PNAS
www.pnas.org

Supplementary Information for

Role of income mobility for the measurement of inequality in life expectancy

Claus Thustrup Kreiner, Torben Heien Nielsen and Benjamin Ly Serena

Claus Thustrup Kreiner.
E-mail: ctk@econ.ku.dk

This PDF file includes:

Supplementary text
Figs. S1 to S10
References for SI reference citations

Supporting Information Text

Data. The analysis is based on Danish administrative data provided by Statistics Denmark and covering the period 1980-2013 for the universe of Danish residents. We link the population registers, BEF and FAIN, containing information on gender and age, with the income register, IND, and the death registers, DODAARS and DODAASG, using personal identifiers (CPR number). Our income variable is household income net of universal transfers defined as:

$$\text{Income} = \text{PERINDKIALT} - \text{QPENSNY} - \text{KONTHJ} - \text{ANDOVERFORSEL} \quad [1]$$

PERINDKIALT is total income, QPENSNY is public pensions (disability pension, public retirement pension), KONTHJ (KONT_GL before 1994) is cash assistance, ANDOVERFORSEL is other universal public transfers. We show later that the results are robust to alternative income definitions. Household income is defined as the mean income of the individual and the spouse (if the individual of interest is either married or cohabiting). Following previous literature (1), we measure income two years before we measure mortality to reduce the importance of reverse causality.

Sample selection. The main sample consists of all individuals aged 40-100 during the years 1983-2013. This includes 81,953,961 person-year observations for 4,789,581 individuals. To estimate the mover-stayer model, we further restrict the main sample to individuals with non-missing lagged, current and lead income data, if alive. This restriction reduces the main sample to 81,344,597 person-year observations for 4,743,719 individuals. The validation sample used to test the predictive power of the methods in Fig. 2 is a subset of the main sample. Specifically, we follow cohorts of men aged 40 during 1983-1993 over 20 years from age 40 to 60 - a sample of 10,141,502 person-year observations for 443,977 individuals. We further restrict the validation sample to individuals with non-missing income from age 40 to 60 if alive. This restriction reduces the validation sample to 9,745,353 person-year observations for 420,813 individuals.

Income in retirement. When estimating the income gradient in life expectancy we use information on household income from age 40 to 100. After retirement, earnings are zero and the main source of variation in our income measure comes from payouts from private and employer-based pensions. Because these pensions are defined by contributions throughout the work life, they are closely tied to previous earnings. We show this below where we also show that the estimates are robust to using a Gompertz approximation (2, 3) from age 65 - the old-age-pension retirement age in Denmark.

Methodology. To execute the data analysis described in the manuscript section "Method to account for income mobility", we construct a distribution matrix containing the initial income distribution at age forty and transition matrices at each age spanning 40-100. Using these matrices we obtain predictions of future mobility and mortality by income class. We use these predictions to calculate the income gradient in life expectancy—with and without accounting for income mobility. Practically, we create the predictions of mobility and mortality in eight steps, corresponding to the steps in the Stata program provided in SI Appendix.

Prediction procedure.

Step 1: *Data construction.* We create lagged, current and lead income variables.

Step 2: *Sample selection.* We restrict the sample to individuals alive on January 1st in the previous year. We further restrict the sample to people with non-missing lagged, current and lead income data. However, we include people, who passed away in the previous or current year. The latter is necessary for constructing the absorbing death state in the transition matrices.

Step 3: *Ranks.* For living people we create income ranks by cohort and year (for lagged, current and lead income). Ties in the ranks are broken using a random draw. The states consist of the 100 income classes and 1 death state.

Step 4: *Transition probabilities and population weights.* We collapse the data by age and lagged, current and lead states. We then calculate conditional probabilities of all possible transitions from lagged and current states to current and lead states (the inputs $P_{j,i,k}$ in Table 2 of the manuscript). These calculations provide the transition probabilities for the transition matrices. We also construct population weights containing the distribution of individuals across lagged and current states. These weights are used to form the first distribution matrix at age 40.

Step 5: *Matrices.* We load the transition probabilities and population weights into matrices in Mata, Stata's integrated matrix programming language. We store the data in two matrices, Q and W. Q contains the 61 age-specific transition matrices, $T_{\hat{a}+t}^{\hat{a}+t+1}$, from age 40 to 100, where $\hat{a}$ is the initial age and t runs from 0 to 60. Each age-specific transition matrix $T_{\hat{a}+t}^{\hat{a}+t+1}$ describes the conditional probability of moving from a given state and lagged state at age $\hat{a}+t$ to all possible states at age $\hat{a}+t+1$. W contains the income distribution for all ages, i.e. the unconditional probabilities of all combinations of lagged and current states, with a row for each age. The first row of W contains the unconditional distribution at age 40—the diagonal elements in the initial distribution matrix $K_{\hat{a}}$ described in the manuscript (note that the initial distribution matrix $K_{\hat{a}}$ is a diagonal matrix, since it holds the relationship between states at age $\hat{a}$ and $\hat{a}$).

Step 6: *One-year death rates.* We change the death probabilities in the transition matrices, which depend both on lagged and current income, to depend on current income only. We do this to make our estimates comparable to standard estimates of life expectancy by income, where death rates depend only on current income.

Step 7: *Mover-stayer assumption.* We impose the mover-stayer assumption that individuals who stay in the same income percentile from the previous to the current year stay in that income percentile until death. Practically, we replace the transition probabilities for these cases with zeroes for all movements out of the income percentile, and one minus the death rate for staying in the current income percentile.

Step 8: *Predictions.* We predict mobility and mortality by income class. We first create a distribution matrix $K_{\hat{a}+t}$ based on the matrix W. The distribution matrix describes the bivariate distribution of lagged and current states between the initial age $\hat{a}$ and age $\hat{a}+t$, where t is 0 initially. We start a loop, continuing the following operations for each age (40-100):

8a We rank the distribution matrix $K_{\hat{a}+t}$. This is necessary because the income percentiles in the distribution matrix—in the no income mobility model, the mover-stayer model, and the second-order Markov model (but not in the first-order Markov model)—do not always contain one percent of survivors. This is most easily seen in the no income mobility model. Consider the first percentile of income, with the highest probability of death. Initially one percent of individuals are placed in each percentile. From the first to the second period, more people die in the 1$^{\text{st}}$ percentile than in the 100$^{\text{th}}$ percentile. Assuming there is no income mobility, less than one percent of survivors are therefore in the 1$^{\text{st}}$ percentile in the second period.

8b We calculate mobility rates, defined as the average income percentile at age $\hat{a}+t$ for each initial income percentile, using the unconditional probabilities in the distribution matrix $K_{\hat{a}+t}$.

8c We extract the age-specific transition matrix $T_{\hat{a}+t}^{\hat{a}+t+1}$ from the matrix Q.

8d We multiply the distribution matrix $K_{\hat{a}+t}$ with the transition matrix $T_{\hat{a}+t}^{\hat{a}+t+1}$ to obtain a new distribution matrix $K_{\hat{a}+t+1}$ describing the bivariate distribution of lagged and current states between age $\hat{a}$ and age $\hat{a}+t+1$.

8e We calculate death rates based on the new distribution matrix $K_{\hat{a}+t+1}$ as the share in each income state that moved to the death state from age $\hat{a}+t$ to age $\hat{a}+t+1$.

This loop provides the predictions of age-specific mobility and mortality rates by initial income class.

Calculating life expectancy. We calculate life expectancy by constructing survival curves, S_I, based on the predicted mortality rates. Life expectancy is defined as the area under the survival curve, $LE_I = \sum_{t=0} S_I^{\hat{a}+t}$.

Estimating the no income mobility model. When we estimate the income gradient in life expectancy without income mobility, we follow the same steps as in the mover-stayer model, described above. However, we skip steps 6-7 and set the transition probabilities of movements out of the initial income class to zero, and the probability of staying in the initial income class to one minus the death rate. Hence, individuals are assumed to stay in their initial income class until death.

Creating the figures. We create figures 1 and 2 in the manuscript in the following way:

Fig. 1 (*Main sample*):

- We implement the prediction procedure described above on the main sample and predict mobility and mortality rates by income class and age.
- We construct survival curves and calculate life expectancy by income class.
- For Fig. 1 A and B, the prediction procedure is implemented on pooled data for all years during 1983-2013.
- For Fig. 1 C and D, the prediction procedure is implemented separately for each year during 1983-2013.
- Fig. 1 A and B plot the estimated life expectancies by income class during 1983-2013 with and without accounting for income mobility, for males and females.
- Fig. 1 C and D plot the difference between the estimated life expectancy at percentiles 80 and 20 with and without accounting for income mobility from 1983 to 2013, for males and females.

Fig. 2 (*Validation sample*):

- We estimate predictions of mobility and mortality rates by income class and age (40-60) on the validation sample.
- We calculate actual mobility and mortality rates from age 40 to 60, which are directly observed in the data.
- We construct actual and predicted survival curves by income class.
- Fig. 2 A and B plot the survival curves averaged within the bottom 5 and top 5 percentiles of the income distribution.
- Fig. 2C plots actual and predicted mobility rates, defined as the average position in the income distribution at age 60, for each income class at age 40.
- Fig. 2D plots the prediction error in mortality from age 40 to 60 for each income class. We calculate accumulated mortality from age 40 to 60 using the survival curves. The prediction error is defined as the percentage deviation from actual to predicted mortality.

Results with alternative models of income mobility. The predictions of mobility and mortality presented in the main text are based on the mover-stayer model of social mobility. Here, we show that other approaches to incorporating income mobility also provide better estimates of future mortality than assuming no income mobility. Still, all alternative approaches are dominated by the mover-stayer model. Fig. S1 compares the predictions of the mover-stayer model, first- and second-order Markov models, and the no income mobility model. The estimations of the first- and second-order Markov models follow the same prediction procedure as described above. For the first-order Markov model we skip steps 6-7. Instead, we change the transition matrices $T_{\hat{a}+t}^{\hat{a}+t+1}$ to depend on current income only. For the second-order Markov model we skip step 7. Fig. S1A displays the prediction error (root mean squared error measured in percentiles) in actual versus predicted income percentiles from age 40 to 60 for each method. The smaller the prediction error, the more accurate is the prediction. Clearly, the mover-stayer model dominates the other models of mobility. At the end of the 20-year period, the prediction error is two percentiles. In comparison, the 20-year prediction error is approximately 6 percentiles for the second-order Markov model and 12 percentiles for the first-order Markov model, while the no income mobility assumption leads to the largest 20-year prediction error of 13 percentiles. Fig. S1B plots the corresponding prediction errors for the mortality rates. The prediction error in actual versus predicted mortality from age 40 to 60 is 0.01 percentage points for the mover-stayer model, 0.02 for the second-order Markov model, 0.03 for the first-order Markov model, and 0.04 for the no income mobility model. Hence, even though the first-order Markov model is a poor predictor of income mobility, the model provides better estimates of mortality than the no income mobility model.

Income classes based on average income over several years. The intergenerational mobility literature has shown that using three- or five-year income averages—rather than one-year income measures—reveal higher intergenerational correlation. This result arguably occurs because individual averages over a long horizon reduce the importance of temporary variations in income and better proxy permanent income (4). Therefore, we ask whether the no income mobility model improves simply by using income from several years, and whether this outperforms the mover-stayer model, which relies on data from fewer periods. Fig. S2 illustrates the effectiveness of the alternative approach by estimating mobility and mortality rates based on the average of income over 1, 2, 3, 4 and 5 years. To create the five-year income average at age 40, we use data on income for the ages 34-38. For our main analyses based on the mover-stayer model, we use income at age 37 (lagged income) and 38 (current income). Because of the three additional years required to create 5-year income averages, the results in Fig. S2 are estimated on cohorts aged 40 in 1986-1993 (note that Fig. 2. samples cohorts aged 40 in 1983-1993).

Fig. S2A plots the average income percentile at age 60 by the initial income percentile at age 40, using the different income averages. Each year added to the income measure slightly reduces mobility rates. Still, taking income averages does not eliminate concerns caused by income mobility. Fig. S2B plots the difference between the actual and predicted mortality from age 40 to 60 when using the no income mobility model with the different income averages. We compare this to the mover-stayer model estimated using one-year income. The prediction error from invoking the no mobility assumption decreases as we add years to the income measure. But this reduction in the bias does not amount to the improvements obtained from using the mover-stayer model applied with the baseline one-year income variable. Hence, compared to using income averages, the mover-stayer model produces far more precise estimates of mortality by income class, even though it requires fewer income observations per individual.

Education as a socio-economic determinant of mortality. Education length is an alternative and commonly used proxy for socioeconomic status (5). In the following we find that income classes are informative of life expectancy conditional on education length and that these education-specific income gradients are also halved when incorporating income mobility. Fig. S3 plots the income gradient in life expectancy by income, split by education groups of 0-12, 13-14, and 15-21 years of education. Information on education length is only available for cohorts born after 1920. Therefore, we use education-specific death rates up until age 83 measured in 2003-2013. From age 83, we apply population-wide death probabilities to all education and income groups. Fig. S3A plots the results of the no income mobility model and Fig. S3B plots the results for the mover-stayer model. Both methods conclude that more years of education is associated with higher life expectancy. However, large differences in life expectancy based on income percentile persist, even within each education group. For the mover-stayer estimates, the unconditional difference in life expectancy between percentile 20 and percentile 80 is 2.7 years. Within education groups, the difference is 2.1-2.3 years. This suggests that much of the variation in mortality across socioeconomic groups is not captured solely by differences in education levels. Moreover, for each education group the difference between the 20^{th} and 80^{th} percentile of income is half as large when using the mover-stayer model compared to the no income mobility model. This suggests that even conditional on education, it is important to take income mobility into account.

