Difference-in-Difference Analysis: Employment Support to Non-Custodial Parents

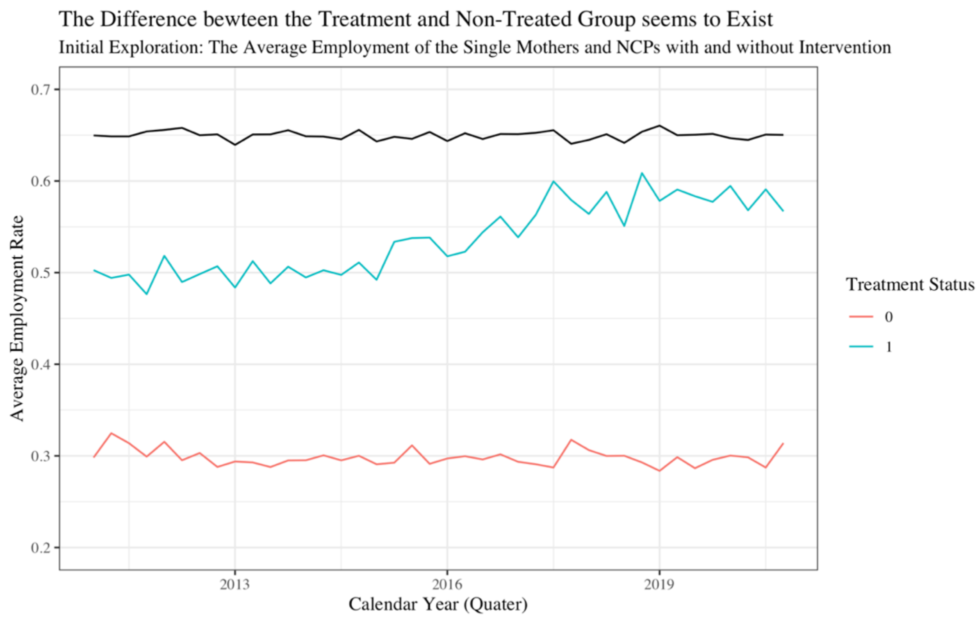
MinAh Kim

1) Setting up the Difference-in-Difference Analysis

This project aims to estimate how providing case management services to non-custodial parents improves their employment circumstances. Here, I want to use the difference-in-difference analysis to examine the project's impact without assigning treatment to randomized groups. To conduct a difference-in-difference analysis, I need panel data, which shows *if* and *when* the intervention happened for each individual. Furthermore, I require a control group, which is a comparing group that shows a stable parallel trend with the group of interest before the intervention happens. In this study, I can observe that the given data is panel data since it provides each site's treatment status (treatment) based on the calendar quarter (year_qrt). The control group will be single mothers, who are not subjected to the policy but are socially impacted similarly to the non-custodial parents. This parallel trend assumption is essential to ensure the causality of the program to the result.

Another condition to assume causality from the research is that the Non-Custodial Parents (NCPs) involved in the study should not anticipate that the intervention will happen so that they would not alter their behaviors due to this expectation. I cannot prove if this condition was held using the data, but it will be an underlying assumption for the study. I also know that each site starts its intervention in a different calendar quarter. The treatment_year_qtr attribute in the data set informs us when the treatment has begun in each site. Hence, this study will implement a generalized difference-in-difference model, using the time since the treatment in the analysis.

1-1) Initial Exploration: Is there a difference between the Treatment and Control?



Before kicking off the difference-in-difference analysis, I create a graph to have a general idea of the difference between the control and the treatment. In this graph, I average the employment rate of the single mothers and the NCPs in sites that apply the policy or not for each calendar year. The black line on this graph represents single mothers, while the blue and orange lines represent NCPs with and without treatment. From this graph, I can see that the difference between the single mothers and the non-treated NCPs tends to keep the 0.35 range stably. I can also observe that the difference between the NCPs group with and without treatment ranges from 0.2 to 0.3. However, each point of the calendar year is a combination of a different set of sites as the sites shift treatment status in separate points. For a more accurate insight, I will dive into the generalized difference-in-difference analysis.

