Readme

This document describes the content of the replication dataset for the article "The Geography of Intergenerational Mobility - Danish Evidence", by Eriksen, J. and Munk, M.D. (2020), Economics Letters, Volume 189, 109024.

The dataset available in the repository is the final, compiled dataset consisting of intergenerational mobility estimates for Danish Municipalities used in the article, as well as additional mobility estimates referred to in the paper. The original administrative data was accessed through a researcher agreement with Statistics Denmark, and cannot be made public due to confidentiality. To learn more about the use of Danish administrative data for research purposes in collaboration with Statistics Denmark, please see the following website: https://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang.

Description of Dataset

The dataset is in a long-format. A set of identifying variables describe the sample from which the estimate is construed, and two variables (**var** and **val**) describe the mobility estimate variable and its value contained in the row, an

Identifying variables

- **muni_name**: The name of the municipality
- **muni_post2007reform**: An indicator for whether the municipality is a post-2007-administrative reform municipality. In 2007, a large administrative reform merged Danish administrative municipalities, reducing the total number of municipalities from 271 to 98. We include estimates for municipalities based on the old and new municipal boundaries.
- **age_child**: The child-age at which the last information about parental income is estimated.
- **yvar**: The child income measure. The left-hand-side from regression estimates of intergenerational mobility, and the rank-measure for quintile based mobility estimates. The variable includes children's income rank (*c_rank*), children's family income rank (*fam_rank*), the log of children's income (*c_inc_log*) where observations with less than 0 DKK income is reset to 1000 DKK, and the log of the child's family income (*fam_inc_log*) similarly adjusted.
- **xvar**: The parental income measure. The right-hand-side from regression estimates of intergenerational mobility, and the rank-measure for quintile based mobility estimates.

The variable includes parents' summed income rank (*parent_rank*), the log of parents summed income (*parent_inc_log*), mother's income rank (*mom_rank*), mother's log income (*mom_inc_log*), father's income rank (*dad_rank*), father's log income (*dad_inc_log*).

- **N**: The total number of observations for the combination of **muni_name**, **age_child**, **yvar**, and **xvar**.

Mobility estimate variables

- **var**: A variable indicating what is measured in the **val** column. The variables include:
  - *alpha*: The intercept from regression of yvar on xvar.
  - *alpha_se*: The standard error of the alpha estimate.
  - *beta*: The slope coefficient from regression of yvar on xvar.
  - *beta_se*: The standard error for the slope coefficient from regression of yvar on xvar.
  - *r_25*: The expected income of a child born to parents at the 25th income percentile, calculated as *alpha* + *beta* * 0.25
  - *r_50*: The expected income of a child born to parents at the 50th income percentile, calculated as *alpha* + *beta* * 0.5
  - *r_75*: The expected income of a child born to parents at the 25th income percentile, calculated as *alpha* + *beta* * 0.75
  - *x_q1_N*: The number of parental observations in the first quintile (only available when xvar and yvar are rank-variables).
  - *x_q5_N*: The number of parental observations in the first quintile (only available when xvar and yvar are rank-variables).
  - *q1q1*: The share of children from municipality growing up in quintile 1 and ending up in quintile 1 (only available when xvar and yvar are rank-variables).
  - *q1q5*: The share of children from municipality growing up in quintile 1 and ending up in quintile 1 (only available when xvar and yvar are rank-variables).
  - *q5q1*: The share of children from municipality growing up in quintile 1 and ending up in quintile 1 (only available when xvar and yvar are rank-variables).
  - *q5q5*: The share of children from municipality growing up in quintile 1 and ending up in quintile 1 (only available when xvar and yvar are rank-variables).
- **val**: The value of the variable described in the column **var**

Sample and income definition

Sample

The sample used to constructing the dataset is defined as follows: Children and parents that can be linked in the Danish administrative datasets, of which the children were born between 1973 and 1977. Parents are observed at least once from 1980, the first year of available income statistics, to the year the child turns 15, 18, or 20.

The dataset has been cleaned to remove all observations for which we observe less than 25 child-parent(s) matches for the municipality. We also remove information about quintile group observations (e.g. children growing up in quintile 1) and movements (e.g. the estimated likelihood that a child moves from quantile 1 to quantile 5) if any observed quintile group contains less than 5 observations. This ensures that no child can be identified in the dataset as a result of disagregated data.

Income definitions

Children's household income is measured as the average CPI deflated sum of child and any spouses total income from 2010 to 2015. Parents' household income is measured as the average sum of mother's and father's summed income from 1980 to the year the child turns 15, 18, or 20. Total income is constructed using the Statistics Denmark defined variable PERINDKIALT for parents and PERINDKIALT_13 for children (For full description of the income variable, in Danish, see https://www.dst.dk/da/Statistik/dokumentation/Times/personindkomst/perindkialt and https://www.dst.dk/da/Statistik/dokumentation/Times/personindkomst/perindkialt-13).