Sensitivity to number of income classes. The share of stayers in the mover-stayer model is determined by the probability that an individual stays in the same income class from one period to the next. With many income classes, this probability is obviously lower compared to using only a few income classes. We have divided individuals into percentiles (one hundred income classes) in our empirical application. In the following we test the validity of our method using fewer quantiles. To obtain consistent and accurate estimates across different quantiles, we propose and test the following simple procedure: If a person stays in the same income class two periods in a row $(a-1, a)$ then the share of stayers is defined as

$$s = \frac{P_{i,i}}{Q_{i,i}}, \tag{2}$$

where $P_{i,i}$ is the observed probability of staying in the same income percentile two periods in a row and $Q_{i,i}$ is the observed probability of staying in the same income class two periods in a row. We implement this in the transition matrices by adding an absorbing 'stayer state' for each income class. Consequently, the probability of entering the stayer state is equal to s, and the probability of staying there is one minus the death rate.

Fig. S4 shows the accuracy of the mobility prediction of the mover-stayer model when the estimates are based on 5, 10, 20, and 100 quantiles of income. For ease of comparison, we show the estimates collapsed to 5 income classes. In all four panels of Fig. S4, the predicted mobility is very close to the actual data, and the errors are not systematically larger when the estimates are based on fewer income quantiles. Below we use this approach to estimate the mover-stayer model on a survey panel data set from the US where fewer groups are needed.

External validity and the role of disabled individuals. Our main results are all based on data from Denmark. To show that our methods also apply to other countries, we estimate the mover-stayer model on the US survey Panel Study of Income Dynamics (PSID). Restricting the PSID data on a balanced panel of household heads through a ten year period during 1968-1997, the sample consists of 3,500 individuals. Given the relatively smaller sample size (compared to the population-wide administrative data) we use income quintiles. Fig. S5A plots the predicted and actual average income quintile after ten years for each initial income quintile. Similar to the results on Danish data, the mover-stayer model provides very accurate predictions of mobility on PSID data. Fig. S5 B and C investigate how the American life expectancy gradient changes when accounting for income mobility. A recent study (1) estimates the income gradient in life expectancy in the US during 2001-2014 using tax return data. While we do not have access to this data and, therefore, cannot estimate mobility-adjusted counterparts to these estimates directly, we make an imputation. Specifically, we account for income mobility in the estimation of the income gradient in life expectancy in the US by combining mobility rates estimated on Danish data with death rates from the US study, obtained from www.healthinequality.org. Hence, we assume that mobility rates are similar in the Danish and the US environments. We evaluate this crucial assumption in Fig. S5B by plotting the rank-rank correlation between the initial and current income rank over a ten year period. The data on rank-rank correlations for the US is also obtained from www.healthinequality.org. In both data sets, the correlation between the current and initial income rank decreases over time, reflecting income mobility. More importantly, the degree of income mobility observed in the Danish and the US data is close to identical. In Fig. S5C, we show the estimated income gradient in life expectancy in the US, with and without accounting for income mobility. Without mobility, the difference in life expectancy between the 20^{th} and the 80^{th} income percentile is 6.5 years. When accounting for income mobility this difference reduces to 3 years.

The steep increase in life expectancy in the lower tail of the income distribution in Denmark, shown in Fig. 1, is more pronounced than in the results for the US, shown in Fig. S5C. This discrepancy occurs because the US study excludes non-filers (individuals with income below the IRS filing threshold, e.g. recipients of Social Security Disability benefits) and individuals with zero income, who constitute up to 30 % of an age group in the US data (calculated based on data obtained from www.healthinequality.org). The income gradient in life expectancy in Denmark is estimated using the entire population. Fig. S5D shows the income gradient in life expectancy in Denmark during 2001-2013, excluding individuals with disability benefits or zero income (87 % of the excluded individuals are disability benefit recipients). When we impose this sample restriction, the steep increase in life expectancy in the lower tail of the income distribution is almost gone. Moreover, the income gradient in life expectancy is still halved when accounting for income mobility.

Out-of-sample predictions of mobility and mortality. The predictions of mobility and mortality shown in Fig. 2 are in-sample predictions, meaning that the predictions are created and evaluated on the same data. In-sample predictions can be subject to overfitting problems. To address this concern, we test the out-of-sample prediction of the models. We partition our panel dataset into two equally sized and randomly drawn samples. The first sample is used to estimate the parameters of the models and create a prediction of mobility and mortality. The second random sample is used to test the accuracy of the prediction. Fig. S6 A and B replicate Fig. 2 C and D using this split-sample approach. The results are similar to the original estimates. The no income mobility model provides systematically biased results, while the mover-stayer model provides precise predictions of both mobility and mortality.

Predicting mobility and mortality at other ages. Table 3 in the manuscript presents estimates of the income gradient in life expectancy at age 40, 50, 60, 70 and 80. Income mobility is most important for life expectancy measurement at age 40, 50 and 60. In Fig. S7 we replicate the validation exercise in Fig. 2 C and D in the manuscript for age 50 and age 60 and show that the mover-stayer model also provides accurate predictions of mobility and mortality from age 50 to 70 and age 60 to 80. Fig. S7 A and B show the results for age 50. Income mobility is substantial from age 50 to 70 and the predictions of the mover-stayer model are close to the actual mobility rates. When assuming no income mobility, mortality from age 50 to 70 is overestimated by up to 50 % in the bottom of the income distribution and underestimated by about 20 % in the top of the distribution. The mover-stayer model outperforms the no income mobility model with small prediction errors across income classes. Fig. S7 C and D show the accuracy of the predictions of mobility and mortality from age 60 to 80. The mover-stayer model produces accurate predictions of income mobility and much more precise predictions of mortality from age 60 to 80 than assuming no income mobility. There is slightly less income mobility from age 60 to 80 than in earlier ages and the percentage prediction error in mortality from age 60 to 80 from assuming no income mobility is lower compared to estimates from age 40 to 60 and age 50 to 70. The difference in prediction errors is due to a higher base level of mortality from age 60 to 80. In percentage points, the prediction errors in mortality from assuming no income mobility are not much smaller from age 60 to 80 than from age 40 to 60 and 50 to 70.

Sensitivity to income definition and income class measurement after retirement. The predictions of mobility and mortality in Fig. 2 are based on household income net of universal transfers. As a sensitivity test, Fig. S8 A and B show the accuracy of predictions of mobility and mortality using a more narrow income measure including only earnings (LOENMV from Statistics Denmark). The predictions of mobility of the mover-stayer model based on this income concept, displayed in Fig. S8A, are slightly less precise in the top and bottom of the income distribution compared to the predictions based on the standard income measure illustrated in Fig. 2. The additional prediction error is caused by self-employed individuals, who are often temporarily in the tails of the earnings distribution. However, as shown in Fig. S8B, the mover-stayer model produces very accurate mortality predictions across the income distribution.

When estimating the income gradient in life expectancy, we use information on household income from age 40 to 100 to group individuals by income classes and age. After retirement, earnings are zero and the main source of variation in our income measure derives from payouts from private and employer-based pensions. Because these pensions are based on contributions made throughout working life, they are closely tied to previous earnings. This is demonstrated in Fig. S8C, which plots the relationship between the income percentile one year before and after retirement. We add a placebo estimate, showing the relationship between the income percentile three years before retirement and one year before retirement. This provides a baseline estimate of mobility against which we judge the additional mobility caused by retirement. The 45-degree line in Fig. S8C represents the case with no income mobility. Retirement does not fundamentally change the income position of individuals in our sample. The red dots, representing the average post-retirement income percentile for each pre-retirement income percentile, lie close to the 45-degree line and the placebo estimates. This suggests that our income measure is able to identify differences in earnings capacity even after retirement.

The Gompertz approximation is an often-used strategy to impute death rates at progressed ages (2, 3). The Gompertz approximation assumes that log death rates are linear in age. Fig. S8D shows the income gradient in life expectancy for males using a Gompertz approximation from age 65—the old-age-pension retirement age in Denmark. The Gompertz approximation does not provide good predictions of mortality in very old ages, and from age 80 to 100 we instead assume that all individuals have the same population-wide death probability. The resulting estimates of the income gradient in life expectancy, shown in Fig. S8D, are very similar to our baseline results from Fig. 1. The difference in life expectancy between the 20$^{\text{th}}$ and 80$^{\text{th}}$ percentile of income is 4.8 years without accounting for income mobility and 1.8 years when accounting for income mobility.

Mobility rates in the cross-section. When incorporating income mobility in the estimation of period life expectancy we use mobility rates of older individuals as an estimate of future mobility of current 40-year-olds. By leveraging our panel data we test that this cross-section assumption is valid. In particular, we estimate mobility rates based on cross-section data using older cohorts to predict future mobility. From these cross-sectional mobility rates we predict panel data mortality rates. We compare these mobility and mortality predictions to mobility and mortality predictions based on the panel data and to actual observed mobility and mortality rates. The result is shown in Fig. S9, which replicates Fig. 2 C and D, including estimates based on cross-section data. The mobility predictions in Fig. S9A based on the cross-section data are almost identical to the panel data estimates (red and green dots coincide) and to actual mobility. The prediction error in mortality in Fig. S9B when estimating the mobility rates on cross-section data is also very close to the panel data estimates, and in both cases the average prediction error is very small.

Changes in mobility over time. Fig. 1C in the main text shows that the difference in life expectancy between income percentiles 20 and 80 increased from 2.5 years in 1983 to 6.5 years in 2013 if we do not account for income mobility. When we incorporate income mobility this trend is half as large. Therefore, the discrepancy between estimates of life expectancy inequality with and without income mobility is increasing. In Fig. S10 we show that this growing gap is not caused by an increase in income mobility over time. We hold mobility constant at the 1983-level by applying mobility rates estimated in 1983 to all years during 1983-2013 in the mover-stayer model. In this way, only changes in death rates over time can affect the estimated inequality in life expectancy. The result is shown in Fig. S10A, along with the original estimates from Fig. 1C. While the level of inequality is slightly higher when we hold mobility rates constant at the 1983-level, the estimated change in inequality over time is exactly the same as in the original estimates. The reason is that mobility is more or less constant during the period. This is seen in Fig. S10B, which plots the predicted average income percentile at age 60 by income percentile at age 40 for the years 1983, 1993, 2003 and 2013. Income mobility from age 40 to 60 is close to identical in all four years.

Note also that since the difference in life expectancy between percentiles 80 and 20 in each of the years, for both men and women, is consistently about half as big when accounting for mobility (the horizontal distance between the black and red dots) it follows that the change in life expectancy inequality over time is also about half as big (the difference in the slopes of the black and red prediction lines). This explains the growing gap between the estimates of the income gradient in life expectancy with and without accounting for income mobility.

References

1. Chetty R, et al. (2016) The association between income and life expectancy in the United States, 2001-2014. *JAMA - Journal of the American Medical Association* 315(16):1750–1766.
2. Gompertz B (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 115:513–583.
3. Preston S, Heuveline P, Guillot M (2000) *Demography: Measuring and Modeling Population Processes.* (Wiley-Blackwell).

4. Solon G (1999) Chapter 29 Intergenerational mobility in the labor market in *Handbook of Labor Economics.* (Elsevier Masson SAS) Vol. 3, pp. 1761–1800.
5. Hummer R, Lariscy J (2011) Chapter 12 Educational Attainment and Adult Mortality in *International Handbook of Adult Mortality.* Vol. 2, pp. 289–310.
6. PSID (2017) Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Institute for Social Research, University of Michigan, Ann Arbor, MI, 2017.

