**How Do Inheritances Shape Wealth Inequality?**

**Theory and Evidence from Sweden**

README

**Data availability statement**

*The data for this project are confidential. They may be obtained from Statistics Sweden (SS) upon a confidentiality assessment congruent with the Public Access to Information and Secrecy Act. Researchers interested in access to the data may contact mikrodata@scb.se , also see*

[*https://www.scb.se/en/services/guidance-for-researchers-and-universities/*](https://www.scb.se/en/services/guidance-for-researchers-and-universities/)

*It takes between around 6 months from the data request to the delivery of microdata. Most orders cost between SEK 50,000 and SEK 100,000.*

*All replication scripts, including detailed explanations of data construction, etc. are available at the following DOI: 10.5281/zenodo.5849620. https://doi.org/10.5281/zenodo.5849620*

**Registers used**

*Statistics Sweden. 2020. “Longitudinal integrated database for health insurance (LISA), 1990-2012 [database]”, accessed 2014.*

*Referred to below as: LISA.*

*Statistics Sweden. 2020. “Wealth (Formogenhetsregistret)), 1999-2007 [database]”, Statistics Sweden Research Service, accessed 2014.*

*Referred to below as: Wealth.*

*Statistics Sweden. 2020. “KURU, 1999-2007 [database]”, accessed 2014.*

*Referred to below as: KURU.*

*Statistics Sweden. 2020. “IoT, 1991-2012 [database]”, accessed 2014.*

*Referred to below as: IoT.*

*Statistics Sweden. 2020. “Vehicle Register (Fordonregistret), 1999-2007 [database]”, accessed 2014.*

*Referred to below as: Vehicle.*

*Statistics Sweden. 2020. “RAMS (Employer-Employee Register), 1985-2012 [database]”, accessed 2014.*

*Referred to below as: RAMS.*

*Statistics Sweden. 2020. “Belinda (Inheritance Register), 2001-2004 [database]”, accessed 2014.*

*Referred to below as: BELINDA.*

*Statistics Sweden. 2020. “Multigenerational Register [database]”, accessed 2014.*

*Referred to below as: MULTI.*

*National Board of Health and Welfare. 2020 “Cause-of-death Register [database]”, accessed 2014.*

*Referred to below as: COD.*

**Stata dofiles / packages used**

weighting\_demean.do – This dofiles reweights observations according to DiNardo, Fortin, Lemieux (1996) in the context of the demeaning-method.

weighting\_deltademean.do – This dofiles reweights observations according to DiNardo, Fortin, Lemieux (1996) in the context of the delta-demeaning-method.

graphexportpdf

* Type ssc install graphexportpdf to install this package.

**List of analysis datasets**

**Eventdesign.dta**

Population definition:

* All children losing a parent in 2000, 2003, 2006, 2009, 2012 and 2013 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* + Annual panel data during 1999-2007. Contains wealth (at market values).

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* Reweight observations according to the attached do-file weighting\_demean.do. This requires as input a reference year-of-death cohort, which is the first parent-death cohort, 2000. The dofiles reweights each of the other cohorts 2003, 2006, 2009, 2012 and 2013 to match the 2000-cohort.
* Then take the weighted median of wealth at market values by year of death of the parent and by calendar year. This transforms the data from 8,114,362 observations to 54 (9 calendar years times 6 cohorts).

**Diffdeltadata.dta**

Population definition:

* All children losing a parent in 2000 and in 2002-2010 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* + Annual panel data during 1999-2007. Contains wealth (at market values).

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* Reweight observations according to the attached do-file weighting\_demean.do. This requires as input a reference year-of-death cohort, which is the first parent-death cohort, 2000. The dofiles reweights the cohorts 2002-2010 to match the 2000-cohort.
* Then take the weighted median of wealth at market values by year of death of the parent and by calendar year. This step converts 13,349,208 observations to 81.
* Finally, compute the difference of median annual wealth between the 2000-cohort and each of the other ones.

**Demean\_data.dta**

Population definition:

* All children losing a parent in 2000-2004 and in 2008-2012 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* Annual panel data during 1999-2007. Contains the following key variables:
  + wealth (at market values, at constant prices); wealth transformations (Arcsinh; proportional; rank within cohort).
  + Car purchases; car values.
  + Unearned income; consumption and savings.

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* Reweight observations according to the attached do-file weighting\_demean.do. This requires as input a reference year-of-death cohort, which is the first parent-death cohort, 2000. The dofiles reweights each of the other treatment-cohorts (2001-2004) and the control group (2008-2012) collectively, to match the 2000-cohort.
* Then take the weighted mean of each outcome of interest among the control group by calendar year. Merge these means to the treatment panels by year. For each variable except proportional wealth, subtract the weighted mean from each treatment unit’s observations. For proportional wealth, instead divide the outcome with the mean.