2) Applying Generalized Difference-in-difference

2-1) Differences in NCP employment between the sites

To analyze the differences between the groups, I plan to apply a fixed effect regression to mathematically describe the relationship between the observation. In this analysis, I am comparing the employment outcome of the two groups: the site that received the intensive case management service (treatment) and the site that never received the intensive case management service (control). I also need to compare the changes within the sites that receive treatment based on the introduction of the service. Since all sites applied the policy at different moments, I need to create a variable, k, that records the relative time since the introduction of the policy.

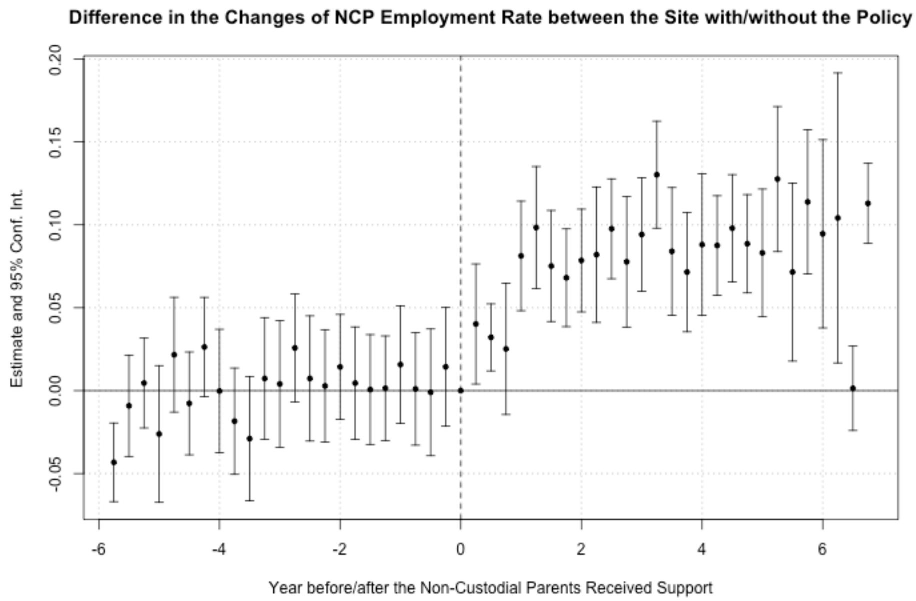
$$k = calendar\ time \text{ (year\_qtr)} - time\ of\ treatment \text{ (treatment\_year\_qtr)}$$

(k = 0 for the control sites since the site never implemented the policy)

Using this relative time indicator, I express the relationship between the NCP employment (y), the indicator for each observation (i), the treatment status (T), the change after the intervention for the control group (D), the control variables to capture the calendar year-wide changes (X) – such as economic recession –, and the error terms as an equation below:

$$\mathrm{y_{it}} = \alpha_0 + \alpha_1 \mathrm{T_i} + \sum_{k \neq 0} \beta_k D_{ik} + \delta_k T_i D_{ik} + \gamma X_{it} + \varepsilon_{it}$$

Here, I am interested in the term δ_k because it represents the difference between the changes in NCP employment between the site with and without policy implementation since the intervention. When I plot the coefficient δ_k, I find this coefficient has no effect until k reaches 0. This trend is within the boundary of my expectation because it aligns with the diff-in=diff assumption that the treatment and control group shows parallel trends. Then, when the k > 0, the coefficients become statistically significant, indicating that the confidence interval of the employment difference ranges around 0.10%. This number is lower than my initial exploration estimate, but both agree that the employment rate increased in the site that applies the policy.