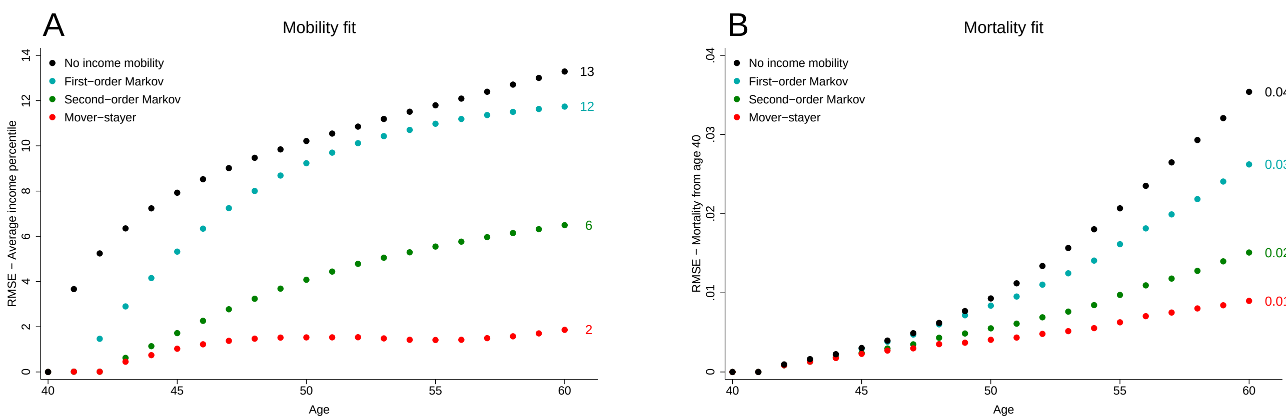

Fig. S1. Alternative models of income mobility

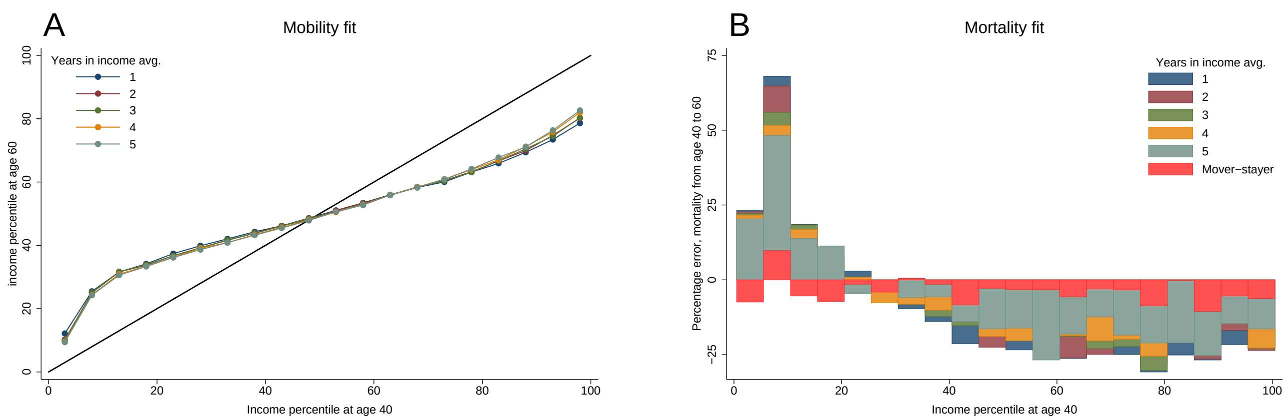

Fig. S2. No mobility model with income classes based on several years

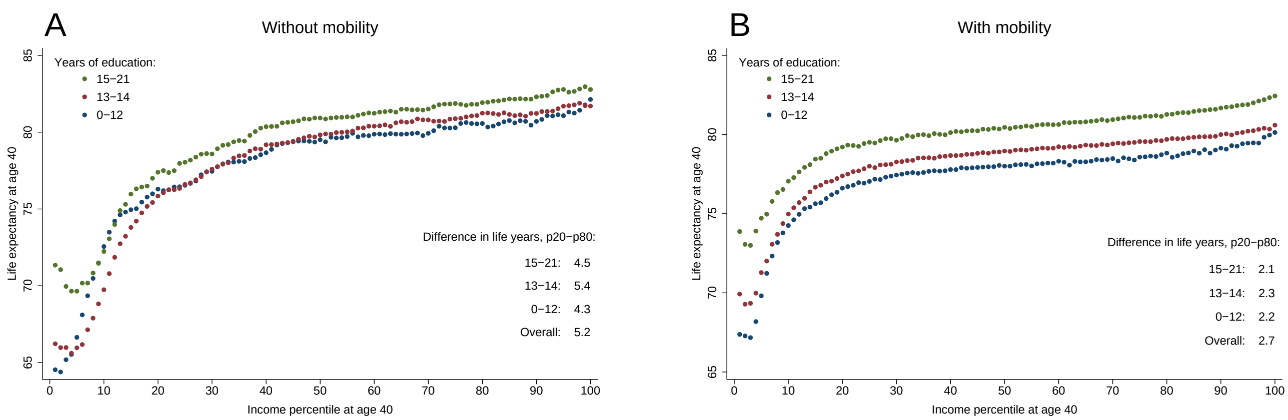

Fig. S3. Estimates of the income gradient in life expectancy by education groups

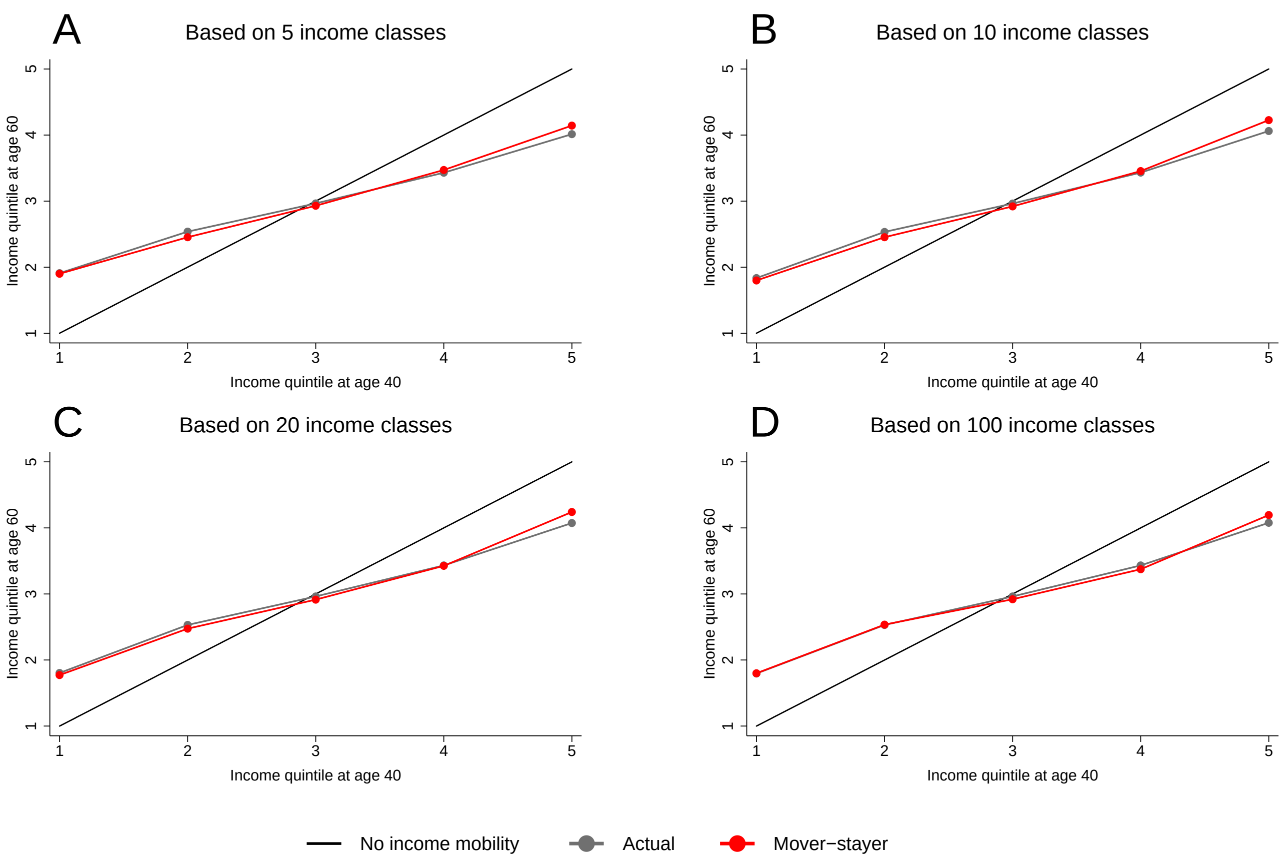

Fig. S4. Estimates of the mover-stayer model based on 5, 10, 20, and 100 income groups

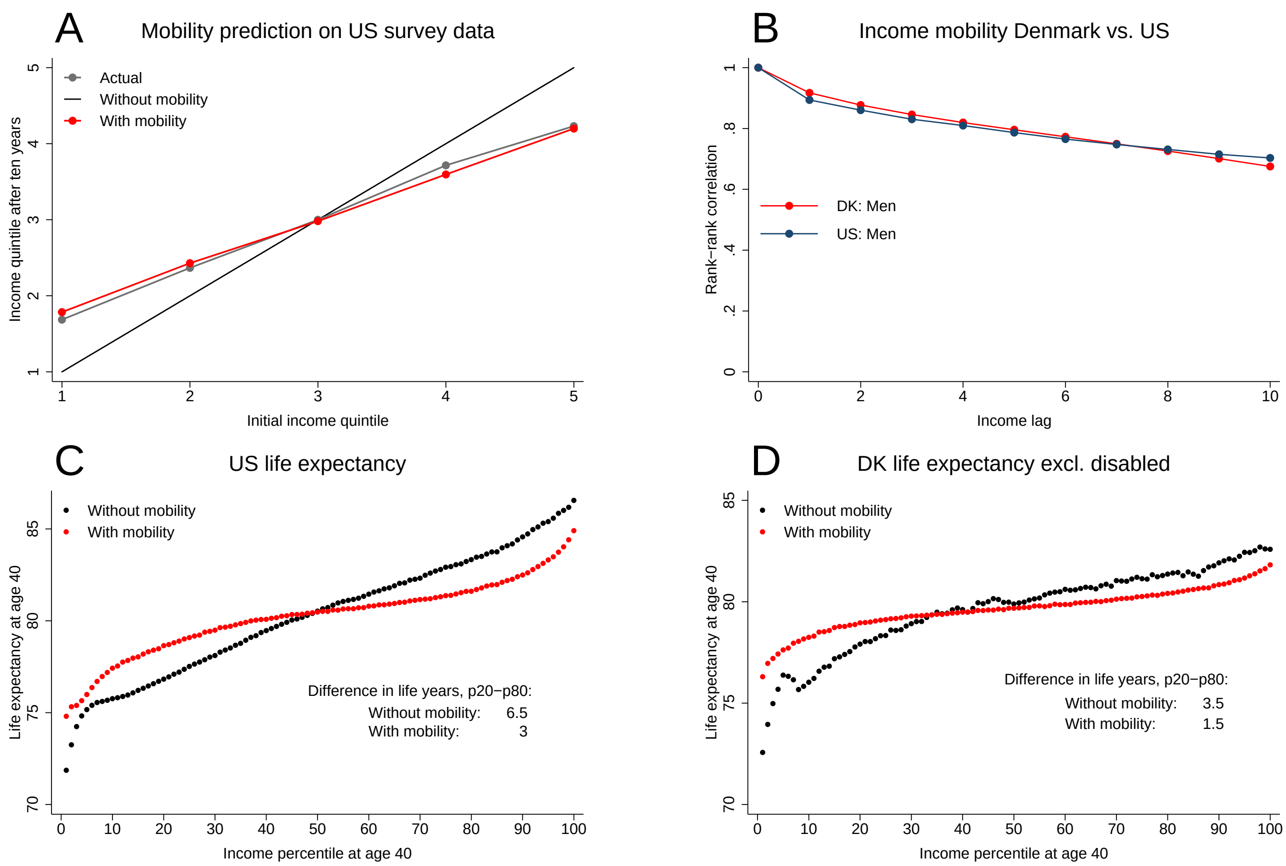

Fig. S5. External Validity. Comparison with the US (PSID and www.healthinequality.org).

Note: Fig. S5 contains results based on external data sources. Fig. S5 A uses publicly available PSID (Panel Survey of Income Dynamics) data (6), downloaded on 12/11/2017. Figs. S5 B and C use publicly available data from the authors of (1), made available at www.healthinequality.org under a CC0 license, downloaded on 07/27/2017. Fig. S5 B uses data on rank-rank correlations from www.healthinequality.org to plot rank-rank correlations of income in the US measured with different lags alongside comparable estimates based on Danish data. Fig. S5 C uses data on income and age-specific mortality rates from www.healthinequality.org, combined with mobility rates estimated on Danish data, to compute the income gradient in life expectancy in the US with and without accounting for income mobility.