**Delta\_demean\_data.dta**

Population definition:

* All children losing a parent in 2000-2004 and in 2008-2015 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* Annual panel data during 1991-2012. Contains the following key variables:
  + wealth (at constant quantities; holding a baseline portfolio).
  + Labor earnings and transformations such as labor earnings rank (within cohort), log-earnings.

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* In the associated estimation strategy we attach a particular set of control cohorts to each treatment cohort. For children who lose their parents in 2000, the control group is 2008-2011. For the 2001-cohort, it comprises 2009-2012 deaths. Therefore duplicate the control observations and define a variable gtreat that associated each unit to a particular treatment year 2000-2004.
* Then, reweight observations according to the attached do-file weighting\_deltademean.do. This requires as input a reference year-of-death cohort, which is – as for the demeaning-method -- the first parent-death cohort, 2000. The dofiles reweights each of the other treatment-cohorts (2001-2004) and the control groups (2008-2015) collectively, to match the 2000-cohort.
* Then take the weighted mean of each outcome of interest among the control groups by calendar year. Merge these means to the treatment panels by both year of death and calendar year. For each variable, subtract the weighted mean from each treatment unit’s observations.

**Demean\_data\_hetero.dta**

Population definition:

* All children losing a parent in 2000-2004 and in 2008-2012 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* Annual panel data during 1999-2007. Contains the following key variables:
  + wealth (at market values, at constant prices); wealth transformations (Arcsinh; proportional; rank within cohort).
  + Car purchases; car values.
  + Unearned income; consumption and savings.

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* Define groups of own wealth and parents’ wealth in 1999 by their corresponding ranking in that year. Create four groups as in the notes to Figure 5 describes.
* Reweight observations according to the attached do-file weighting\_demean.do, separately for each wealth group. This requires as input a reference year-of-death cohort, which is the first parent-death cohort, 2000. The dofiles reweights each of the

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **List of different datasets** | **Nr of obs** | **Unique identifier** | **Inputs** | **Years** | **Reweighting** | **Input parent death years** |
| Eventdesign.dta | *54* | Parent-death-year, calendar year | COD, LISA, Wealth | 1999-2007 | Reweight to match cohort 2000. | 2000, 2003, 2006, 2009, 2012, 2013 |
| Diffdeltadata.dta | *81* | Parent-death-year, calendar year | COD, LISA, Wealth | 1999-2007 | Reweight to match cohort 2000. | 2000, 2002-2010 |
| Demean\_data.dta | *13,404,527* | Individual-id, parent-id, year | COD, LISA, Wealth, KURU, MULTI, Vehicle | 1999-2007 | Reweight to match cohort 2000. | 2000-2004  2008-2012 |
| Delta\_demean\_data.dta | *42,972,178* | Individual-id, parent-id, year | COD, LISA, Wealth, KURU, MULTI, IoT, RAMS | 1991-2012 | Reweight to match cohort 2000. | 2000-2004  2008-2015 |
| Demean\_data\_hetero.dta | *13,404,527* | Individual-id, parent-id, year | COD, LISA, Wealth, KURU, MULTI, Vehicle | 1999-2007 | Reweight to match cohort 2000 within strata. | 2000-2004  2008-2012 |
| Delta\_demean\_data\_hetero.dta | *42,972,178* | Individual-id, parent-id, year | COD, LISA, Wealth, KURU, MULTI, IoT, RAMS | 1991-2012 | Reweight to match cohort 2000 within strata. | 2000-2004  2008-2015 |
| shortrun.dta | *433,652* | Individual-id | COD, Wealth, BELINDA | Cross-section | N/A | 2001-2004 |
| longrun.dta | *13,013* | Individual-id, parent-id, year | Demean\_data.dta | 1999-2007 | Reweight to match cohort 2000 | 2000-2004  2008-2012 |

* other treatment-cohorts (2001-2004) and the control group (2008-2012) collectively, to match the 2000-cohort.
* Then take the weighted mean of each outcome of interest among the control group by calendar year. Merge these means to the treatment panels by year. For each variable except proportional wealth, subtract the weighted mean from each treatment unit’s observations. For proportional wealth, instead divide the outcome with the mean.

**Delta\_demean\_data\_hetero.dta**

Population definition:

* All children losing a parent in 2000-2004 and in 2008-2015 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Panel structure:

* Annual panel data during 1991-2012. Contains the following key variables:
  + wealth (at constant quantities; holding a baseline portfolio).
  + Labor earnings and transformations such as labor earnings rank (within cohort), log-earnings.

