This document provides a description of the data and do-files used for the article "Infant Health, Cognitive Performance and Earnings: Evidence from Inception of the Welfare State in Sweden", by Sonia Bhalotra, Martin Karlsson, Therese Nilsson and Nina Schwarz.

Data availability

To undertake our analysis we purposively digitised individual birth certificates from historical parish records to obtain a census of births in every treated area and a set of matched controls. To this data, we match school records that we collected and digitized and administrative information from census and tax records to track individuals through the life-course. We also digitised aggregate information on election results.

The data are available for any researcher, but given security rules for Swedish administrative data, they are only available in a secure remote access system and subject to an application process requiring ethical approval from the Swedish Ethical Review Authority https://etikprovningsmyndigheten.se. Once ethical approval has been granted, we are happy to make the data available to anyone who would like to replicate the results. As mentioned before, these data have to be accessed through a secured server. To get server access please contact Therese Nilsson at therese.nilsson@nek.lu.se.

The administrative data builds on several data sources that can be ordered from:

- Ratsit (https://www.ratsit.se/): Provides tax records for the years 2001-2005
- Statistics Sweden (SCB, scb.se): Provides Census 1970
- the Swedish National Archive (Riksarkivet, https://riksarkivet.se/start): Provides birth records for cohorts 1930-1934
- Arkiv Digital (https://www.arkivdigital.se/): Provides Census 1950
- Federation of Swedish Genealogical Societies (Sveriges släktforskarförbund, https://www.rotter.se/): Provides death register

The aggregate data comes from:

- Kommunala valen år 1946 (http://share.scb.se/ov9993/data/historisk%20statistik//SOS%201911-/Valstatistiken/Kommunala%20valen%20(SOS)%201919-1966/Valstatistik-Kommunala-valen-1946.pdf): Provides election results and local female political representation

Data

To replicate the results in the paper, the following datasets are required:

Datafinal.dta: This data set includes individual level information for the sample cohorts 1930-34 from birth registers, administrative school records, census 1970 and tax records for the years 2001-2005. Furthermore, it includes information on school reforms, treatment status and eligibility.

ParishkeyFinal: This data set includes aggregate parish level information for treatment and control parishes.

Indsample.dta: This data set includes the Age 7 Index, the Age 10 Index, and the Adult Index.

Childc.dta: This data set includes information on the growth in the number of childcare workers in each municipality.

Elections.dta: This data set includes aggregate information on the number of elected women and seats in each municipality level council for the 1946 election.

Utilisation.dta: This data set includes the individual level information on detailed infant and maternal care utilisation.

Autor.dta: This data set is based on the classification by Autor et al. (2003). Each 1970 occupation code is assigned to a certain skill content. See Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. The Quarterly journal of economics 118(4), 1279–1333.

Widow_mID.dta: Individual level information on widow status between 2001-2005 derived from tax records.

Attadj.dta: Probability weights and censoring probabilities for the calculation of Lee bounds for the school data sample, the census 1970 sample and the pension sample.

Placebo4044.dta: This data set includes individual level information on cohorts born 1940-44. The information stems from Census 1950 and is used for the placebo analysis.

Parishpanelfinal.dta: This data set contains municipality-level data from yearbooks covering the period 1918-1938.

Förvärvsarbetande1930-1975.dta: Information on females working in public sector jobs taken from Statistiska Centralbyrån (2009). Förvärvsarbetande: Folk- och bostadsräkningarna 1910-1985. National Bureau of Statistics.

PublicOccupations1960-1975.dta: Information on working population by gender taken from Statistiska Centralbyrån (2009). Förvärvsarbetande: Folk- och bostadsräkningarna 1910-1985. National Bureau of Statistics.

birthparshare.dta, parish46share.dta: The files include, for each parish-cohort, the number of boys and girls in each secondary school catchment (given as the nearest location with a secondary school). One file is based on birth parish and the other on parish of residence 1946.

Cshares.dta Employment shares for 4-digit industry codes by subpopulation according to the 1950 and 1970 censuses.

CshiftsharesAgeC.dta: Shift-share instrumental variables by parish, age, sex and qualification.

rotemberg2sam.dta: Rotemberg weights by industry, sex and qualification.

Do files (Stata)

The files titled Figure*.do and Table*.do are do-files for Stata. These show the code for the construction of all tables and figures in the paper and the online Appendix based on our estimation dataset. The number refers to the corresponding figure/table in the paper.

Furthermore, the following do-file is included:

MASTER

- Master do-file which calls the Table and Figure do-files.

Software

Stata 15

Operating System

Windows Server 2012 R2