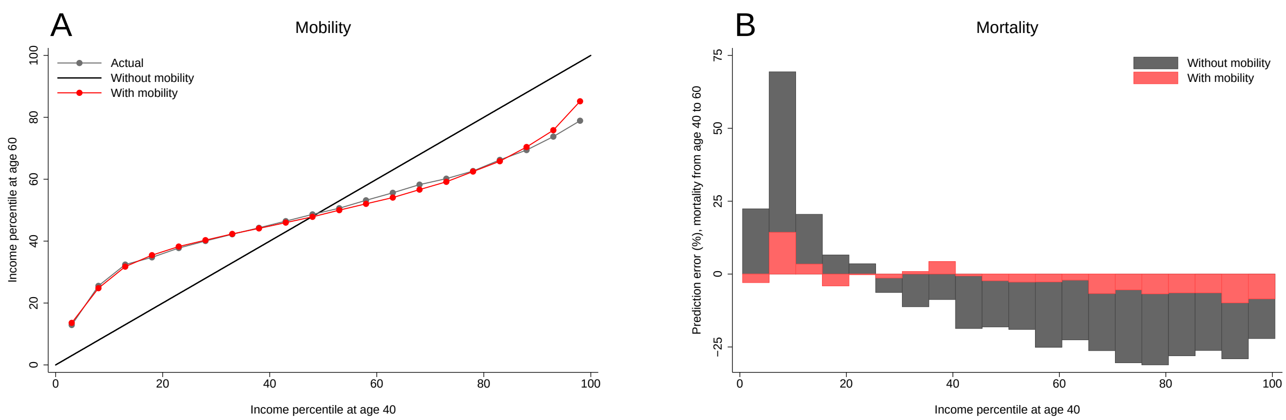

Fig. S6. Out-of-sample prediction of the mover-stayer model

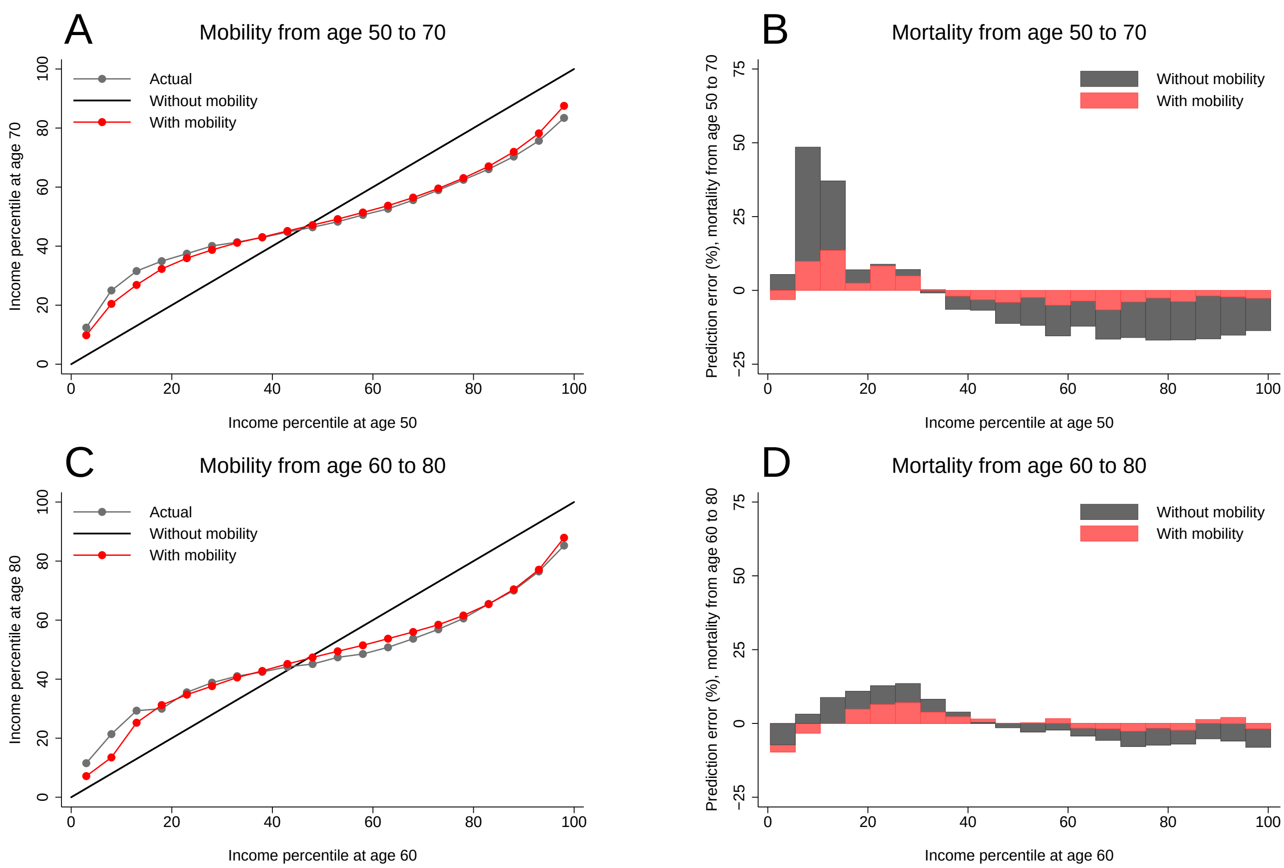

Fig. S7. Predicting mobility and mortality at other ages

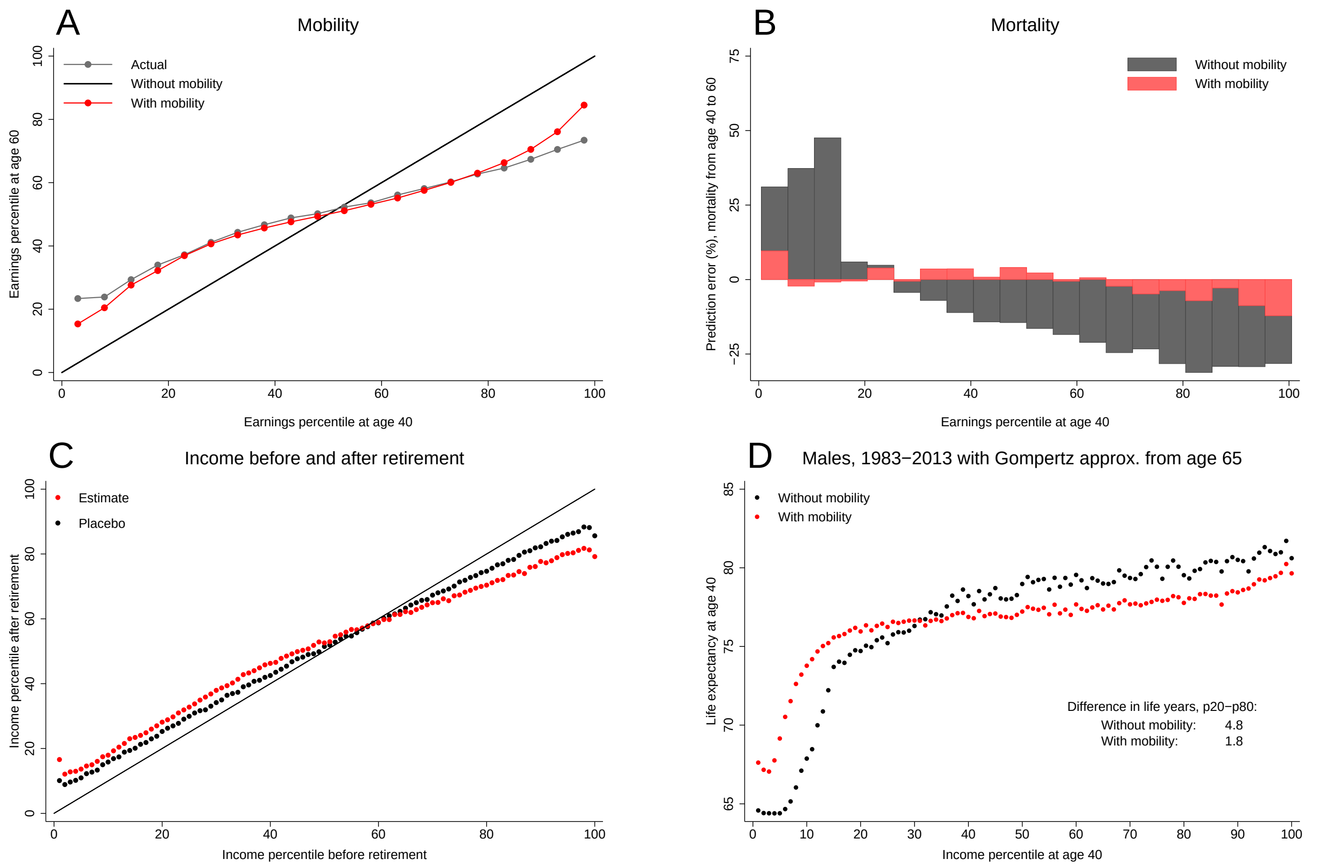

Fig. S8. Accuracy of income class measurement after retirement

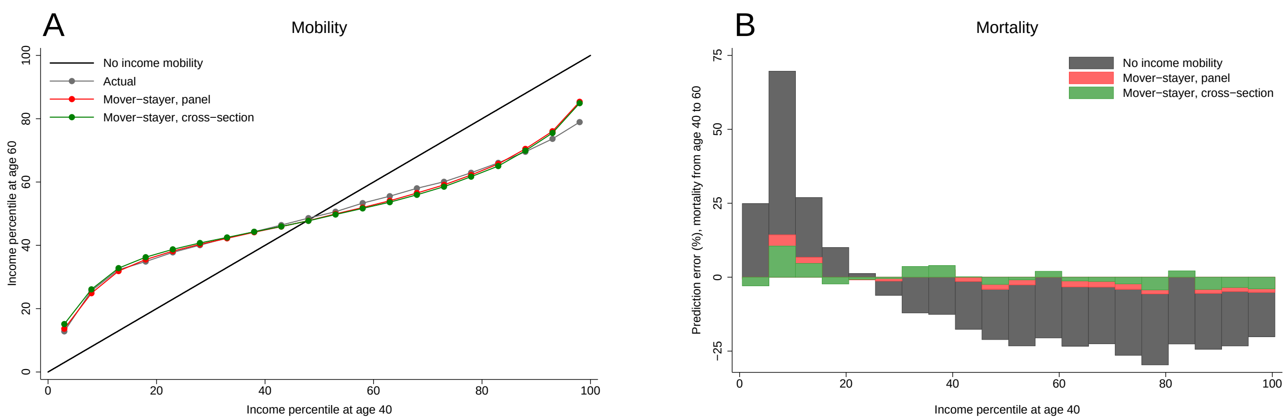

Fig. S9. Estimates of the mover-stayer model when based on cross-section data

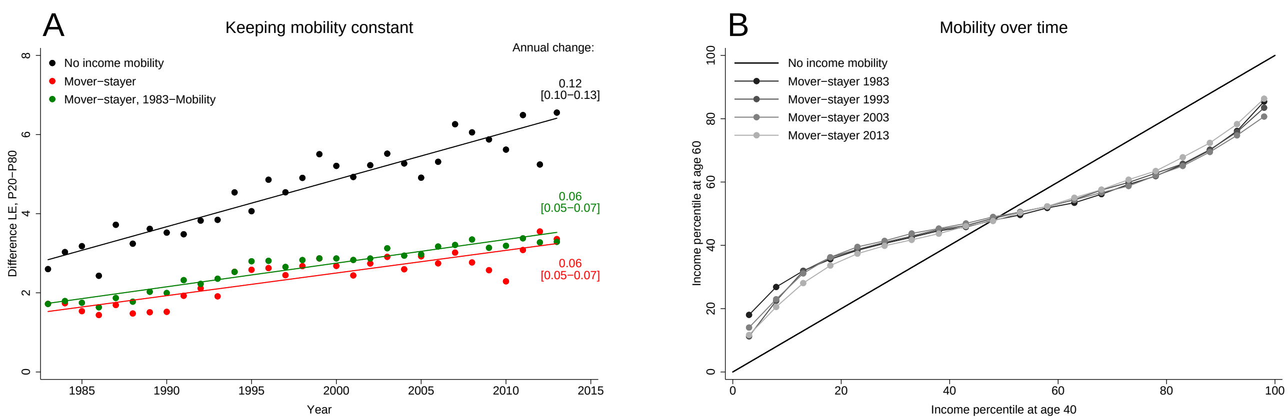

Fig. S10. Change in mobility over time

```
/////////////////////////////////////////////////////////////////
// STATA Program to implement mover-stayer model.              //
// Last edited 02/07/2018 by Benjamin.Ly.Serena@econ.ku.dk     //
// Steps in program follow steps described in SI "Methodology" //
/////////////////////////////////////////////////////////////////

cap program drop LifeExp
program LifeExp, eclass

syntax varlist(min=2 max=2) [if] , idvar(varlist max=1) agevar(varlist max=1) yearvar(varlist max=1) from(numlist
max=1) to(numlist max=1) ///
[preserve byvar(varlist max=1) QUANtiles(numlist max=1) method(namelist min=1 max=4) save(string) ///
INPUTDeath(namelist max=1) INPUTTrans(namelist  min=3 max=3) OUTPUTTrans(namelist min=3 max=3) INPUTRanks(namelist
min=3 max=3) ratio(numlist max=1 min=1)]

/* Varlist
1 Income
2 Death dummy
*/

/** Errors **/
if "`byvar'"!="" {
cap confirm numeric variable `byvar'
if _rc {
di as error "Error: Only numeric byvars allowed"
exit
}
}
if ("`inputranks'"!="" & "`ratio'"=="") {
di as error "Option inputranks must be combined with ratio"
exit
}
/** Options **/
* Default number of income groups = 100 *
if "`quantiles'"=="" {
local quantiles=100
}
* Default methods *
if "`method'"=="" {
local method "M1 NIM M2 MS"
}
* Input ranks and stayer share *
if ("`inputranks'"!="") {
qui tokenize `inputranks'
local lagincrank "`1'"
local incrank "`2'"
local leadincrank "`3'"
global ratio=`ratio'
}
* Retrieve methods *
qui tokenize `method'
local m1="`1'"
local m2="`2'"
local m3="`3'"
local m4="`4'"

if ("`inputranks'"=="") {
qui tokenize `varlist'
}

/********************/
/*** Estimation   ***/
/********************/
if ("`preserve'"!="") {
preserve
}

quietly {
if ("`inputranks'"=="") {
qui marksample touse

* Years in data *
qui sum `yearvar'
local ymin=r(min)
local ymax=r(max)

* Create observation for dead people the year after death *
tempvar lagdead
qui expand 2 if `2'==1 , gen(`lagdead')

qui replace `yearvar'=`yearvar'+1 if `lagdead'==1
qui replace `agevar'=`agevar'+1 if `lagdead'==1
qui replace `1'=. if `lagdead'==1

/* STEP 1: Create lagged, current and lead income */
tempvar laginc inc leadinc

sort `idvar' `yearvar'
```

```
by `idvar': gen `laginc' =`1'[_n-3]     if `idvar'==`idvar'[_n-3] & `yearvar'==`yearvar'[_n-3]+3
by `idvar': gen `inc'    =`1'[_n-2]     if `idvar'==`idvar'[_n-2] & `yearvar'==`yearvar'[_n-2]+2
by `idvar': gen `leadinc'=`1'[_n-1]     if `idvar'==`idvar'[_n-1] & `yearvar'==`yearvar'[_n-1]+1

/* STEP 2: Restrict sample to individuals with non-missing lagged, current and lead income if alive */
qui keep if !mi(`laginc') & ((!mi(`inc') & `lagdead'==0) | `lagdead'==1)  & ((!mi(`leadinc') & `lagdead'==0 &
`2'==0) | `lagdead'==1 | `2'==1)

* Additional sample restrictions *
keep if `touse'

/* STEP 3: Create income ranks for lagged, current and lead income */
set seed 112
tempvar random
qui gen double `random'=runiform()
qui replace `random'=(`random'-0.5)/_N

* Lagged income *
tempvar nonmiss
qui gen `nonmiss'=(!mi(`laginc') & !mi(`yearvar') & !mi(`agevar') )

sort `agevar' `yearvar' `nonmiss' `laginc' `random'

tempvar lagincrank maxincrank
by `agevar' `yearvar' `nonmiss': gen double `lagincrank'=_n
by `agevar' `yearvar' `nonmiss': gen double `maxincrank'=_N
qui replace `lagincrank'=ceil((((`lagincrank'-0.5)/`maxincrank')*100+`random')/(100/`quantiles'))
qui replace `lagincrank'=. if `nonmiss'==0

drop `maxincrank' `nonmiss'

* Current income *
tempvar nonmiss
qui gen `nonmiss'=(`lagdead'==0 & !mi(`inc') & !mi(`yearvar') & !mi(`agevar') )

sort `agevar' `yearvar' `nonmiss' `inc' `random'

tempvar incrank
by `agevar' `yearvar' `nonmiss': gen double `incrank'=_n
by `agevar' `yearvar' `nonmiss': gen double `maxincrank'=_N
if ("`m1'"=="MS" | "`m2'"=="MS" | "`m3'"=="MS" | "`m4'"=="MS" )  {
tempvar ori_incrank
gen `ori_incrank'=ceil((((`incrank'-0.5)/`maxincrank')*100 +`random'))
}
qui replace `incrank'=ceil((((`incrank'-0.5)/`maxincrank')*100 +`random')/(100/`quantiles'))
qui replace `incrank'=. if `nonmiss'==0

drop `nonmiss' `maxincrank'

* Lead income *
tempvar nonmiss
qui gen `nonmiss'=(`2'==0 & `lagdead'==0 & !mi(`leadinc') & !mi(`yearvar') & !mi(`agevar') )

sort `agevar' `yearvar' `nonmiss' `leadinc' `random'

tempvar leadincrank
by `agevar' `yearvar' `nonmiss': gen double `leadincrank'=_n
by `agevar' `yearvar' `nonmiss': gen double `maxincrank'=_N
if ("`m1'"=="MS" | "`m2'"=="MS" | "`m3'"=="MS" | "`m4'"=="MS" )  {
tempvar ori_leadincrank
gen `ori_leadincrank'=ceil((((`leadincrank'-0.5)/`maxincrank')*100 +`random'))
}
qui replace `leadincrank'=ceil((((`leadincrank'-0.5)/`maxincrank')*100 +`random')/(100/`quantiles'))
qui replace `leadincrank'=. if `nonmiss'==0

drop `nonmiss' `maxincrank'

* Create Death Category *
qui sum `incrank'
qui replace `lagincrank'=r(max)+1 if mi(`lagincrank')
qui replace `incrank'=r(max)+1 if mi(`incrank')
qui replace `leadincrank'=r(max)+1  if mi(`leadincrank')

qui keep if inrange(`agevar',`from',`to')
qui keep if inrange(`yearvar',`ymin'+3,`ymax')

* Determine share of stayers in mover stayer (See SI section "Sensitivity to number of income classes")*
if ("`m1'"=="MS" | "`m2'"=="MS" | "`m3'"=="MS" | "`m4'"=="MS" )  {
qui sum `incrank'
tempvar match100 match
gen `match100'=(`ori_incrank'==`ori_leadincrank') if `ori_incrank'!=101 & `ori_leadincrank'!=101
gen `match'=(`incrank'==`leadincrank') if `incrank'!=r(max) & `leadincrank'!=r(max)

qui sum `match100'
local m100=r(mean)
qui sum `match'
global ratio=`m100'/r(mean)
}
```

```
}

if "`byvar'"!="" {
qui sum `byvar'
global Bmin=r(min)
global Bmax=r(max)
}
if "`byvar'"=="" {
global Bmin=1
global Bmax=1
}

/* STEP 4: Collapse data and calculate conditional probabilities */
tempvar tot
qui gen `tot'=1
qui collapse (sum) `tot' if !mi(`agevar') & !mi(`incrank') & !mi(`lagincrank') & !mi(`leadincrank'), by(`byvar'
`agevar' `incrank' `lagincrank' `leadincrank') fast

qui sum `lagincrank'
local max=r(max)
local max1=r(max)+1
local min=r(min)

qui count
local c=r(N)+1
qui set obs `c'
tempvar num
qui gen `num'=_n

* Add dead-dead to dead-dead transition - not observed in the data *
qui replace `lagincrank'=`max1'  if `num'==`c'
qui replace `incrank'=`max1'     if `num'==`c'
qui replace `agevar'=`from'      if `num'==`c'
qui replace `leadincrank'=`max1' if `num'==`c'
if "`byvar'"!="" {
qui replace `byvar'=${Bmin} if `num'==`c'
}

* Lagged-current income category *
tempvar laggroup
qui gen `laggroup'=(`lagincrank'-1)*`max1'+`incrank'

* Current-lead income category *
tempvar leadgroup
qui gen `leadgroup'=(`incrank'-1)*`max1'+`leadincrank'

* Transition probabilities *
qui sum `laggroup'
local max1=r(max)

tempvar pers pers2
bysort `byvar' `agevar' `laggroup': egen `pers'=sum(`tot')
bysort `byvar' `agevar': egen `pers2'=sum(`tot')

tempvar Dperc
qui gen `Dperc'=`tot'/`pers'
qui replace `Dperc'=0 if mi(`Dperc')

* Set dead-dead to dead-dead transition probability equal to 1 *
qui replace `Dperc'=1 if `laggroup'==`max1' & `leadgroup'==`max1'

* Weights for initial distribution matrix *
tempvar weight
qui gen `weight'=`pers'/`pers2'
qui replace `weight'=0 if mi(`weight')

* Create globals with number of age and income groups *
qui sum `leadgroup'
global d=r(max)-r(min)+1

qui sum `agevar'
global a=r(max)-r(min)+1

* Create placement in matrix indicator *
tempvar panelvar
qui gen double `panelvar'=(`agevar'-r(min))*`max1'*`max1'+(`laggroup'-1)*`max1'+`leadgroup'

sort `byvar' `panelvar'
keep `byvar' `panelvar' `Dperc' `weight'
order `byvar' `panelvar' `Dperc' `weight'
}

// Conduct estimation //
forvalues B=${Bmin}/${Bmax} {
if "`byvar'"=="" {
local B=.
}
    /* STEP 5: Load transition probabilities and population weights */
```

```
    mata Q=J(${d},${d}*${a},.)
    mata W=J(1,${d}*${a},.)
    mata Wa=J(1,sqrt(${d})*${a},.)

    mata LoadMatrix(${a},${d},Q,W,Wa,`B')

    if ("`inputtrans'"!="") {
    tokenize `inputtrans'
    mata InputT(${a},${d},Q,W,`1')
    mata W=`2'
    mata Wa=`3'
    }
    if ("`inputdeath'"!="") {
    mata InputD(${a},${d},Q,`inputdeath')
    }
    if ("`outputtrans'"!="") {
    tokenize `outputtrans'
    mata `1'=Q
    mata `2'=W
    mata `3'=Wa
    }
    /* The rest of the steps (6-8) are performed in the programs defined below, depending on the method */

    /* STEP 8: First order markov (skips steps 6-7) */
    if ("`m1'"=="M1" | "`m2'"=="M1" | "`m3'"=="M1" | "`m4'"=="M1" )   {
    mata TMM1=J(${a},sqrt(${d})-1,.)
    mata DRM1=J(${a},sqrt(${d})-1,.)

    mata M1(${a},${d},TMM1,DRM1,Q,W)
    }
    /* STEP 8: No income mobility  (skips steps 6-7) */
    if ("`m1'"=="NIM" | "`m2'"=="NIM" | "`m3'"=="NIM" | "`m4'"=="NIM" )   {
    mata DRNIM=J(${a},sqrt(${d})-1,.)

    mata NIM(${a},${d},DRNIM,Q,W,Wa)