Construction:

* To construct this dataset, start with the population definition. Merge in the variables of interest (e.g. wealth) and demographic information from LISA.
* Define reweighting cells as follows. Define a new year of birth variable that top-codes all year-of-birth years occurring after 1980 to 1980. Use the education codes (SUN-2000-codes) to define a mutually exclusive education variable taking on four values as follows:
  + 1 if 100<=SUN2000<=317 or SUN2000 = 999
  + 2 if 320<=SUN2000<=337
  + 3 if 410<=SUN2000<=527
  + 4 if 530<=SUN2000<=640.
* Define cells as combinations of the new year-of-birth variable and the coarsened education variable.
* Define groups of own wealth and parents’ wealth in 1999 by their corresponding ranking in that year. Create four groups as in the notes to Figure 5 describes.
* In the associated estimation strategy we attach a particular set of control cohorts to each treatment cohort. For children who lose their parents in 2000, the control group is 2008-2011. For the 2001-cohort, it comprises 2009-2012 deaths. Therefore duplicate the control observations and define a variable gtreat that associated each unit to a particular treatment year 2000-2004.
* Then, reweight observations according to the attached do-file weighting\_deltademean.do, separately by each of the four groups. This requires as input a reference year-of-death cohort, which is – as for the demeaning-method -- the first parent-death cohort, 2000. The dofiles reweights each of the other treatment-cohorts (2001-2004) and the control groups (2008-2015) collectively, to match the 2000-cohort.
* Then take the weighted mean of each outcome of interest among the control groups by calendar year. Merge these means to the treatment panels by both year of death and calendar year. For each variable, subtract the weighted mean from each treatment unit’s observations.

**Shortrun.dta**

Population definition:

* All children losing a parent in 2001-2004 (retrieved from *COD*, *MULTI* and *BELINDA*) who are alive at the end of the sample period (2012).

Data structure:

* This dataset contains the wealth of the sample population in the year before and after the death of the parent as well as the inheritances received (retrieved from BELINDA). We also augment the data with taxable gifts received from the same datasource.

Construction:

* This dataset contains many different rank-variables: wealth prior to inheritance receipt; inheritance amount; wealth after inheritance receipt and counterfactual rankings. After constructing the ranks, the cumulative shares of these different objects need to be calculated as they are the inputs to Figure 7.
* The counterfactual rankings (and shares) are constructed as follows: For extreme inheritance inequality, sum total inheritances and assign them to the top-1% inheritance recipients. Then compute the rankings and shares of wealth after inheritance receipt (in this counterfactual scenario), difference it against pre-inheritance wealth. For no intergenerational mobility, give each individual the average inheritance in the inheritance percentile that her wealth ranking corresponds to. For full intergenerational mobility, give everyone the average inheritance.

**Longrun.dta**

Population definition:

* All children losing a parent in 2000-2004 and 2008-2012 (retrieved from *COD* and *MULTI*) who are alive at the end of the sample period (2012).

Data structure:

* To construct this dataset, start from demean\_data.
* Adjust the top of the wealth distribution according to the Pareto tail of the wealth distribution (see Appendix Figure C.28).
* Collapse the share accruing to the top 1%, 5% and 10% for treatment and control groups separately, using the weights.
* Subtract the shares of the treatment group with that of the control group.
* Repeat this procedure 1000 times, drawing bootstrapped samples of individuals (i.e. keeping individuals over the sample period).

|  |  |  |
| --- | --- | --- |
| **Do-files** | **Input** | **Output** |
| Figure2A.do | eventdesigndata.dta | Figure2A.eps |
| Figure2B.do | Diffdeltadata.dta | Figure2B.eps |
| Figure2C.do | eventdesigndata.dta | Figure2C.eps |
| Figure3A.do | Demean\_data.dta  Delta\_demean\_data.dta | Figure3A.eps |
| Figure3B.do | Demean\_data.dta | Figure3B.eps |
| Figure4A.do | Demean\_data.dta  Delta\_demean\_data.dta | Figure4A.eps |
| Figure4BC.do | Delta\_demean\_data.dta | Figure4B.eps  Figure4C.eps |
| Figure5.do | Demean\_data\_hetero.dta | Figure5A.eps  Figure5B.eps  Figure5C.eps  Figure5D.eps |
| Figure6.do | Demean\_data\_hetero.dta  Delta\_demean\_data\_hetero.dta | Figure6.eps |
| Figure7.do | Shortrun.dta | Figure7.eps |
| Figure8AB.do | Demean\_data.dta | Figure8A.eps  Figure8B.eps |
| Figure8CD.do | Longrun.dta | Figure8C.eps  Figure8D.eps |
| Descriptives.do | Demean\_data.dta | Descriptives.tex |
| Regtable.do | Demean\_data.dta  Delta\_demean\_data.dta | Regtable.tex |