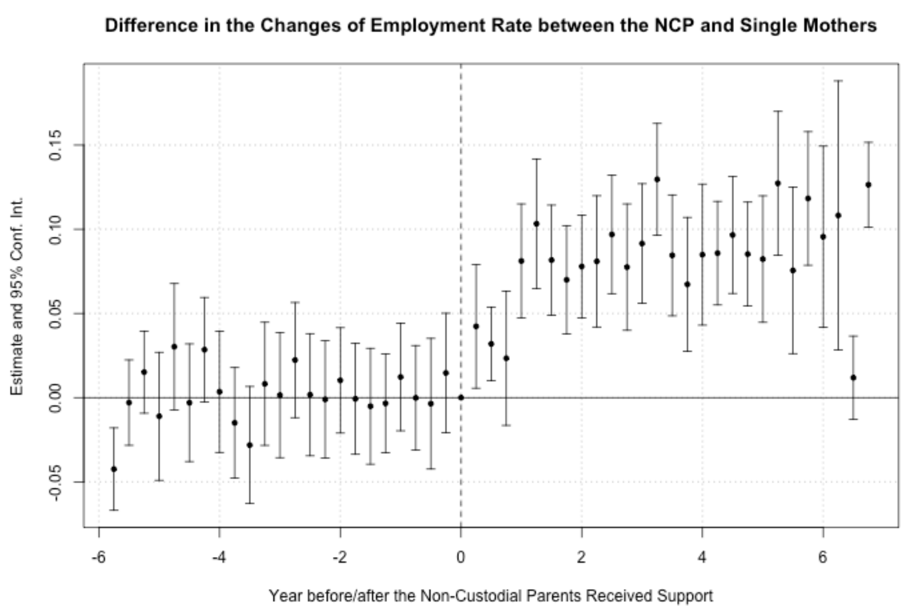
2-2) Differences in the employment rate of the NCP and Single Mothers in the Same site

One may still argue that the difference may be due to the difference between the sites instead of the policy. For instance, the sites that adopt the policy experience similar political changes in their leadership, which causes policy implementation. Then, the increase in the NCP's employment rate may not be the result of the intensive care management service but the change in the political leadership. Therefore, I run another fixed effect regression that compares the employment rate between the NCPs and single mothers in the site that implement the policy. Single mothers can be a good comparison within the site since the NCPs also tend to be unmarried parents but mostly men. Due to the resemblance in their social status, their employment status is likely to fluctuate similarly. I also check that the employment trend between the single mother and the NCP without the policy implementation tends to be similar through the initial exploration. If the difference from the first difference in difference analysis is due to the circumstance of the site, *not* the policy, the difference in the employment rate between the NCPs and the single mothers will be constant regardless of the time the intervention kicks in.

The equation that reflects the estimated relationship between the single mother and the NCP is the following:

$$\mathrm{Emp_{it}} = \alpha_0 + \alpha_1 \mathrm{\ Group_i} + \sum\nolimits_{k\neq 0} \beta_k D_{ik} + \delta_k Group_i D_{ik} + \gamma X_{it} + \varepsilon_{it}$$

Here, Emp refers to the employment rate. The variable Group equals 1 if the employment rate of the observation regards the NCP of the site. The same variable will be 0 if the employment rate of the observation is on the single mothers. After selecting the observation from the site that implemented the policy, I run this regression and observe the new δ_k. Below is the graph that illustrates the value of the δ_k.