    }

    /* STEP 6: Change to one-year death rates */
    if ("`m1'"=="M2" | "`m2'"=="M2" | "`m3'"=="M2" | "`m4'"=="M2" ) | ("`m1'"=="MS" | "`m2'"=="MS" | "`m3'"=="MS" |
    "`m4'"=="MS" )  {
    mata OneYDR(${a},${d},Q,W)
    }

    /* STEP 8: Second order markov (skips step 7) */
    if ("`m1'"=="M2" | "`m2'"=="M2" | "`m3'"=="M2" | "`m4'"=="M2" )  {
    mata TMM2=J(${a},sqrt(${d})-1,.)
    mata DRM2=J(${a},sqrt(${d})-1,.)

    mata M2(${a},${d},TMM2,DRM2,Q,W,Wa)
    }

    /* STEP 7: Impose mover stayer assumption */
    if ("`m1'"=="MS" | "`m2'"=="MS" | "`m3'"=="MS" | "`m4'"=="MS" )  {
    mata IMSA(${a},${d},Q,${ratio})

    /* STEP 8: Mover stayer */
    mata TMMS=J(${a},sqrt(${d})-1,.)
    mata DRMS=J(${a},sqrt(${d})-1,.)

    mata MS(${a},${d},TMMS,DRMS,Q,W,Wa)
    }

    * Save results *
    foreach m of local method {
        if "`byvar'"!="" {
        if (`B'==${Bmin}) mata DR`m'F=J(sqrt(${d})-1,1,`B'),DR`m'';
        if (`B'>${Bmin})  mata DR`m'F=DR`m'F\J(sqrt(${d})-1,1,`B'),DR`m'';
        if (`B'==${Bmin} & "`m'"!="NIM") mata TM`m'F=J(sqrt(${d})-1,1,`B'),TM`m'';
        if (`B'>${Bmin}  & "`m'"!="NIM") mata TM`m'F=TM`m'F\J(sqrt(${d})-1,1,`B'),TM`m'';
        }
        if "`byvar'"=="" {
        mata DR`m'F=DR`m'';
        if ("`m'"!="NIM") mata TM`m'F=TM`m'';
        }
    }
}
* Drop matrices *
mata mata drop Q W Wa

/******************/
/*** Save data  ***/
/******************/
quietly {

local savevars=""
forvalues a=`from'/`to' {
local savevars="`savevars'"+" a`a'"
}
```

```
* Save Death Rate Data *
local j=1
foreach m of local method {
if (`j'==1) mata Dres=DR`m'F;
if (`j'!=1) mata Dres=Dres\DR`m'F;
local j=`j'+1
}

clear
getmata (`byvar' `savevars')=Dres

tempvar var
gen `var'=floor((_n-1)*wordcount("`method'")/_N)

gen method=""
local j=0
foreach m of local method {
replace method="`m'" if `var'==`j'
local j=`j'+1
}

tempvar num
gen `num'=_n
sort `byvar' `var' `num'
by `byvar' `var': gen incg=_n

drop `var' `num'
sort `byvar' method incg
order `byvar' method incg *
reshape long a, i(`byvar' method incg) j(`agevar')
rename a deathrate

if "`save'"!="" {
save "`save'_DeathRates", replace
}

* Save mobility rate Data *
if ("`m1'"!="NIM" | ("`m2'"!="NIM" & "`m2'"!="") ) {

local j=1
foreach m of local method {
if (`j'==1 & "`m'"!="NIM") mata Mres=TM`m'F;
if (`j'!=1 & "`m'"!="NIM") mata Mres=Mres\TM`m'F;
local j=`j'+1
}

clear
getmata (`byvar' `savevars')=Mres

tempvar var
if ("`m1'"=="NIM" | "`m2'"=="NIM" | "`m3'"=="NIM" | "`m4'"=="NIM" ) {
gen `var'=floor((_n-1)*(wordcount("`method'")-1)/_N)
}
if ("`m1'"!="NIM" & "`m2'"!="NIM" & "`m3'"!="NIM" & "`m4'"!="NIM" ) {
gen `var'=floor((_n-1)*wordcount("`method'")/_N)
}

gen method=""
local j=0
foreach m of local method {
    if ("`m'"!="NIM"){
    replace method="`m'" if `var'==`j'
    local j=`j'+1
    }
}
tempvar num
gen `num'=_n
sort `byvar' `var' `num'
by `byvar' `var': gen incg=_n

drop `var' `num'
sort `byvar' method incg
order `byvar' method incg *
reshape long a, i(`byvar' method incg) j(`agevar')
rename a avincg

save "`save'_MobilityRates", replace
}
}
if ("`preserve'"!="") {
restore
}
end


/***************************/
/****** Help Programs ******/
```

```
/***************************/

/* Load Matrices (STEP 5) */
cap mata: mata drop LoadMatrix()
mata
function LoadMatrix(a,d,Q,W,Wa,B) {
d2=sqrt(d)
d2m=d2-1
TW=J(d2*a,1,((1..d*a):-J(1,a,(floor((0..(d-1))/d2)):*d2))):== J(a,1,(1::d2))+(floor((0..(d2*a-1))/d2))':*d

Dat=J(d*d*a,2,0)
if (B==.) Dat[st_data(.,1),1]=st_data(.,2);
if (B==.) Dat[st_data(.,1),2]=st_data(.,3);
if (B!=.) Dat[select(st_data(.,2),st_data(.,1):==B),1]=select(st_data(.,3),st_data(.,1):==B);
if (B!=.) Dat[select(st_data(.,2),st_data(.,1):==B),2]=select(st_data(.,4),st_data(.,1):==B);

Q=colshape(Dat[,1],d)'
W=colmax(colshape(Dat[,2],d)')

Wa=W*TW'
Wa=rowshape(Wa,a)'
Wa=Wa:/colsum(Wa[1..d2m,])
Wa=Wa[1..d2m,]

Q=editmissing(Q,0):+(colsum(Q):==0):*J(1,a,I(d))
Q[,J(1,a,d2*(1..d2))+floor((0..a*d2-1)/d2):*d]=J(1,a,(J(d-1,d2,0)\J(1,d2,1)))
}
end

/* Change death rates to one-year death rates (STEP 6) */
cap mata: mata drop OneYDR()
mata
function OneYDR(a,d,Q,W) {
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))
RW=J(1,d2,I(d2))
RW[,(d2m*d2+1)..d]=J(d2,d2,0)

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    Q2=Q[,l..u]

    Q2[(1..d2m):*d2,]=RW[1..d2m,]:*(Q[(1..d2m):*d2,l..u]*W[,l..u]'):/(RW[1..d2m,]*W[,l..u]')


    Q2[RR,RR]=(Q2[RR,RR]:*editmissing((colsum(Q[,RR:+(l-1)])-colsum(Q2[d2:*(1..d2),RR])):/(colsum(Q2[RR,RR])),0)):*(co
    lsum(Q[(1..d2m):*d2,RR:+(l-1)]):!=1)+(I(d2m*d2m)-diag(colsum(Q2[d2:*(1..d2),RR]))):*(colsum(Q[(1..d2m):*d2,RR:+(l-
    1)]):==1)

    Q[,l..u]=Q2
    }

Q=editmissing(Q,0):+(colsum(Q):==0):*J(1,a,I(d))
}
end

/* Impose mover stayer assumption (STEP 7) */
cap mata: mata drop IMSA()
mata
function IMSA(a,d,Q,ratio) {
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))

DR=J(1,a-2,(1..d2m):*d2-(d2m..1))+floor((d2m*2..a*d2m-1)/(d2m)):*(d+d2m)
DRQ=J(1,a-2,(1..d2m):*d2-(d2m..1))+floor((d2m*2..a*d2m-1)/(d2m)):*(d)
DM=J(1,a-2,d:+(1..d2m))+floor((d2m*2..a*d2m-1)/(d2m)):*(d+d2m)

Q3=J(1,a,I(d+d2m))
Q3[1..d,J(1,a,(1..d))+floor((0..a*d-1)/d):*(d+d2m)]=Q
Q3[RR,DR]=Q[RR,DRQ]:*(1-ratio)
Q3[d:+(1..d2m),DR]=J(1,a-2,I(d2m)):*colsum(Q[RR,DRQ])*ratio
Q3[d2:*(1..d2m),DM]=Q[d2:*(1..d2m),DRQ]
Q3[d:+(1..d2m),DM]=J(1,a-2,I(d2m))-Q[d2:*(1..d2m),DRQ]

Q=Q3
}
end

/* Input transition matrix (See SI section "Changes in mobility over time")*/
cap mata: mata drop InputT()
mata
function InputT(a,d,Q,W,Qb) {
```

```
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))
RW=J(1,d2,I(d2))
RW[,(d2m*d2+1)..d]=J(d2,d2,0)

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    Q2=Qb[,l..u]

    Q2[(1..d2m):*d2,]=RW[1..d2m,]:*(Q[(1..d2m):*d2,l..u]*W[,l..u]'):/(RW[1..d2m,]*W[,l..u]')


    Q2[RR,RR]=(Q2[RR,RR]:*editmissing((colsum(Q[,RR:+(l-1)])-colsum(Q2[d2:*(1..d2),RR])):/(colsum(Q2[RR,RR])),0)):*(co
    lsum(Q[(1..d2m):*d2,RR:+(l-1)]):!=1)+(I(d2m*d2m)-diag(colsum(Q2[d2:*(1..d2),RR]))):*(colsum(Q[(1..d2m):*d2,RR:+(l-
    1)]):==1)

    Q[,l..u]=Q2
    }

Q=editmissing(Q,0):+(colsum(Q):==0):*J(1,a,I(d))
}
end

/* Input death matrix (See SI section "External validity and the role of disabled individuals") */
cap mata: mata drop InputD()
mata
function InputD(a,d,Q,Db) {
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))
RW=J(1,d2,I(d2))
RW[,(d2m*d2+1)..d]=J(d2,d2,0)

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    Q2=Q[,l..u]

    Q2[(1..d2m):*d2,]=J(1,d2,(diag(Db[i,]),J(d2m,1,0)))


    Q2[RR,RR]=(Q2[RR,RR]:*editmissing((colsum(Q[,RR:+(l-1)])-colsum(Q2[d2:*(1..d2),RR])):/(colsum(Q2[RR,RR])),0)):*(co
    lsum(Q[(1..d2m):*d2,RR:+(l-1)]):!=1)+(I(d2m*d2m)-diag(colsum(Q2[d2:*(1..d2),RR]))):*(colsum(Q[(1..d2m):*d2,RR:+(l-
    1)]):==1)

    Q[,l..u]=Q2
    }

Q=editmissing(Q,0):+(colsum(Q):==0):*J(1,a,I(d))
}
end

/*********************************/
/****** Estimation Programs ******/
/*********************************/

/* First order markov */
cap mata: mata drop M1()
mata
function M1(a,d,TMM1,DRM1,Q,W) {
d2=sqrt(d)
d2m=d2-1
// Define initial distribution matrix //
K1=diag(colsum((W[,1..d]:*(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2)))'))
LK1=K1

T1=J(d2,d2,0)

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    // STEP 8b - Calculate and save mobility rates (STEP 8a is redundant in Markov 1) //
    TMM1[i,]=colsum(diag(1..d2m)*K1[1..d2m,1..d2m]):/colsum(K1[1..d2m,1..d2m])

    // STEP 8c - Extract age-specific transition matrix //

    T1=(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2))*Q[,l..u]*(W[,l..u]:*(J(d2,1,((1..d):-(floor((0..(d-1))
    /d2)):*d2)):==(1::d2)))':/(colsum((W[,l..u]:*(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2)))'))
    T1=editmissing(T1,0):+(colsum(T1):==0):*I(d2)

    // STEP 8d - Multiply distribution matrix with transition matrix //
```

```
    LK1=K1
    K1=T1*K1

    // STEP 8e - Calculate and save death rates //
    DRM1[i,]=(colsum(LK1[1..d2m,1..d2m])-colsum(K1[1..d2m,1..d2m])):/colsum(LK1[1..d2m,1..d2m])
    }
}
end

/*  No income mobility */
cap mata: mata drop NIM()
mata
function NIM(a,d,DRNIM,Q,W,Wa) {
d2=sqrt(d)
d2m=d2-1
// Define initial distribution matrix //
KN=diag(colsum((W[,1..d]:*(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2)))'))
LKN=KN

T1=J(d2,d2,0)

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    // STEP 8a - Rank distribution matrix //
    KN=RankNIM(KN,d2m,Wa[,i])

    // STEP 8c - Extract age-specific transition matrices (STEP 8b is redundant without mobility) //

    T1=(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2))*Q[,l..u]*(W[,l..u]:*(J(d2,1,((1..d):-(floor((0..(d-1))
    /d2)):*d2)):==(1::d2)))':/(colsum((W[,l..u]:*(J(d2,1,((1..d):-(floor((0..(d-1))/d2)):*d2)):==(1::d2)))'))
    T1=editmissing(T1,0):+(colsum(T1):==0):*I(d2)
    TN=I(d2)-diag(T1[d2,])
    TN[d2,]=T1[d2,]

    // STEP 8d - Multiply distribution matrix with transition matrix //
    LKN=KN
    KN=TN*KN

    // STEP 8e - Calculate and save death rates //
    DRNIM[i,]=(colsum(LKN[1..d2m,1..d2m])-colsum(KN[1..d2m,1..d2m])):/colsum(LKN[1..d2m,1..d2m])
    }
}
end

/** Second order markov **/
