A comment on "They Never had a Chance: Unequal Opportunities and Fair Redistributions"*

Valentin Klotzbücher Tim Krieger Marco Wallner

March 25, 2024

Abstract

We reproduce the study by Dong et al. (2025)
Keywords:

JEL codes:

*Authors: Klotzbücher: University of Freiburg and University Hospital Basel. E-mail: valentin.klotzbuecher@usb.ch; Krieger: University of Freiburg and CESifo. E-mail: tim.krieger@econ.uni-freiburg.de; Wallner: Independent researcher

1 Introduction: Example

Analyst et al. (2022), henceforth SSD, investigate the impact of a program called PROSCOL. The setting is the country of Labas. In 2000, its government introduced an antipoverty program in northwest Labas (Ravallion 2001). The program aims to provide cash transfers to poor families with school-age children. To be eligible to receive the transfer, households must have certain observable characteristics that suggest they are poor.

SSD tested the effect of the policy (PROSCOL) on school enrollment and fertility for low-income families, using a difference-in-differences approach comparing the treated region (Labas) to untreated regions before/after the implementation of the policy. The main data set comes from the Labas Social Survey from 1998 to 2002. SSD's describe their main results on p.7 as follows: "We show that the policy (PROSCOL) increased school enrollment rates for the treated group by 30 percentage points and decreased the number of children born by 0.10 per family (mean of the dependent variable is 3.4). Our point estimates are statistically significant at the 5% level."[1]

In the present report prepared for the Institute for Replication (Brodeur *et* al. 2024), we investigate whether their analytical results are computationally reproducible and further test their replicability and robustness to: (1) adding more years to the sample and (2) changing how standard errors are clustered. In their original analysis, SSD rely on data from 1998 to 2002 and cluster their standard errors at the region/year level. In our re-analysis, we extend the time period to 1998 to 2004 and cluster the standard errors at the region level. We are grateful to the original authors for providing the raw data, which was not included in the replication package.

In terms of reproducibility, we would like to acknowledge that the original study was successfully reproduced by the data editor's team at the American Economic Review. We also successfully reproduced SSD's main tables (Tables 4 and 5) using their codes, although there were very small discrepancies in the magnitude of the main point estimates for Table 5 due to coding errors. We uncovered two minor coding errors; (1) coding the control variable Age and (2) the gender dummy was

[1]Report the statistical significance used by the original authors.

included as a continuous variable in one regression.

In terms of replication, we add more years to the sample. We find that adding more years to the sample decreases the size of the main point estimate by one-third for educational attainment and by one-fourth for fertility. The point estimate for fertility becomes statistically insignificant at the 5% level, while it remains significant at the 5% level for education.

We then turn to sensitivity analysis [robustness reproduction]. We test the robustness of the results to changing how standard errors are clustered. Clustering at the region level makes the point estimates to be statistically insignificant at the 5% level.

2 Computational Reproducibility (Example)

We used the replication package here: hyperlink. The cleaning codes were provided in the replication package, but only the analysis data were provided. The Institute for Replication reached out to the authors who provided us the raw data. We successfully computationally reproduced all the main results (*i.e.*, Tables 2, 3 and 4) from the raw data. See Table 1 for details.

Table 1: Replication Package Contents and Reproducibility

Replication Package Item	**Fully**	**Partial**	**No**
Raw data provided	✓		
Analysis data provided	✓		
Cleaning code provided			✓
Analysis code provided	✓		✓
Reproducible from raw data			✓
Reproducible from analysis data	✓		

Notes: This table summarizes the replication package contents contained in Analyst et al. (2022). **Displayed here only for replicators knowledge. Please put a similar table in your Tables appendix.**

We describe in this section two minor coding errors that we uncovered while reproducing the study. First, we noticed that the coding of the control variable Age was incorrect. Age was defined as the age of the mother in the paper but coded using the variable age of the head of the household in the codes. Second, the gender dummy was included as a continuous variable in one regression. Our codes/programs are available here (*e.g.* OSF webpage with DOI).[2] The original authors' updated codes are available here (*e.g.* OSF webpage with DOI).

We re-run the codes correcting these two errors and reproduce the results for the outcome variable fertility in Table 1. (The specification for educational attainment does not include these control variables.) The structure of the table is the same as in the original study. We find that the point estimates are strikingly similar, with the sign, magnitude and statistical significance being remarkably similar.

2.1 Discrepancies Between Pre-analysis Plan and Article (Example)

Mandatory for *Psychological Science*. Optional for all other journals.

The authors did not register a pre-analysis plan.