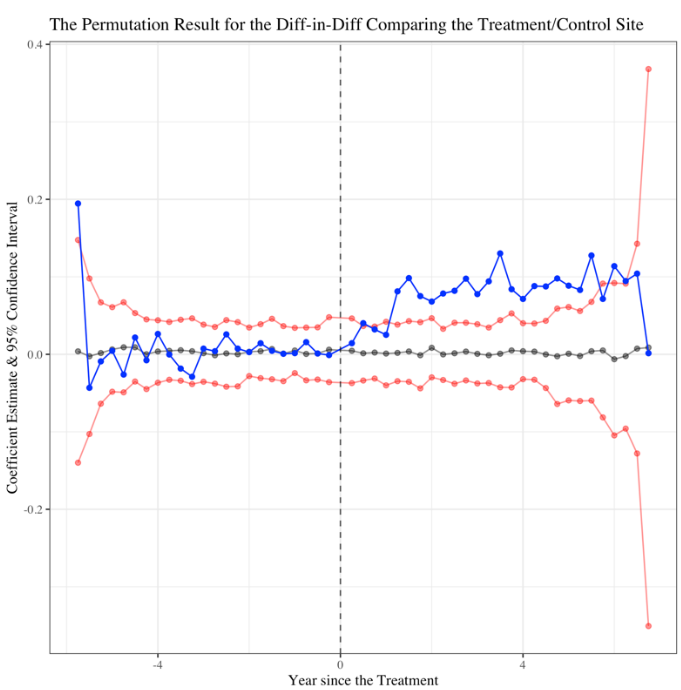
This graph shows that the coefficient becomes statistically significant after k = 0, particularly after k = 1. In other words, the employment rate difference between the NCPs and the single mothers widens after the intervention, especially after one year of implementation. The confidence ranges of the difference mainly locate around 0.10%. This result aligns with the conclusion of the first difference-in-difference analysis. Hence, I can conclude that the increase in the employment rate of the NCPs is due to the program that provides employment support.

3) Permutation Test

One may question if the founding from the two difference-in-difference analyses a result of random chance can be. It may be a coincidence that the control and the treatment group are divided in a way that their result is statistically significant. To this address this concern, I conduct the permutation test for each difference-in-difference analysis.

3-1) Site with Policy Intervention vs Site without Intervention

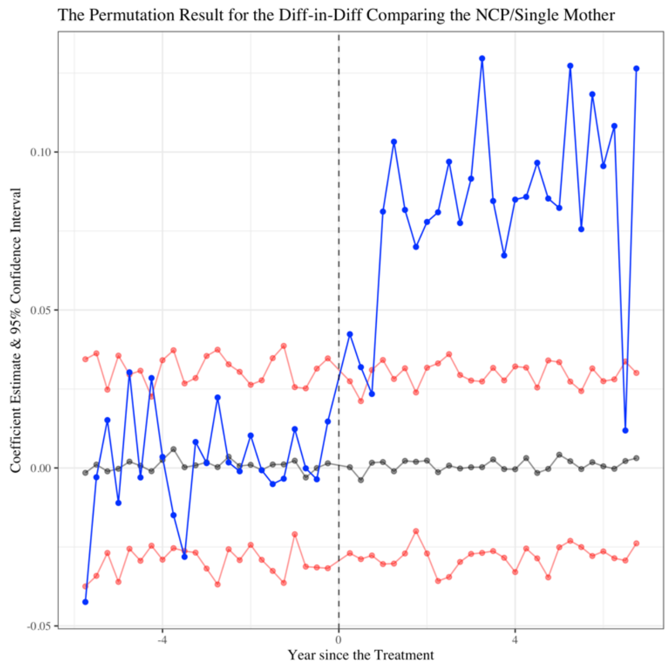
Here, I randomly assign the treatment status and the year of the treatment to each site so that the total number of sites treated in each year stays the same from the original analysis. Then, based on the random assignment, I average the coefficients of the same fixed effect regression that compares the sites with and without intervention. I repeat this process thirty times and compare the average of the thirty estimated coefficient values with the original coefficient. If the original result was due to random assignment, it is 95% likely that the coefficient discovered in the previous section will fall into the confidence interval.



