ECONOMETRIC ANALYSIS REPORT FOR A LANGUAGE PROFICIENCY PROGRAM ON STUDENTS' STANDARDIZED LANGUAGE SCORES

Author list

William Banda

1.0 Introduction

This report presents the results of an econometric analysis conducted to evaluate the impact of a language proficiency program implemented in two states by the World Bank. The program involved distributing books to certain schools (treatment schools) while leaving others without books (control schools). The primary objective is to assess the effect of this intervention on students' language proficiency, measured through standardized language scores.

2.0 Data Overview

Data was collected from two states, with each dataset containing information on students, including their treatment status, gender, age, and standardized language scores. The key variables analyzed include:

- **TREATED**: Indicator variable for whether the school received books (1 = treated, 0 = control).

- **LANGUAGE_SCORE_STD**: Standardized language proficiency score ranging from -6 to 6.

3.0 Summary Statistics

The dataset contains information on key student characteristics, including gender, age, and treatment status. Overall, 54.9% of the students were female, with a total count of 11,575 female students. Among male students in treated schools, the average age was found to be 11 years. Additionally, an analysis of students aged 12 revealed that the school with the highest number of 12-year-old students had 18 such students. These summary statistics provide a snapshot of the demographic composition and age distribution across schools included in the study.

4.0 Regression Analysis

To estimate the effect of the program on language proficiency, a regression analysis was performed using the following model

$$LANGUAGE\_SCORE\_STD = \beta_0 + \beta_1 \, TREATED + \varepsilon$$

5.0 Results

Results for State 1

In State 1, the regression analysis produced an estimated coefficient for the treatment indicator (TREATED) of approximately 0.00337. This suggests that being in a treated school is associated with a very slight increase in standardized language scores compared to control schools. However, the standard error for this coefficient was 0.00781, resulting in a p-value of 0.6659. This high p-value indicates that the effect is not statistically significant, meaning we cannot reject the null hypothesis that there is no difference in language scores between treated and control schools in State 1.

Results for State 2

Conversely, in State 2, the regression results revealed an estimated coefficient for TREATED of approximately 0.132. This indicates a positive association between being in a treated school and higher standardized language scores. The standard error for this coefficient was calculated at 0.064, leading to a p-value of 0.038. This p-value is below the conventional threshold of 0.05, suggesting that the effect of the treatment on language scores in State 2 is statistically significant.

6.0 Conclusion

The analysis reveals contrasting results between the two states regarding the impact of the language proficiency program. In State 1, there was no significant improvement in language scores among treated schools compared to control schools. Conversely, in State 2, a significant positive effect was observed, indicating that the program may have been effective in enhancing language proficiency in that state.

Future research could explore potential factors contributing to these differences, such as variations in implementation fidelity, student demographics, or contextual factors influencing learning outcomes. This study contributes valuable insights into the effectiveness of educational interventions aimed at improving language proficiency among students.

7.0 Data Availability

The data sets that were used in this research including the R code script for the analysis are openly available at the following Github Repository: https://github.com/WilliamBanda/Econometric-Analysis