Data Cleaning

In the R Markdown file shared here, I have performed the following steps as part of a project to evaluate the impact of childcare provision on the mental well-being of mothers. The major task that had to be accomplished to achieve this was to create an index by aggregating the following parameters:

1. PHQ-4 well-being: quantifies depression and anxiety levels
2. Locus of Control: quantifies how much control individuals feel they have over their lives
3. Cantrill's Ladder: quantifies the current and expected future levels of life satisfaction

The PHQ-4 and Cantrill's Ladder variables were calculated as per conventions. The set of variables that required the most work was the Locus of Control. After standardizing the scores, a Principal Component Analysis and Factor Analysis were conducted to account for the individual impacts each of the standardized questions had on the total score. All of these indicators were aggregated into a composite index, which was standardized with respect to the mean of the control group.

To accommodate for attrition through the duration of the study, Lee Bounds were calculated and performed before running the regressions.

Regressions and Findings

After the creation of the well-being index, for each stage of the study, the variable was regressed on the treatment indicator and controls. The controls were determined through a LASSO test. Since there were two types of treatment, this was done at the pooled treatment and separated treatment levels. Fixed effects were added to the regressions to control for unobservable changes by location.

It was found that the treatment effect positively impacted the mental health of mothers while it was provided at the pooled (0.170 points) and separated levels (0.250 and 0.150 points for treatments 1 and 2, respectively). This is supported by the Lee Bounds being above 0. However, these effects dissipated within 3 months of the treatment ending, which shows that they didn't last. The Lee Bounds include 0 in their range, which affirms this conclusion.

To strengthen the confidence in these conclusions, a panel model and blank regression models were run without fixed effects and controls (except for lagging effects). The coefficients and statistical significance from this series were consistent with the previous models.