The black line of the graph above shows the average of the thirty coefficients derived through the random assignment of control and treatment groups. The red lines represent the confidence interval of the values. Finally, the blue line illustrates the coefficient values of the difference-in-difference analysis using the original data set. Until the year from the treatment reaches 0, which is also statistically insignificant from the initial analysis, the original coefficient values stay near 0 and between the confidence intervals of the randomized coefficient values. However, after the intervention, most of the original coefficients leave the confidence interval of the randomized coefficients. This trend is particularly apparent after one year of the intervention. In other words, the probability of getting the original result due to random treatment and control assignment is less than 5%. Hence, I can argue that the conclusion from the difference-in-difference analysis between the site with and without intervention is very unlikely to be by chance.

3-2) NCPs vs Single Mothers in the Treated Sites

For the Difference in Differences analysis between the NCPs and the single mothers, all sites receive treatment at one point in the study. Thus, I randomly reassign the variable k, the years since the intervention, for each site and apply the same fixed effect regression from the difference in difference analysis for NCPs and single mothers. Similar to the first permutation test, I repeat the process 30 times to average the coefficient estimate produced by the randomization of k. In the graph below, I indicate the averages of the coefficient estimates based on random assignment in black, their confidence intervals in red, and the original coefficient values of the difference-in-difference analysis between the NCPs and single mothers in blue.



From the graph, I can observe that the coefficients from the original difference-in-difference analysis begin to exceed the confidence level after the year since the treatment becomes positive. Since the values outside the red line are less than 5% likely to result from randomization, I can interpret that the coefficients derived from the difference-in-difference analysis are unlikely to be a product of chance, particularly after the first year. Therefore, I constantly support the result of my difference-in-difference analysis using the NCPs and the single mothers of the treatment sites.

4) Estimation of the Policy Impact: NCP Employment & NCP with Child Support Debt

Another goal of this analysis is to learn how much the NCP employment rate increased due to the implementation of the policy. To calculate the impact, I create a variable that indicates only the observation after the intervention (post_t in the data set and T in the equation). Using this variable, I run the regression that represents the following mathematic equation when y is the NCP employment:

$$y_{it} = \alpha_0 + \beta_{\text{emp}} T_{it} + \gamma X_{it} + \varepsilon_{it}$$

When I fit the regression, I find that β_{emp} is around 0.267 with a significantly large t-value and a p-value lower than 0.05. From this result, I can infer that the policy intervention increases the average NCP employment rate by around 0.267.

This policy ultimately aims for the NCPs to have better relationships with their children and the custodial parents by helping them improve their capacity to pay child support. None of the difference-in-difference analyses in this paper proves that this policy causes the number of NCPs with child support debt to decrease. However, I can still assume this relationship and calculate how much the number of NCPs with child support debt decreases based on the increase in the NCP employment rate using the data set.

I derive the number of NCPs with child support debt (Nnpc_wdebt) by multiplying the number of NCPs of each site (Nncp) and the proportion of the NCPs with child support debt (ncp_wdebt). Then, I calculate the change in the number of NCPs with children support debt by fitting the fixed effect regression expressed below:

$$Nnpc_w debt_{it} = \alpha_0 + \beta_{nnpc} T_{it} + \gamma X_{it} + \varepsilon_{it}$$

(T equals 1 for the observation after the intervention and 0 for others)

The coefficient of this regression is around -4.745, implying that the number of NCPs with child support debt decreases by around -4.745 after the intervention. Since I already know that employment increases by 0.267 after the intervention, I can now calculate how much the increase in employment rate by intervention reduces the number of NCPs with child support debt. When I divide β_{nnpc} by β_{emp}, it represents the change in the number of NCP when the employment rate increases by 1. Therefore, I can divide this number by 100 and learn the decrease in the number of NCPs by 0.01 increase in the employment rate. The result of this calculation is around -0.178. Hence, if the causality between the number of NCPs with child support and the employment rate is valid, I can argue that the intervention also reduces the number of NCPs with child support debt to around -4.745 by reducing the number by -0.178 for each 0.01 increase in employment.