Cross Section Project: Difference-in-Difference Analysis

Samantha Scott[a]

[a] *Stellenbosch University, Cape Town, South Africa*

Keywords: Difference-in-difference analysis, PIRLS data, Child Support Grant

*Corresponding author: Samantha Scott
Email address: 20945043@sun.ac.za (Samantha Scott)

1. Introduction

The following paper is an investigation of the Child Support Grant (CSG) and its impact on learning outcomes in South Africa. More specifically, this paper aims to estimate the effect that broader eligibility criteria would have on reading outcomes of grade 4 learners between the years 2011 and 2016. The broader criteria was a result of an increase in age eligibility of children to under 18 years of age in 2012. To investigate the causal effect of this policy change on the reading scores of children, a difference-in-difference estimation is conducted. In the estimation, individuals who qualified for the grant in their respective years are described as the treatment group, whereas the individuals who lie above the threshold are described as the control group. This paper uses the Progress in International Reading Literacy Study (PIRLS) data for 2011 and 2016 to investigate the research question at hand, and although the investigation is subject to limitations, the results are still noteworthy. The paper concludes that the policy change, namely the rise in age eligibility, has positively affected the learning scores of the treatment group.

2. Difference-in-Differnce Estimation

Difference-in-difference methods are widely used in the field of Economics to conduct policy evaluation with observational data (Sant'Anna & Zhao 2020:101-102). A difference-in-difference estimation makes use of the outcome of a control group as a proxy for the outcome for the treatment group, if the treatment had not occurred. This method requires there to be a treatment group, who is not treated in period one but treated in period two, as well as a control group who is not treated in either period. To obtain the difference-in-difference estimate, it is required to obtain the difference in before and after the treatment for the treated, and the control (untreated) group. The difference between these two differences is known as the difference-in-difference estimate, which is depicted in mathematical form below:

$$\rho = (Y_{treated2016} - Y_{treated2011}) - (Y_{untreated2016} - Y_{untreated2011})$$

3. Data and Methodology

According to previous literature, there was an increase in the maximum income per caretaker per annum between 2011 and 2016, which could have resulted in an increase in the eligibility of the caretakers in the population. After 2008, the threshold placed on the national income per caretaker per annum by the government was R27 600 (Beukes, Jansen, Moses, & Yu, 2017:513). By 2016, this threshold rose to R39 600 per caretaker per annum to qualify for the grant (Bhorat & Köhler, 2020:23).

According to Statistics South Africa, expenditure per household in 2011 and 2016 for quintile two was R32 569 and R43 452, respectively (Statistics South Africa, 2015:) & (Statistics South Africa, 2017:). Although there was an increase in the threshold amount, the jump was not large enough for another quintile to qualify. Therefore, for this investigation, the treatment group in both 2011 and 2016 is quintile one.According to Beukes et al. (2017:513), the eligibility criteria that qualified mothers for the CSG became broader in 2012. This was due to a rise of the maximum age the child could be for the mother to be eligible. The increase in age was up to 18 years of age. This increase in the number of individuals eligible for the CSG may result in an improvement of learning outcomes, for example, reading outcomes. To investigate this research question, the data used is the Progress in International Reading Literacy Study (PIRLS) for South Africa in 2011 and 2016. The samples are made of 12 810 grade 4 learners and are assessments of reading comprehension of the learners. In 2011, the sample was nationally representative, but was only stratified by language. In 2016, the sample was also nationally representative, but was stratified by both language and province (Howie, Combrinck, Roux, Tshele, Mokoena, & McLeod Palane, 2017:2).

To approximate wealth for the creation of the treatment group, the possessions that learner's indicated having in their homes are used. In doing so, an asset index of wealth is created. From this variable, a peer quintile variable is formulates,distinguishing between the different wealth groups of the population. Using these quintiles, it may be assessed whether an individual would have qualified for the CSG in 2011, and in 2016. If they qualified for the grant, the individuals are described as the treatment group. The individuals who fall above the threshold, are described as the control group. Using the treatment variable described above, an OLS regression is run to estimate the effect of the quintile the child falls under on their reading outcomes. In this first model, the learner's age and gender is included. A simple difference-in-difference regression is then run. The model includes only the treatment and control groups, the year before and after treatment, as well as the interaction term of the year after treatment and the treated. The third and fourth models introduce controls to the simple difference-in-difference estimation. The controls that are added is the learner's age and gender.

4. Results and Discussion

In Figure 4.1 below, model (1) is an OLS regression. In this model, the coefficient on the treatment group is negative. This is because the model is estimating the effect of the learner being in quintile one on their reading outcomes. As such, the coefficient on treatment group does not provide an accurate view of impact of the policy change on the treatment group. The coefficient on the learner's age is also negative. This is because PIRLS assesses grade 4 learners. If a child is older than 10 years of age, it means the child could have either failed one or more years. This may be indicative of the child's lower learning ability. By controlling for gender in this model, it is evident that girls have higher reading scores than boys. Although controls have been added to the model, there may still be unobservables in the error term which may bias the coefficient on the learner's quintile.

In the figure, model (2) is a simple difference-in-difference estimation. This model indicates that there is a positive effect of the broader eligibility criteria of the CSG that was implemented in 2012 on the treatment group. This method allows a better understanding of the impact of the policy change on the affected group, as opposed to the OLS regression. In checking the robustness of the results from model (2), controls are added to the simple difference-in-difference estimation. In model (3), the learner's age is added. As a result, the coefficient on the interactive term remains positive, but slightly decreases. When another variable is added, namely gender, in model (4), the coefficient drops slightly more, but still remains positive. This indicates that the strength of the effect is not accurate, however, the direction of the effect is accurate.

	(1) reading_score	(2) reading_score	(3) reading_score	(4) reading_score
treated	-30.14^{***}			
	(5.801)			
student_age	-17.20^{***}		-16.40^{***}	-13.22^{***}
	(2.152)		(1.562)	(1.561)
female	35.74^{***}			36.97^{***}
	(2.790)			(2.249)
2016		-146.5^{***}	-143.8^{***}	-143.2^{***}
		(7.195)	(7.098)	(7.041)
treated		-50.69^{***}	-48.83^{***}	-48.73^{***}
		(5.576)	(5.521)	(5.438)
2016 x treated		31.79^{***}	28.47^{***}	25.94^{**}
		(8.119)	(8.293)	(8.178)
_cons	558.8^{***}	470.8^{***}	643.5^{***}	592.1^{***}
	(24.24)	(5.298)	(18.06)	(18.56)
N	28162	28554	28236	28162
R^2	0.055	0.331	0.348	0.369
adj. R^2				

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 4.1: Regression Table

Figure 4.2 depicts the decrease in reading scores from 2011 to 2016, for the control group as well as the treatment group. As seen in the figure below, the reading scores for the control group are greater than those of the treatment group in both 2011 and 2016. However, the decrease in reading scores for the control group is larger than that of the the treatment group. Figure 4.2 therefore justifies the positive result presented by the difference-in-difference estimations, as the difference between the differences in treated and in untreated would be positive.

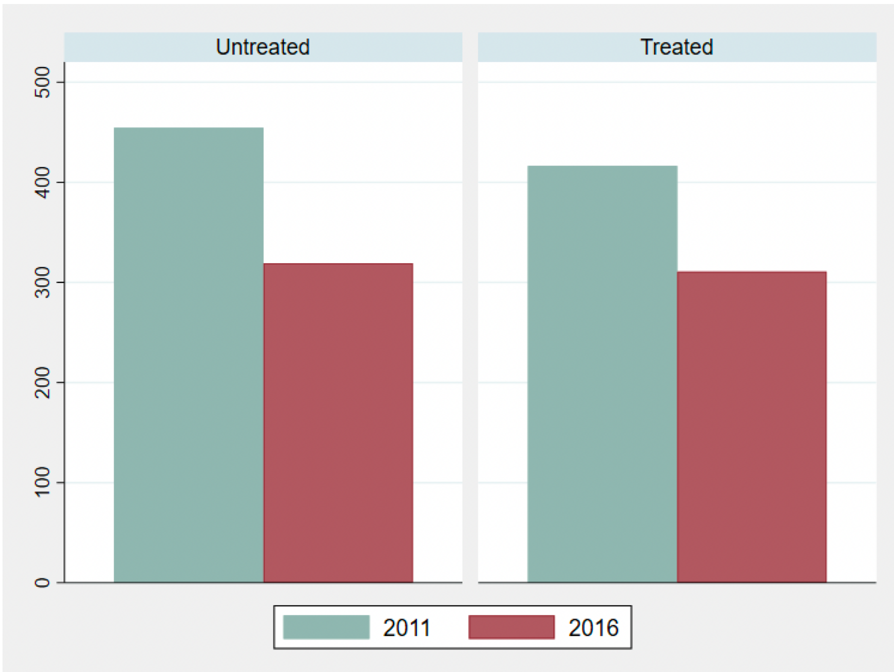

Figure 4.2: Reading Scores: Untreated vs.Treated

Figures 4.3 and 4.4 depict the reading scores per quintile in 2011 and 2016, respectively. As seen in the graphs below, the reading scores of the population drops over all quintiles. This indicates that the drop in reading scores for quintile one is not unique to the rest of the population. Using the graphs below, it is evident that the percentage change of the mean of reading scores between 2011 and 2016 is the lowest for quintile one in comparison to the other quintiles, as the difference in the average reading score of quintile one between 2011 and 2016 is 105. This value increases for each quintile.