3 Robustness Reproduction and Replication with New Data (Example)

We now turn our attention to our replication. We test the robustness of the results to a direct replication by adding more years to the sample and a robustness replicability by changing how standard errors are clustered. We add more years to the sample to increase the sample size as the original study was underpowered (e.g., 40% power). We cluster the standard errors at the region level instead of at the region/year level to account for non-independence between years within each region.

The decision to conduct these two robustness checks was taken after reading the paper but prior to observing the codes/programs.[3] We pre-registered our sensitivity analysis here (*e.g.* hyperlink to your pre-analysis plan).

3.1 Regression model

For our analysis, we rely on the same specifications and a difference-in-differences analysis comparing the treated region (Labas) to untreated regions before/after the

[2]Make sure to cite and clearly reference your data and codes.

[3]You should be honest about whether your sensitivity analysis was conceived before or after looking at the programs of the original author(s) and state whether your replication or sensitivity analysis was pre-registered.

implementation of the policy, restricting the sample to low-income families.[4] The analysis is at the region/year level for the educational attainment outcome and at the family level for the fertility outcome. See the original study for more details and equations. See an example of what to write if the specification, model or method is different in Appendix I.

3.2 Results educational attainment

3.2.1 Extending the time period We first investigate whether extending the time period to 2004 has an impact on the sign, magnitude and statistical significance of the difference-in-differences model for educational attainment. For this analysis, an observation is a region/year. The sample is restricted to low-income families and collapsed at the region/year level. There are 20 regions and seven years. The dependent variable, educational attainment, is the fraction of low-income families with all school-age children attending school in a region in a year. Robust standard errors are in parentheses, clustered by region/year. Our findings are reported in Table 2. Column 1 reports the preferred estimates from the original study.[5]

The preferred specification, as determined by the original authors, is presented in column 2. Column 3 adds control variables as in the original study. We find that the policy (PROSCOL) increased school enrollment rates for the treated group by 21 percentage points (in comparison to 30 percentage points in the original study). The point estimate remains statistically significant at the 5

3.2.2 Clustering We then investigate whether changing the clustering technique affects the main point estimates for educational attainment. Robust standard errors are in parentheses, clustered by region are reported in Table 3. Recall that the original study clustered by region/year. Column 1 reports the preferred estimates from the original study. We find that the point estimate for the preferred specification (column 2) is not anymore statistically significant at the 5% level. The standard

[4]Make sure to mention the main statistical or econometric method used to examine each claim and whether the method that you use is the same as in the original study. Also, you should state and rely on the original authors' preferred specification (or yours, if the authors are not clear).

[5]It is useful to offer a direct comparison to the original authors' estimates, for instance, in the first column of your tables. Another option is to reproduce the entire table. See the Appendix for a reproduction of the original authors' point estimates (Table 4 in their paper).

error is now much larger (0.235) than in the original study (0.151).[6]

3.3 Fertility

3.3.1 Extending the time period We turn to the second outcome variable, fertility, in this subsection. We first extend the time period to 2004. For this analysis, an observation is a family. The sample is restricted to low-income families. The dependent variable is the number of children and the mean of the dependent variable is 3.3 (3.4 in the original study). Robust standard errors are in parentheses, clustered by region/year. Our findings are reported in Table 4. Column 1 reports the preferred estimates from the original study. We find that adding more years to the sample decreases the size of the main point estimate by one-fourth; the point estimate in column 2 (preferred specification) is now -0.075, while it was -0.098 in the original study. The point estimate becomes statistically insignificant at the 5% level (standard error 0.058).

3.3.2 Clustering We now turn to changing the clustering technique for fertility. Robust standard errors are in parentheses, clustered by region are reported in Table 5. We find that the point estimate for the preferred specification (column 2) is not anymore statistically significant at the 5% level. The standard error is now much larger (0.075) than in the original study (0.055).

3.4 Extending time period and clustering

We now explore the effect of changing the time period and clustering simultaneously on the two main dependent variables. The estimates are presented in Tables 6 and 7. Overall, we confirm our previous conclusions.

4 Conclusion

[6]Be careful to not label differences between your estimates and the original author(s) estimates as mistakes or errors. Rather, help the reader better understand why you conducted a specific robustness check or modified the setting/model.

References

Analyst, S., Analyst, S. and Doe, J.: 2022, The impact of PROSCOL on educational attainment and fertility, *The Journal* **1**(1), 1–10.

Brodeur *et* al.: 2024, Mass Reproducibility and Replicability: A New Hope, *I4R Discussion Paper Series* .

Dong, L., Huang, L. and Lien, J. W.: 2025, 'they never had a chance': Unequal opportunities and fair redistributions, *The Economic Journal* **135**(667), 914–942.