cap mata: mata drop M2()
mata
function M2(a,d,TMM2,DRM2,Q,W,Wa) {
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))
FW=J(d,d2m,0)
for (k=1; k<=d2m; k++) {
FW[d2:*(1..d2)-J(1,d2,d2-k),k]=J(d2,1,1)
}
CD=J(1,d2,I(d2))
NZ=select(1..d,W[1..d]:!=0)

// Define initial distribution matrix //
K2=diag(W[,1..d])
K2=K2[,NZ]

    for (i=1; i<=a; i++) {
    l=(i-1)*d+1
    u=i*d

    // STEP 8a - Rank distribution matrix //
    K2=RankM2(K2,d2,d2m,d,CD,Wa[,i])

    // STEP 8b - Calculate and save mobility rates //

    TMM2[i,]=colsum(diag((1..dg)-((ceil((1..dg)/d2m))-J(1,dg,1))*d2m)*K2[RR,]*CD[1..d2m,NZ]'):/colsum(K2[RR,]*CD[1..d2
    m,NZ]')

    // STEP 8c - Extract age-specific transition matrices //
    T=Q[,l..u]

    // STEP 8d - Multiply distribution matrix with transition matrix //
    K2=T*K2

    // STEP 8e - Calculate and save death rates //
    DRM2[i,]=colsum(K2[d2:*(1..d2m),]*FW[NZ,]):/colsum(K2[1..d2m*d2,]*FW[NZ,])
    }
}
end
```

```
/** Mover-stayer **/
cap mata: mata drop MS()
mata
function MS(a,d,TMMS,DRMS,Q,W,Wa) {
d2=sqrt(d)
d2m=d2-1
dg=(d-d2)*(d2-1)/d2
RR=(1..dg)+ floor((0..d-2*d2)/(d2-1))
FW=J(d,d2m,0)
for (k=1; k<=d2m; k++) {
FW[d2:*(1..d2)-J(1,d2,d2-k),k]=J(d2,1,1)
}
CD=J(1,d2+1,I(d2))
CD2=CD[,1..d+d2m]
CD=CD[1..d2m,1..d]
NZ=select(1..d,W[1..d]:!=0)

// Define initial distribution matrix //
KMS=diag((W[,1..d],J(1,d2m,0)))
KMS=KMS[,NZ]
CKMS=KMS[1..d,]

    for (i=1; i<=a; i++) {
    l=(i-1)*(d+d2m)+1
    u=i*(d+d2m)

    // STEP 8a - Rank distribution matrix //
    KMS=RankMS(KMS,d2,d2m,d,CD2,Wa[,i])

    // STEP 8b - Calculate and save mobility rates //
    TMMS[i,]=colsum(diag((1..dg)-((ceil((1..dg)/d2m))-J(1,dg,1))*d2m)*CKMS[RR,]*CD[,NZ]'):/colsum(CKMS[RR,]*CD[,NZ]')

    // STEP 8c - Extract age-specific transition matrices //
    T=Q[,l..u]

    // STEP 8d - Multiply distribution matrix with transition matrix //
    KMS=T*KMS

    // STEP 8e - Calculate and save death rates //
    CKMS=KMS[1..d,]
    CKMS[(1..d2m):*(d2+1):-d2,]= KMS[(1..d2m):*(d2+1):-d2,]+KMS[(d+1)..d+d2m,]

    DRMS[i,]=colsum(CKMS[d2:*(1..d2m),]*FW[NZ,]):/colsum(CKMS[1..d2m*d2,]*FW[NZ,])
    }
}
end


/*********************/
/* Ranking programs  */
/*********************/

* Second order markov *
cap mata: mata drop RankM2()
mata
function RankM2(XUR,d2,d2m,d,CD,Wag) {
CX=CD*XUR
Mis=Wag:*sum(CX[1..d2m,])-(rowsum(CX[1..d2m,]))
KW=J(d2,cols(XUR),rowsum(CX[1..d2,]))
cW=editmissing(XUR:/KW,0)

if (max(Mis)>0 | min(Mis)<0)
    for (o=1; o<d2m; o++) {
    no=o+1

    if (Mis[o,]!=0) {
    sumx=sum(CX[no,])
    Io=d2:*(1..d2)-J(1,d2,d2-o)
    Ino=d2:*(1..d2)-J(1,d2,d2-no)

    if (sumx>Mis[o,1]) {
    XUR[Io,]=XUR[Io,]+(cW[Ino,]:*Mis[o,1]):*(Mis[o,1]>=0)+(cW[Io,]:*Mis[o,1]):*(Mis[o,1]<0)
    XUR[Ino,]=XUR[Ino,]-(cW[Ino,]:*Mis[o,1]):*(Mis[o,1]>=0)-(cW[Io,]:*Mis[o,1]):*(Mis[o,1]<0)
    }
    if (sumx<=Mis[o,1]) {
    XUR[Io,]=XUR[Io,]+(cW[Ino,]:*sumx):*(Mis[o,1]>=0)
    XUR[Ino,]=J(d2,cols(XUR),0)
    rest=Mis[o,1]-sumx
        for (p=o+2; p<=d2m; p++) {
        Ip=d2:*(1..d2)-J(1,d2,d2-p)
        sumx=sum(CX[p,])
        lrest=rest
        XUR[Io,]=XUR[Io,]+(cW[Ip,]:*rest):*(rest<sumx)+(cW[Ip,]:*sumx):*(rest>=sumx)
        XUR[Ip,]=(XUR[Ip,]-cW[Ip,]:*rest):*(rest<sumx)+J(d2,cols(XUR),0):*(rest>=sumx)
        rest=0*(lrest<sumx)+(lrest-sumx)*(lrest>=sumx)
        if (rest==0) break;
        }
```

```
    }
        CX=CD*XUR
        if (Mis[o,1]<0) {
        cW[Ino,]=editmissing(XUR[Ino,]:/rowsum(CX[no,]),0)
        }
    Mis=Wag:*sum(CX[1..d2m,])-(rowsum(CX[1..d2m,]))
    }
    }
;
_edittozerotol(XUR,0.0000000000000000001)
return(XUR)
}
end

* No income mobility *
cap mata: mata drop RankNIM()
mata
function RankNIM(XUR,d2m,W) {
X=XUR
Mis=W:*sum(X[1..d2m,])-rowsum(X[1..d2m,])
rowW=editmissing(X[1..d2m,]:/rowsum(X[1..d2m,]),0)

     for (o=1; o<d2m; o++) {
     no=o+1

     sumx=sum(X[no,])
     if (sumx>Mis[o,1]) {
     X[o,]=X[o,]+(rowW[no,]:*Mis[o,1]):*(Mis[o,1]>=0)+(rowW[o,]:*Mis[o,1]):*(Mis[o,1]<0)
     X[no,]=X[no,]-(rowW[no,]:*Mis[o,1]):*(Mis[o,1]>=0)-(rowW[o,]:*Mis[o,1]):*(Mis[o,1]<0)
     }
     if (sumx<=Mis[o,1]) {
     X[o,]=X[o,]+(rowW[no,]:*sumx):*(Mis[o,1]>=0)
     X[no,]=J(1,d2m+1,0)
     rest=Mis[o,1]-sumx

             for (p=o+2; p<=d2m; p++) {
             sumx=sum(X[p,])
             lrest=rest
             X[o,]=X[o,]+(rowW[p,]:*rest):*(rest<sumx)+(rowW[p,]:*sumx):*(rest>=sumx)
             X[p,]=(X[p,]-rowW[p,]:*rest):*(rest<sumx)+J(1,d2m+1,0):*(rest>=sumx)
             rest=0*(lrest<sumx)+(lrest-sumx)*(lrest>=sumx)
             if (rest==0) break;
             }
     }
     Mis=W:*sum(X[1..d2m,])-rowsum(X[1..d2m,])
     rowW=editmissing(X[1..d2m,]:/rowsum(X[1..d2m,]),0)
     }
return(X)
}
end

* Mover-stayer  *
cap mata: mata drop RankMS()
mata
function RankMS(XUR,d2,d2m,d,CD,Wag) {
CX=CD*XUR
Mis=Wag:*sum(CX[1..d2m,])-(rowsum(CX[1..d2m,]))
KW=J(d2+1,cols(XUR),rowsum(CX[1..d2,]))
KW=KW[1..d+d2m,]
cW=editmissing(XUR:/KW,0)

if (max(Mis)>0 | min(Mis)<0)
    for (o=1; o<d2m; o++) {
    no=o+1

    if (Mis[o,]!=0) {
    sumx=sum(CX[no,])
    Io=d2:*(1..d2+1)-J(1,d2+1,d2-o)
    Ino=d2:*(1..d2+1)-J(1,d2+1,d2-no)

    if (sumx>Mis[o,1]) {
    XUR[Io,]=XUR[Io,]+(cW[Ino,]:*Mis[o,1]):*(Mis[o,1]>=0)+(cW[Io,]:*Mis[o,1]):*(Mis[o,1]<0)
    XUR[Ino,]=XUR[Ino,]-(cW[Ino,]:*Mis[o,1]):*(Mis[o,1]>=0)-(cW[Io,]:*Mis[o,1]):*(Mis[o,1]<0)
    }
    else {
    XUR[Io,]=XUR[Io,]+(cW[Ino,]:*sumx):*(Mis[o,1]>=0)
    XUR[Ino,]=J(d2+1,cols(XUR),0)
    rest=Mis[o,1]-sumx
        for (p=o+2; p<=d2m; p++) {
        Ip=d2:*(1..d2+1)-J(1,d2+1,d2-p)
        sumx=sum(CX[p,])
        lrest=rest
        XUR[Io,]=XUR[Io,]+(cW[Ip,]:*rest):*(rest<sumx)+(cW[Ip,]:*sumx):*(rest>=sumx)
        XUR[Ip,]=(XUR[Ip,]-cW[Ip,]:*rest):*(rest<sumx)+J(d2+1,cols(XUR),0):*(rest>=sumx)
        rest=0*(lrest<sumx)+(lrest-sumx)*(lrest>=sumx)
        if (rest==0) break;
        }
    }
```

```
        CX=CD*XUR
        if (Mis[o,1]<0) {
        cW[Ino,]=editmissing(XUR[Ino,]:/rowsum(CX[no,]),0)
        }
    Mis=Wag:*sum(CX[1..d2m,])-(rowsum(CX[1..d2m,]))
    }
    }
;
_edittozerotol(XUR,0.0000000000000000001)
return(XUR)
}
end
```

```
///////////////////////////////////////////////////////////////////////
// STATA program to calculate actual mobility and mortality rates  //
// Last edited 03/07/2018 by Benjamin.Ly.Serena@econ.ku.dk          //
///////////////////////////////////////////////////////////////////////

cap program drop CohortLifeExp
program CohortLifeExp, eclass

syntax varlist(min=2 max=2) [if] , idvar(varlist max=1) agevar(varlist max=1) yearvar(varlist max=1) from(numlist
max=1) to(numlist max=1) ///
[byvar(varlist max=1) uncond QUANtiles(numlist max=1) INPUTRanks(namelist min=1 max=1) OUTPUTDeath(namelist min=1
max=1) save(string)]

/* Varlist
1 Income
2 Death dummy
*/

/** Errors **/
if "`byvar'"!="" {
cap confirm numeric variable `byvar'
if _rc {
di as error "Error: Only numeric byvars allowed"
exit
}
}
/** Options **/
* Default number of income groups = 100 *
if "`quantiles'"=="" {
local quantiles=100
}
if ("`inputranks'"!="") {
qui tokenize `inputranks'
local incrank "`1'"
}
qui tokenize `varlist'

/********************/
/*** Estimation   ***/
/********************/
preserve
quietly {
if ("`inputranks'"=="") {
qui marksample touse

* Years in data *
qui sum `yearvar'
local ymin=r(min)
local ymax=r(max)

* Create observation for dead people the year after death *
tempvar lagdead
qui expand 2 if `2'==1 , gen(`lagdead')

qui replace `yearvar'=`yearvar'+1 if `lagdead'==1
qui replace `agevar'=`agevar'+1 if `lagdead'==1
qui replace `1'=. if `lagdead'==1

/* Create lagged, current and lead income */
tempvar laginc inc leadinc

sort `idvar' `yearvar'
by `idvar': gen `laginc' =`1'[_n-3]     if `idvar'==`idvar'[_n-3] & `yearvar'==`yearvar'[_n-3]+3
by `idvar': gen `inc'    =`1'[_n-2]     if `idvar'==`idvar'[_n-2] & `yearvar'==`yearvar'[_n-2]+2
by `idvar': gen `leadinc'=`1'[_n-1]     if `idvar'==`idvar'[_n-1] & `yearvar'==`yearvar'[_n-1]+1

/* Restrict sample to individuals with non-missing lagged, current and lead income if alive */
qui keep if !mi(`laginc') & ((!mi(`inc') & `lagdead'==0) | `lagdead'==1)  & ((!mi(`leadinc') & `lagdead'==0 &
`2'==0) | `lagdead'==1 | `2'==1)

* Additional sample restrictions *
keep if `touse'

/* Create income ranks for current income */
set seed 112
tempvar random
qui gen double `random'=runiform()
qui replace `random'=(`random'-0.5)/_N

tempvar nonmiss
qui gen `nonmiss'=(`lagdead'==0 & !mi(`inc') & !mi(`yearvar') & !mi(`agevar') )

sort `agevar' `yearvar' `nonmiss' `inc' `random'
tempvar incrank maxincrank
by `agevar' `yearvar' `nonmiss': gen double `incrank'=_n
by `agevar' `yearvar' `nonmiss': gen double `maxincrank'=_N
qui replace `incrank'=ceil(((( `incrank'-0.5)/`maxincrank')*100 +`random')/(100/`quantiles'))
qui replace `incrank'=. if `nonmiss'==0
```

```
drop `nonmiss' `maxincrank'

* Create Death Category *
qui sum `incrank'
qui replace `incrank'=r(max)+1 if mi(`incrank')

qui keep if inrange(`agevar',`from',`to')
keep if `lagdead'==0
}

* Initial income class *
sort `idvar' `yearvar'
tempvar firstincrank
by `idvar': gen `firstincrank'=`incrank'[1]

/* Collapse data and calculate conditional probabilities */

