)
iplot(ddreg, xlab = 'Years to Treatment',
      main = 'Figure 2: DD Coefficients -- Single Mother')
```




```
dev.copy(png,"sm_emp.png")
```

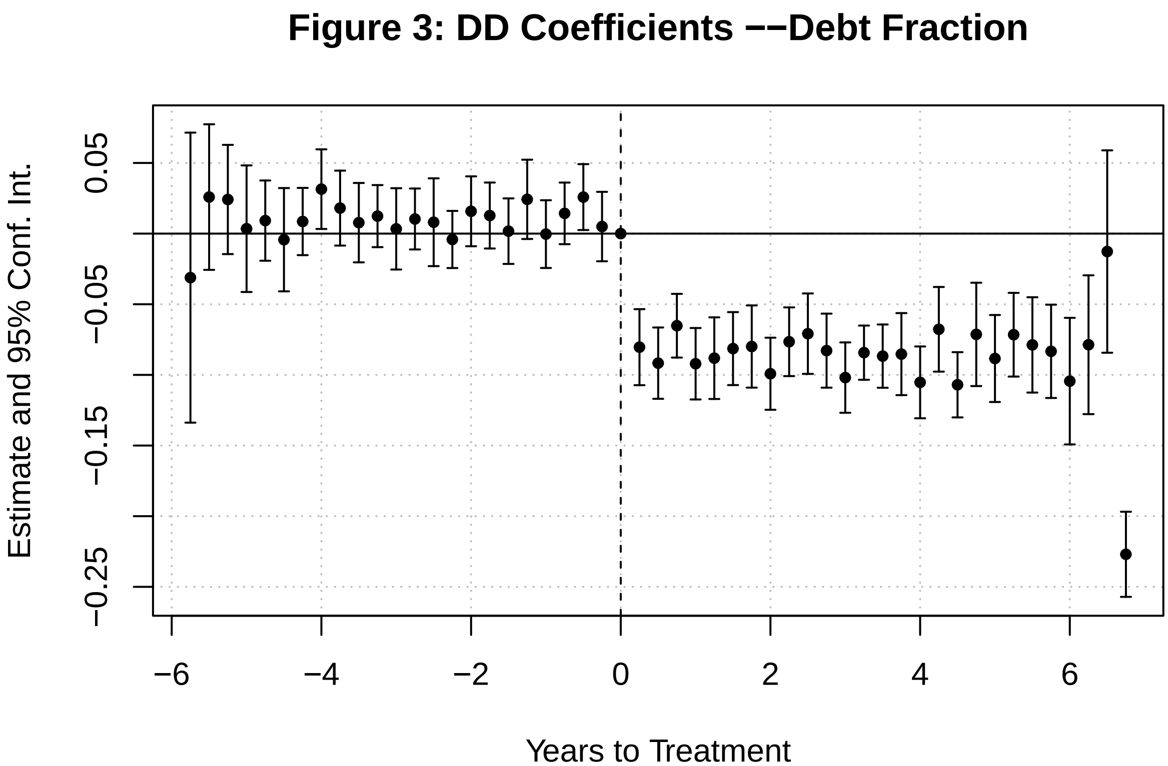
```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

Now we estimate the above regression specification in using ncp_wdebt as the outcome variable. Plot the δ_k coefficients over time since treatment

```
## plot coefficients for NCP fractions that owe debt
ddreg = feols(ncp_wdebt ~ i(time_since_treatment, treatment, ref = 0)
                | Nncp + year_qtr + site_id, data = df)
iplot(ddreg, xlab = 'Years to Treatment',
      main = 'Figure 3: DD Coefficients --Debt Fraction')
```




```
dev.copy(png,"debt.png")
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

Now we estimate how much increase in employment rate is associated with the change of NCPs fractions that owe debt.

```
# create after and TAfter to see the treatment effect
df$after <- ifelse(df$time_since_treatment > 0, 1, 0)
df$TAfter<- df$treatment*df$after
ta_reg <- feols(ncp_wdebt ~ TAfter  | Nncp + year_qtr + site_id, data = df)
ta_reg1 <- feols(ncp_emp_rate ~ TAfter | Nncp + year_qtr + site_id, data = df)
coefficients(ta_reg)[1]/(coefficients(ta_reg1)[1])
```

```
##    TAfter
## -1.289465
```

Therefore, one percent increase in employment rate is associated with 1.28 percent decrease of NCP fraction that owe child debt.