Ravallion, M.: 2001, The mystery of the vanishing benefits: An introduction to impact evaluation, *The World Bank Economic Review* **15**(1), 115–140.

5 Figures

Insert figures here.

6 Tables

Table 2: Coding error: Fertility

OLS	Number of Children Original Study (1)	Number of Children (2)
Treated (Labas)	0.279***	0.278***
	(0.079)	(0.079)
	[0.001]	[0.001]
After Policy	0.047	0.047
	(0.045)	(0.045)
	[0.317]	[0.317]
Treated * After	-0.100**	-0.101**
	(0.042)	(0.042)
	[0.046]	[0.045]
Controls		
Demographic Controls	Y	Y
Observations	12,311	12,311
R-Squared	0.100	0.100

Notes: Authors' calculations using data from 1998–2002. Mention data sources here. An observation is a family. The sample is restricted to low-income families. There are 20 regions and five years. The dependent variable is the number of children. Robust standard errors are in parentheses, clustered by region/year. Significant at the ***[1%] **[5%] *[10%] level.

Table 3: Impact of the treatment on educational attainment: Changing time period

OLS	Educational Attainment Original Study	Educational Attainment	Educational Attainment
	(1)	(2)	(3)
Treated (Labas)	-0.147***	-0.168***	-0.208***
	(0.020)	(0.024)	(0.041)
	[0.001]	[0.001]	[0.001]
After Policy	0.041	0.043	0.050
	(0.030)	(0.029)	(0.047)
	[0.240]	[0.222]	[0.311]
Treated * After	0.333**	0.210**	0.200**
	(0.151)	(0.095)	(0.097)
	[0.047]	[0.043]	[0.049]
Controls			
Demographic Controls	N	N	Y
Observations	100	140	140
R-Squared	0.047	0.048	0.051

Notes: Authors' calculations using data from 1998–2004. Mention data sources here. An observation is a region/year. The sample is restricted to low-income families and collapsed at the region/year level. There are 20 regions and seven years. The dependent variable, educational attainment, is the fraction of low-income families with all school-age children attending school in a region in a year. Robust standard errors are in parentheses, clustered by region/year. Significant at the ***[1%] **[5%] *[10%] level.

Table 4: Impact of the treatment on educational attainment: Changing clustering

OLS	Educational Attainment Original Study	Educational Attainment	Educational Attainment
	(1)	(2)	(3)
Treated (Labas)	-0.147***	-0.147*	-0.188*
	(0.020)	(0.085)	(0.103)
	[0.001]	[0.078]	[0.058]
After Policy	0.041	0.041	0.047
	(0.030)	(0.045)	(0.052)
	[0.261]	[0.342]	[0.331]
Treated * After	0.333**	0.333	0.300
	(0.151)	(0.235)	(0.216)
	[0.047]	[0.171]	[0.183]
Controls			
Demographic Controls	N	N	Y
Observations	100	100	100
R-Squared	0.047	0.047	0.050

Notes: Authors' calculations using data from 1998–2002. Mention data sources here. An observation is a region/year. The sample is restricted to low-income families and collapsed at the region/year level. There are 20 regions and five years. The dependent variable, educational attainment, is the fraction of low-income families with all school-age children attending school in a region in a year. Robust standard errors are in parentheses, clustered by region. Significant at the ***[1%] **[5%] *[10%] level.

Table 5: Impact of the treatment on fertility: Changing time period

OLS	Number of Children Original Study	Number of Children	Number of Children
	(1)	(2)	(3)
Treated (Labas)	0.375**	0.275**	0.249**
	(0.157)	(0.127)	(0.119)
	[0.042]	[0.032]	[0.033]
After Policy	0.041	0.031	0.037
	(0.030)	(0.025)	(0.035)
	[0.252]	[0.272]	[0.291]
Treated * After	-0.098*	-0.075	-0.075
	(0.055)	(0.058)	(0.057)
	[0.062]	[0.248]	[0.245]
Controls			
Demographic Controls	N	N	Y
Observations	15,345	15,345	15,300
R-Squared	0.098	0.098	0.100

Notes: Authors' calculations using data from 1998-2004. Mention data sources here. An observation is a family. The sample is restricted to low-income families. There are 20 regions and seven years. The dependent variable is the number of children. Robust standard errors are in parentheses, clustered by region/year. Significant at the ***[1%] **[5%] *[10%] level.