* Split by initial income class *
if ("`uncond'"=="") {
qui collapse (mean) `2' `incrank' if !mi(`agevar') & !mi(`incrank') , by(`byvar' `agevar' `firstincrank') fast

gen incg=`firstincrank'
gen avincg=`incrank'
gen deathrate=`2'
drop `firstincrank' `incrank' `2'
gen method="Act"

sort  method `byvar' incg `agevar'
order method `byvar' incg `agevar' deathrate avincg
}
* Split by current income class *
if ("`uncond'"!="") {
qui collapse (mean) `2'  if !mi(`agevar') & !mi(`incrank') , by(`byvar' `agevar' `incrank') fast

gen incg=`incrank'
gen deathrate=`2'
drop `incrank' `2'
gen method="Act"

sort  method `byvar' incg `agevar'
order method `byvar' incg `agevar' deathrate
}

if ("`save'"!="") {
save "`save'", replace
}
if ("`outputdeath'"!="") {
keep incg `agevar' deathrate

qui reshape wide deathrate, i(`agevar') j(incg)
drop `agevar'
mata `outputdeath'=st_data(.,.)
}
restore
}
end
```

```
/*****************************************************************************************/
/*** The role of income mobility for the measurement of inequality in life expectancy    ***/
/***                                                                                     ***/
/***                                    Contents:                                        ***/
/***                                                                                     ***/
/***                            1. Generating Data                                       ***/
/***                            2. Main figures and tables                               ***/
/***                            3. SI figures                                            ***/
/***                                                                                     ***/
/*****************************************************************************************/
/*** Last edited 17/09/2018 by Benjamin.Ly.Serena@econ.ku.dk                             ***/
/*****************************************************************************************/


/*******************************************************/
/***                1.  Generating Data                ***/
/*******************************************************/

clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Data"
use deathincome

/* Only use observations between 1980 - 2013 */
keep if inrange(year,1980,2013)

/* Remove persons under age 27 */
rename alder age
drop if age<27

* Drop unnecessary variables *
drop  C_DODSMAADE

/* Generate income variables */
gen ERHVERVSINDK_GL_ori=ERHVERVSINDK_GL
gen ERHVERVSINDK_GLy_ori=ERHVERVSINDK_GLy
gen loenmv_ori=loenmv
gen loenmvy_ori=loenmvy
foreach var in tilbtot andoverforsel overforsindk OVSKEJD02_NY ERHVERVSINDK_GL NETOVSKUD_GL loenmv {
replace `var'=0 if mi(`var')
}
foreach var in  TILBTOTy ANDOVERFORSELy OVERFORSINDKy OVSKEJD02_NYy ERHVERVSINDK_GLy NETOVSKUD_GLy loenmvy {
replace `var'=0 if mi(`var') & !mi(pnry)
}
/* Take out imputed value of rent in houses - causes data break */
replace FORMUEINDK_NY=FORMUEINDK_NY-OVSKEJD02_NY
replace FORMUEINDK_NYy=FORMUEINDK_NYy-OVSKEJD02_NYy

/* Take out negative profits, capital income and wage income */
replace ERHVERVSINDK_GL=ERHVERVSINDK_GL-NETOVSKUD_GL if NETOVSKUD_GL<0
replace ERHVERVSINDK_GL=ERHVERVSINDK_GL-loenmv if loenmv<0
replace ERHVERVSINDK_GLy=ERHVERVSINDK_GLy-NETOVSKUD_GLy if NETOVSKUD_GLy<0
replace ERHVERVSINDK_GLy=ERHVERVSINDK_GLy-loenmvy if loenmvy<0
replace perindkialt=perindkialt-NETOVSKUD_GL if NETOVSKUD_GL<0
replace perindkialt=perindkialt-FORMUEINDK_NY if FORMUEINDK_NY<0
replace perindkialt=perindkialt-loenmv if loenmv<0
replace PERINDKIALTy=PERINDKIALTy-NETOVSKUD_GLy if NETOVSKUD_GLy<0
replace PERINDKIALTy=PERINDKIALTy-FORMUEINDK_NYy if FORMUEINDK_NYy<0
replace PERINDKIALTy=PERINDKIALTy-loenmvy if loenmvy<0

/* Difference between seperate components of income and aggregate income (in a few cases they don't match perfectly)
*/
cap drop restinc restincy
gen restinc=ERHVERVSINDK_GL+overforsindk-qpensny-cashass-andoverforsel +FORMUEINDK_NY
gen restincy=ERHVERVSINDK_GLy+OVERFORSINDKy-qpensnyy-cashassy-ANDOVERFORSELy +FORMUEINDK_NYy

/* Create income without universal transfers */
cap drop inc
gen inc=perindkialt-qpensny-cashass-andoverforsel
replace inc=restinc if inc==RESUINK_GL+restinc &  inc<0
replace inc=inc+qpensny+cashass+andoverforsel-overforsindk if inc<0 & qpensny+cashass+andoverforsel>overforsindk &
!mi(qpensny)

* Spouse's income *
cap drop incy
gen incy=PERINDKIALTy-qpensnyy-cashassy-ANDOVERFORSELy
replace incy=restincy if incy==RESUINK_GLy+restincy &  incy<0
replace incy=incy+qpensnyy+cashassy+ANDOVERFORSELy-OVERFORSINDKy if incy<0 &
qpensnyy+cashassy+ANDOVERFORSELy>OVERFORSINDKy & !mi(qpensnyy)

cap drop altinc altincy
gen altinc=perindkialt
gen altincy=PERINDKIALTy

/* Calculate household income */
egen hhinc=rowmean(inc incy)
egen hhearn=rowmean(loenmv_ori loenmvy_ori)
egen hhearn2=rowmean(ERHVERVSINDK_GL_ori ERHVERVSINDK_GLy_ori)
egen hhincalt=rowmean(altinc altincy)
```

```
egen hhdispinc=rowmean(dispinc dispincy)
egen hhwealth=rowmean(assets assetsy)
egen sumhhinc=rowtotal(inc incy)

cap drop gender
gen gender="men" if koen==1
replace gender="women" if koen==2

/* Death */
cap drop Dall Dcancer Dheart Dother
gen Dall=(!mi(d_dodsdto) & !mi(C_LISTE14))
gen Dcancer=(C_LISTE14=="03" &  !mi(C_LISTE14) & !mi(d_dodsdto))
gen Dheart=(C_LISTE14=="05" &  !mi(C_LISTE14) & !mi(d_dodsdto))
gen Dother=(C_LISTE14!="03" & C_LISTE14!="05" &  !mi(C_LISTE14) & !mi(d_dodsdto))
gen dyear=year(d_dodsdto)

/* Drop observations after death */
sort pnr year
by pnr: egen mdyear=max(dyear)
replace dyear=mdyear
drop mdyear
drop if year>dyear

compress

keep if inrange(year,1980,2013)

/* Drop nonessential variables */
drop pstilly IE_TYPE isced hfaudd assets wealth diag  pnry koen  assetsy wealthy Offtransy cashassy  ///
andoverforsel brutto eftloen ERHVERVSINDK_GL ERHVERVSINDK_GL_ori FORMUEINDK_NY kontskfri loenmv loenmv_ori
NETOVSKUD_GL ///
overforsindk overg OVSKEJD02_NY perindkialt qandpens RESUINK_GL SKATMVIALT_NY tilbtot varmehjalp cashass ///
Offtrans ANDOVERFORSELy bruttoy eftloeny ERHVERVSINDK_GLy ERHVERVSINDK_GLy_ori FORMUEINDK_NYy KONTSKFRIy loenmvy
loenmvy_ori  ///
NETOVSKUD_GLy OVERFORSINDKy overgy OVSKEJD02_NYy PERINDKIALTy qandpensy RESUINK_GLy SKATMVIALT_NYy TILBTOTy
VARMEHJALPy

/* Save Data */
save SocialMobilityData, replace
save F:\StataWork\Benjamin\SocialMobilityData , replace

/************************************************************/
/***            2.  Main figures and tables             ***/
/************************************************************/

/////////////////////////////////////
/*** Table 1: Mover stayer table ***/
/////////////////////////////////////
clear all
cd "F:\StataWork\Benjamin"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1983 and 1993 */
gen birthyear=year-age

keep if inrange(birthyear,1943,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-3
keep if inrange(year,minyear,maxyear)
keep if inrange(age,38,40) | inrange(age,48,50) | inrange(age,58,60)

/* Sample Restrictions */
sort pnr year
by pnr: gen laginc=hhinc[_n-2] if pnr==pnr[_n-2] & year==year[_n-2]+2
drop if mi(laginc) | mi(birthyear)

/* Sample only those alive at age 40 and 50 */
by pnr: egen sumage=sum(age)
drop if (sumage<150 & (mi(dyear) | dyear>=birthyear+60)) | (sumage<90)

gen alive=(year<dyear | mi(dyear))
by pnr: egen sumalive=sum(alive)
keep if sumalive>=2 & !mi(sumalive)
drop sumage alive sumalive
gen deadage=dyear-birthyear
gen dead=(inrange(deadage,51,60))

/* Create ranks */
set seed 112
cap drop random
qui gen double random=runiform()
replace random=(random-0.5)/_N

* Current income *
cap drop nonmiss
```

```
gen nonmiss=(!mi(laginc) & !mi(year)  & !mi(birthyear))

sort birthyear year nonmiss laginc random
cap drop incrank
by birthyear year nonmiss: gen double incrank=_n
by birthyear year nonmiss: gen double maxincrank=_N
replace incrank=ceil((((incrank-0.5)/maxincrank)*100 +random)/5)
replace incrank=. if nonmiss==0

drop nonmiss maxincrank
drop if mi(incrank)

/* Split into movers and stayers */
sort pnr year
by pnr: gen Stayer=(incrank==incrank[_n+1])
keep if age==40

/* Create Table */
mat X=J(2,4,.)

/* Bottom 5 % */
qui count if incrank==1
local tot=r(N)

* Movers *
qui sum dead if incrank==1 & Stayer==0
mat X[2,1]=r(mean)*100
mat X[1,1]=r(N)/`tot'*100

* Stayers *
qui sum dead if incrank==1 & Stayer==1
mat X[2,2]=r(mean)*100
mat X[1,2]=r(N)/`tot'*100

/* Top 5 % */
qui count if incrank==20
local tot=r(N)

* Movers *
qui sum dead if incrank==20 & Stayer==0
mat X[2,3]=r(mean)*100
mat X[1,3]=r(N)/`tot'*100

* Stayers *
qui sum dead if incrank==20 & Stayer==1
mat X[2,4]=r(mean)*100
mat X[1,4]=r(N)/`tot'*100

frmttable using  H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Table1_MoverStayer.tex, replace  ///
ctitle("Income class:", "Bottom 5%", " ", "Top 5 %", " " \ "Social mobility:", "Movers", "Stayers", "Movers",
"Stayers"  )   ///
rtitle("Share of income class (%)" \ "10-year mortality (%)") ///
sdec( 0, 0, 0, 0 \1, 1, 1, 1 )   nocenter  coljust(l c) statmat(X) ///
vlstyle(a) colwidth(16 10 10 10 10) multicol(1,2,2; 1,4,2;) hlines(1 0 1 0  1)