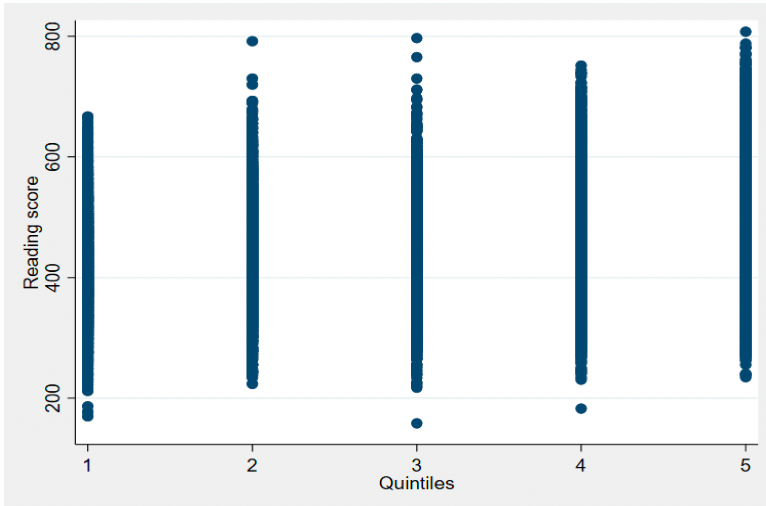

Figure 4.3: Reading Scores per Quintile in 2011

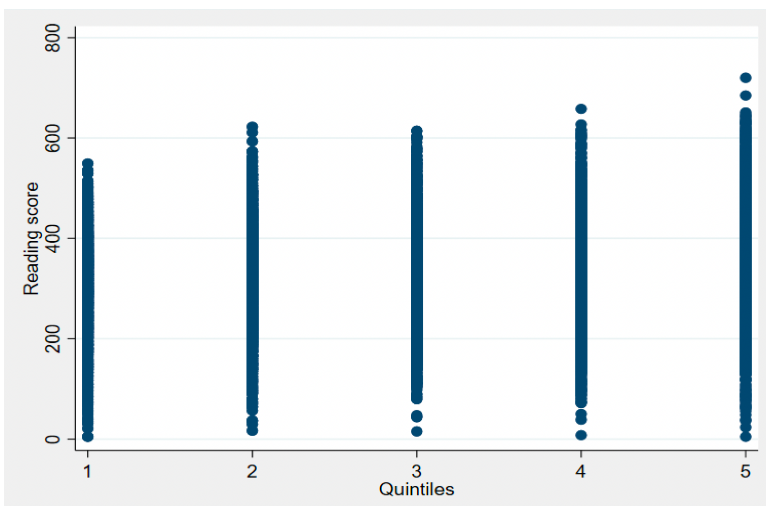

Figure 4.4: Reading Scores per Quintile in 2016

5. Limitations

In this investigation, several limitations are evident. This first limitation is that in the PIRLS 2011 and 2016 datasets, there are no variables indicative of household or parental income levels. In order to develop a measure of wealth, an asset index is created. By using an asset index to determine the treatment group, there is room for error. The wealth measure is presented in quintiles, which when converted to income levels using external data, provides a range of income levels. As such, it is not possible to create an accurate representation of the treatment group.

The next limitation is that there are no variables that indicate if an individual has received a grant in the data. In this investigation, it is assumed that if individuals qualified for the grant, they received the grant. The problem with the intention to treat is that the individuals are not necessarily offered the grant and declining to receive it. There are costs involved with obtaining the grant, which may hinder individuals from receiving it. Another limitation of this investigation is that variable indicating the race of a learner is not included in the data. This is problematic when developing the models as the race of a student is likely to affect the learning outcomes of the learner, due to South Africa's apartheid history which has resulted in systematic issues that need to be accounted for in the models.

6. Conclusion

In conclusion, through conducting a difference-in-difference analysis of PIRLS 2011 and 2016 data, it is evident that the rise in eligibility of Child Support Grant (CSG) recipients due to an increase in maximum age has resulted in a positive effect on the reading scores of grade 4 learners in South Africa. Through the addition of controls to the simple difference-in-difference estimation, the direction of the positive effect of the policy change on reading scores is robust. When the data is investigated further, it is evident that although the overall reading scores of the population dropped from 2016 to 2011, the drop in the treatment group was smaller than that of the untreated group. This reinforces the results of the difference-in-difference estimation. The investigation is faced with several limitations, such as the neglect of variables that hold important information for the question at hand (for example, a grant recipient variable), however the results are still noteworthy.

7. Reference List

Beukes, R., Jansen, A., Moses, M. & Yu, D. 2017. Exploring the Eligibility Criteria of the Child Support Grant and its Impact on Poverty. *Social Indicators Research*, 134:511-529.

Bhorat, H. & Köhler, T. 2020. Social Assistance during South Africa's National Lockdown: Examining the COVID-19 grant, changes to the Child Support Grant and post-October policy options. *Development Policy Research Unit.* Working Paper No. 202009.

Gerwyn, N. 2019. *Generalised Regression Difference in Differences: with Fixed Effects, Multiple Treatment Periods and Dynamic Treatment Effects* [Online]. Available: https://medium.com/eatpredlove/regression-difference-in-differences-208c2e787fd2 [2022, 19 June].

Howie, S.J., Combrinck, C., Roux, K., Tshele, M., Mokoena, G.M., & McLeod Palane, N. (2017). *PIRLS LITERACY 2016: South African Highlights Report.* Pretoria: Centre for Evaluation and Assessment.

Sant'Anna, P. & Zhao, J. 2020. Doubly Robust Difference-in-Differences Estimators. *Journal of Econometrics*, (219):101-122.

Statistics South Africa. 2015. Earnings and Spending in South Africa, 2006–2011.

Statistics South Africa. 2017. Living Conditions of Households in South Africa.