Table 6: Impact of the treatment on fertility: Changing clustering

OLS	Number of Children Original Study	Number of Children	Number of Children
	(1)	(2)	(3)
Treated (Labas)	0.375**	0.375	0.279
	(0.157)	(0.357)	(0.289)
	[0.042]	[0.249]	[0.353]
After Policy	0.041	0.041	0.047
	(0.030)	(0.037)	(0.045)
	[0.252]	[0.257]	[0.311]
Treated * After	-0.098*	-0.098	-0.100
	(0.055)	(0.075)	(0.073)
	[0.062]	[0.266]	[0.241]
Controls			
Demographic Controls	N	N	Y
Observations	12,345	12,345	12,311
R-Squared	0.098	0.098	0.100

Notes: Authors' calculations using data from 1998-2002. Mention data sources here. An observation is a family. The sample is restricted to low-income families. There are 20 regions and five years. The dependent variable is the number of children. Robust standard errors are in parentheses, clustered by region. Significant at the ***[1%] **[5%] *[10%] level.

Table 7: Impact of the treatment on educational attainment: Changing time period and clustering

OLS	Educational Attainment Original Study	Educational Attainment	Educational Attainment
	(1)	(2)	(3)
Treated (Labas)	-0.147***	-0.168*	-0.208*
	(0.020)	(0.085)	(0.103)
	[0.001]	[0.057]	[0.056]
After Policy	0.041	0.043	0.050
	(0.030)	(0.045)	(0.052)
	[0.240]	[0.329]	[0.315]
Treated * After	0.333**	0.210**	0.200**
	(0.151)	(0.235)	(0.216)
	[0.047]	[0.360]	[0.368]
Controls			
Demographic Controls	N	N	Y
Observations	100	140	140
R-Squared	0.047	0.048	0.051

Notes: Authors' calculations using data from 1998–2004. Mention data sources here. An observation is a region/year. The sample is restricted to low-income families and collapsed at the region/year level. There are 20 regions and seven years. The dependent variable, educational attainment, is the fraction of low-income families with all school-age children attending school in a region in a year. Robust standard errors are in parentheses, clustered by region. Significant at the ***[1%] **[5%] *[10%] level.

Table 8: Impact of the treatment on fertility: Changing time period and clustering

OLS	Number of Children Original Study (1)	Number of Children (2)	Number of Children (3)
Treated (Labas)	0.375**	0.275	0.249
	(0.157)	(0.357)	(0.289)
	[0.042]	[0.412]	[0.429]
After Policy	0.041	0.031	0.037
	(0.030)	(0.037)	(0.045)
	[0.252]	[0.422]	[0.521]
Treated * After	-0.098*	-0.075	-0.075
	(0.055)	(0.075)	(0.073)
	[0.062]	[0.312]	[0.309]
Controls			
Demographic Controls	N	N	Y
Observations	15,345	15,345	15,300
R-Squared	0.098	0.098	0.100

Notes: Authors' calculations using data from 1998-2004. Mention data sources here. An observation is a family. The sample is restricted to low-income families. There are 20 regions and seven years. The dependent variable is the number of children. Robust standard errors are in parentheses, clustered by region/year. Significant at the ***[1%] **[5%] *[10%] level.

7 APPENDIX

Example of what to write if the replicator(s)' specification, model or method differs from the original study for the sensitivity analysis.

We test the claim that PROSCOL impacted enrollment rates and the number of children born for treated regions in comparison to control regions using a triple differences model comparing low-income and high-income families in treated and control regions before/after the implementation of the policy. This setting contrasts with the original study which relies on a difference-in-differences comparing treated regions to untreated before/after the implementation of the policy for low-income families. Our equation is:

$$\begin{aligned} Y_{irt} = \alpha + \beta_1 Treated_{rt} + \beta_2 LowIncome_{irt} + \beta_3 Aftert + \beta_4 Treated_t \times LowIncome_{irt} \\ + \beta_5 Treated_{rt} \times After_t + \beta_6 LowIncome_{irt} \times After_t \\ + \beta_7 Treated_{rt} \times LowIncome_{irt} \times After_t + \zeta X_{irt} + \epsilon_{irt} \quad (1) \end{aligned}$$

where the dependent variable is the number of children born for individual i in region r in time t. The time period is 1998–2004 (1998–2002 in the original study) and an observation is a family (head of the family for the control variable). Treated region is an indicator which is set to 1 if the respondent lives in the treated region (northwest region of Labas), Low Income indicates whether the family is low-income, and After is a variable that takes the value 1 if the year is after 1999. The interaction of Low Income and Treated Region determines the treated group, while the additional interaction with After shows the effect of the treatment. The coefficient of interest is β_7. X_{irt} represents a vector of family socioeconomic variables (head of the family). We use robust standard errors and cluster the error term at the region level (region/year level in the original study).

8 Appendix Figures

Insert figures here.

9 Appendix Tables

Here, you may reproduce the main tables (Tables 4 and 5) from the original study.