/////////////////////////////////////
/***** Figure 1: Main estimates ****/
/////////////////////////////////////

/// Estimate income gradient 1983-2013 ///
foreach g in women men {

clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

LifeExp hhinc Dall if gender=="`g'", idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(100) method(MS
NIM) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_1_`g'")

}

/* Bootstrap standard errors */

foreach g in women men {

    forvalues r=1/100 {

    clear
    tmpdir
    local dir=r(tmpdir)
    cd "`dir'"
    use SocialMobilityData
    keep if gender=="`g'"
```

```
    set seed `r'
    bsample , cluster(pnr)

    mata mata clear
    run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

    LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(100) method(MS NIM)
    save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Bootstrap\Main_1_`g'_b`r'")

    }
}

/// Estimate income gradient seperately for each year during 1983-2013 ///
foreach g in women men {

clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

LifeExp hhinc Dall if gender=="`g'", idvar(pnr) agevar(age) yearvar(year) byvar(year) quantiles(100) from(40)
to(100) method(MS NIM) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_1_ByYear_`g'")

}


/// Create figures ///

/* Cross-section Gradient */

* Panel A - Men *

* Calculate life expectancy *
forvalues r=1/100 {

clear
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Bootstrap"
use Main_1_men_b`r'_DeathRates

* Construct survival curves by method and income class *
sort  method incg  age
gen surv=1 if age==40
by  method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep method incg lifeexp
duplicates drop

save "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\BootstrapLE\Main_1_men_b`r'_LE", replace
}

* Append and find standard deviation *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\BootstrapLE"
use Main_1_men_b1_LE

gen it=1
forvalues r=2/100 {
cap {
append using Main_1_men_b`r'_LE
replace it=`r' if mi(it)
}
}

sort it method incg
by it method: gen LE20=lifeexp[20]
by it method: gen LE80=lifeexp[80]

gen dif8020=LE80-LE20

* Standard error *
sum dif8020 if method=="NIM"
global se_NIM=r(sd)

sum dif8020 if method=="MS"
global se_MS=r(sd)


* Plot cross-section gradient *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_men_DeathRates
```

```
/* Construct survival curves by method and income class */
sort  method incg  age
gen surv=1 if age==40
by  method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep method incg lifeexp
duplicates drop

/* Estimate gradient */
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
local dif=r(mean)-`p20'
global DifNIM="`:di %04.1g `dif''"
global DifNIM_CL="`:di %04.1g `dif'-1.96*${se_NIM}'"
global DifNIM_CU="`:di %04.1g `dif'+1.96*${se_NIM}'"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
local dif=r(mean)-`p20'
global DifMS="`:di %04.1g `dif''"
global DifMS_CL="`:di %04.1g `dif'-1.96*${se_MS}'"
global DifMS_CU="`:di %04.1g `dif'+1.96*${se_MS}'"

twoway (scatter lifeexp incg if method=="NIM" , mcolor(black) msymbol(smcircle) msize(medsmall)) ///
       (scatter lifeexp incg if method=="MS" , mcolor(red) msymbol(smcircle) msize(medsmall)) ///
 , graphregion(color(white)) xlabel(0(10)100) xtitle(Income percentile at age 40) ytitle(Life expectancy at age 40)
 ///
 legend(order(1 2) label(1 "Without mobility") label(2 "With mobility") region(color(white)) rows(2) pos(11) ring(0)
 rowgap(*0.5)) ///
 yscale(titlegap(*1.5)) xscale(titlegap(*6)) ylabel(,nogrid) xsize(8.4) ysize(6) ///
 text(71.2 70 "Difference in life years, p20-p80:", bcolor(white) box)  text(70 91 "Without mobility:  ${DifNIM}
 [${DifNIM_CL}-${DifNIM_CU}]" , bcolor(white) box justification(right) placement(west)) ///
 text(69 91 "With mobility:       ${DifMS} [${DifMS_CL}-${DifMS_CU}]", bcolor(white) box justification(right)
 placement(west)) ///
 t2title("Males, 1983-2013", position(12) ring(1) size(large)) title("A", ring(1) position(11) size(vhuge)
 color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1_CrossSectionGradient_Men.gph, replace

* Panel B - Women *

* Calculate life expectancy *
forvalues r=1/100 {

cap {
clear
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Bootstrap"
use Main_1_women_b`r'_DeathRates

* Construct survival curves by method and income class *
sort  method incg  age
gen surv=1 if age==40
by  method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep method incg lifeexp
duplicates drop

save "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\BootstrapLE\Main_1_women_b`r'_LE", replace
}
}

* Append and find standard deviation *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\BootstrapLE"
use Main_1_women_b1_LE

gen it=1
forvalues r=2/100 {
cap {
append using Main_1_women_b`r'_LE
replace it=`r' if mi(it)
}
}

sort it method incg
by it method: gen LE20=lifeexp[20]
by it method: gen LE80=lifeexp[80]
gen dif8020=LE80-LE20

/* Standard error */
sum dif8020 if method=="NIM"
```

```
global se_NIM=r(sd)

sum dif8020 if method=="MS"
global se_MS=r(sd)

* Plot cross-section gradient *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_women_DeathRates

/* Construct survival curves by method and income class */
sort  method incg  age
gen surv=1 if age==40
by  method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep method incg lifeexp
duplicates drop

/* Estimate gradient */
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
local dif=r(mean)-`p20'
global DifNIM="`:di %04.1g `dif''"
global DifNIM_CL="`:di %04.1g `dif'-1.96*${se_NIM}'"
global DifNIM_CU="`:di %04.1g `dif'+1.96*${se_NIM}'"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
local dif=r(mean)-`p20'
global DifMS="`:di %04.1g `dif''"
global DifMS_CL="`:di %04.1g `dif'-1.96*${se_MS}'"
global DifMS_CU="`:di %04.1g `dif'+1.96*${se_MS}'"

twoway (scatter lifeexp incg if method=="NIM" , mcolor(black) msymbol(smcircle) msize(medsmall)) ///
       (scatter lifeexp incg if method=="MS" , mcolor(red) msymbol(smcircle) msize(medsmall)) ///
 , graphregion(color(white)) xlabel(0(10)100) xtitle(Income percentile at age 40) ytitle(Life expectancy at age 40)
 ///
 legend(order(1 2) label(1 "Without mobility") label(2 "With mobility") region(color(white)) rows(2) pos(11) ring(0)
 rowgap(*0.5)) ///
 yscale(titlegap(*1.5)) xscale(titlegap(*6)) ylabel(,nogrid) xsize(8.4) ysize(6) ///
 text(76.6 70 "Difference in life years, p20-p80:", bcolor(white) box)  text(75.5 91 "Without mobility:  ${DifNIM}
 [${DifNIM_CL}-${DifNIM_CU}]" , bcolor(white) box justification(right) placement(west)) ///
 text(74.5 91 "With mobility:       ${DifMS} [${DifMS_CL}-${DifMS_CU}]", bcolor(white) box justification(right)
 placement(west)) ///
 t2title("Females, 1983-2013", position(12) ring(1) size(large)) title("B", ring(1) position(11) size(vhuge)
 color(black))

 graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1_CrossSectionGradient_Women.gph,
 replace

/* Change over time */

/* Panel C - Men */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_men_ByYear_DeathRates

/* Construct survival curves by year, method and income class */
sort year method incg  age
gen surv=1 if age==40
by year method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by year method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep year method incg lifeexp
duplicates drop

/* Calculate difference between 20th and 80th percentile */
keep if incg==20 | incg==80
reshape wide lifeexp, i(year method) j(incg)
gen Dif_20_80=lifeexp80-lifeexp20

keep year method Dif*
duplicates drop

qui reg Dif_20_80 c.year if method=="NIM"
local bNIM=_b[c.year]
local cNIM=_b[_cons]
local rbNIM="`:di %04.2f _b[c.year]'"
local rclNIM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuNIM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

qui reg Dif_20_80 c.year if method=="MS"
local bM=_b[c.year]
```

```
local cM=_b[_cons]
local rbM="`:di %04.2f _b[c.year]'"
local rclM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

twoway (scatter Dif_20_80 year if method=="NIM", mcolor(black) msize(medsmall)) ///
       (scatter Dif_20_80 year if method=="MS", mcolor(red) msize(medsmall)) ///
       (function y=`cNIM'+`bNIM'*x , range(1983 2013) lcolor(black)) ///
       (function y=`cM'+`bM'*x , range(1983 2013) lcolor(red)), ///
graphregion(color(white)) xlabel(1985(5)2015) ylabel(0(2)8, nogrid) xtitle(Year)  ytitle("Difference LE, P20-P80") ///
legend(order(1 2 ) label(1 "Without mobility") label(2 "With mobility") rows(2) ring(0) pos(11) region(color(white))
rowgap(*0.5)) ///
yscale(titlegap(*1.5)) xscale(titlegap(*6)) xsize(8.4) ysize(6) ///
text(8.2 2012.8 "Annual change:", bcolor(white) box) ///
text(7.15 2013.8 "`rbNIM'" "[`rclNIM'-`rcuNIM']", bcolor(white) box) ///
text(4.2 2013.8 "`rbM'" "[`rclM'-`rcuM']", color(red) bcolor(white) box) ///
t2title("Males, change over time", position(12) ring(1) size(large)) title("C", ring(1) position(11) size(vhuge)
color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1_CrossSecByYearP20P80_Men.gph , replace

/* Panel D - Women */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_women_ByYear_DeathRates

/* