**READ ME: Taking an Extra Moment to Consider Treatment**

**Effects on Distributions; Empirical Application II:**

Gawain Heckley & Dennis Petrie

This document provides a description of the data, do-files, and variables used for the analyses of empirical application II: Swedish comprehensive school reform, in “Taking an Extra Moment to Consider Treatment Effects on Distributions” by Gawain Heckley & Dennis Petrie.

# Data Availability

Swedish administrative data cannot be shared for ethical/privacy reasons. Our data contains sensitive information on individual outcomes and, hence, Swedish law requires users of the data to hold a permission from the Swedish Ethical Review Authority (“Etikprövningsmyndigheten”). This means that we are unable to make our data available online. Researchers interested in obtaining the data can apply for a permission from the Swedish Ethical Review Authority at <https://etikprovningsmyndigheten.se/>. Conditional on approval, researchers can apply for and buy the data from Statistics Sweden ([www.scb.se](http://www.scb.se)). Statistics Sweden removes personal identifiers and replaces these with a personal ID used to merge the data sets. The data on the introduction of the comprehensive school reform is provided as part of the replication package for Hjalmarsson et al (2015). The data obtained from Statistics Sweden does not contain any personal information such as names and addresses.

Formal references to all the data sets used are provided at the end of the document. Below, we describe the various data sets used in the analyses.

# Individual-level data

|  |  |
| --- | --- |
| *Data set* | *Reference* |
| Population register (RTB) (1968-2016) | Statistics Sweden (2020a) |
| Census (FoB) (1960, 1965, 1970) | Statistics Sweden (2020b) |
| Income register (1968-2016) | Statistics Sweden (2020c) |
| Education register (1990-2016) | Statistics Sweden (2020d) |
| Multigeneration register (1932-2016) | Statistics Sweden (2020e) |

# Historical Reform Data

The data on the introduction of the comprehensive school reform in Sweden is provided as part of the replication package for Hjalmarsson et al (2015).

# Do-files

STATA version 18 was used for the analyses. The STATA do-files are:

*did\_multiplegt\_PERM.ado* – the STATA ado file that estimates DiD using the method of De Chaisemartin & d’Haultfoeuille (2020) and also provides the scalars required to estimate PERM DiD, storing them in STATA’s temporary memory to be used by subsequent commands.

The do-files listed below produce the analysis dataset from the raw register data as well as the tables and figures in the paper. The individual tables and figures are found in the subfolders “Tables” and “Figures”.

Data manipulation:

1. 0\_NewSIPGenSamp.do
2. 1\_NewSIPGenEarn.do
3. 2\_NewSIPGenEduc.do
4. 3\_NewSIPGen\_Parentbackground.do
5. Data\_merge\_PERM.do

Figures and Tables:

1. *Ana\_PERM\_Reform\_RollOut.do* – produces Figure L.1
2. *Ana\_PERM\_descriptives.do* - produces Table M.1
3. *Ana\_PERM\_earningsHisto\_LT9years.do* - produces Figure M.1
4. *Ana\_PERM\_bstrap\_edu.do* - Estimates PERM DiD, bootstraps the whole procedure and provides years of education and covariance of earnings and years of education event study Figures 3, 4 N.1, N.2, N.3
5. *Ana\_PERM\_bstrap\_inc.do* - Estimates PERM DiD, bootstraps the whole procedure and provides earnings event study Figures 3, 4 N.1, N.2, N.3
6. *Ana\_PERM\_bstrap\_MainTable.do* - produces Table 4 using results from *Ana\_PERM\_bstrap\_edu.do* and *Ana\_PERM\_bstrap\_inc.do*
7. *Ana\_PERM\_balancing\_occ.do* - produces Table O.1
8. *Ana\_PERM\_bstrap\_inc.outliers.do* - produces Figures O.1
9. *Ana\_PERM\_fig\_covariance.do* - produces Figure P.2

# Variables in do-files

The variables used in the do-files are described in Table 1 below.

**Table 1. Variable description**

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| Firstcohort60 | First birth cohort exposed to reform by municipality (Hjalmarsson et al. 2015) |
| corigin | ==29 if born born in Sweden |
| SUN2000Niva, SUN2000Niva\_old | Highest level of education achieved (education register) |
| SkolUtbAlm | Highest level of schooling (1970 census) |
| DAGPE |  |
| PENS | Pension income (from 1974) |
| CSFVI | Total pre-tax income |
| ARBINK/ARBINSJ/CARB | Income from employment minus, sick leave and working age pension) |
| DodDatum | Death date |
| FodelseArMan | Birth year and month |
| Kon | Gender |
| FelPersonNr | Wrong personal identifier |
| AterAnv | Re-used personal identifier |
| fodelseland | Country of Birth |
| FAR\_LopNr\_PersonNr | Personal identifier Father |
| MOR\_LopNr\_PersonNr | Personal identifier Mother |
| Forsamling | Parish of residence (1960, 1965) |
| Kommun | Municipality of residence (1960, 1965) |
| Yrke | Occupation (1960 census) |

# Stata packages used

did\_multiplegt is used to estimate Figure 4

Installation:

ssc install did\_multiplegt

# References

De Chaisemartin, C. and d’Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. American economic review, 110(9), pp.2964-2996.

Hjalmarsson, R., Holmlund, H. and Lindquist, M.J., 2015. The effect of education on criminal convictions and incarceration: Causal evidence from micro‐data. *The Economic Journal*, *125*(587), pp.1290-1326.

Statistics Sweden (SCB). (2020a). “Registret över totalbefolkningen, RTB (Population register), 1968-2012 [database]”, SCB:s Statistikservice, accessed (2024).

Statistics Sweden (SCB). (2020b). “Folk- och bostadsräkning, FOB (Census), 1960, 1965 [database]”, SCB:s Statistikservice, accessed (2024).

Statistics Sweden (SCB). (2020c). “Inkomst- och taxeringsregistret, IoT, (Income register), 1968-2011 [database]”, SCB:s Statistikservice, accessed (2024).

Statistics Sweden (SCB). (2020d). “Utbildningsregistret (Education register), 1990-2011 [database]”, SCB:s Statistikservice, accessed (2024).

Statistics Sweden (SCB). 2020e. “Flergenerationsregistret (Multigeneration register), 19322011 [database]”, SCB:s Statistikservice, accessed (2024).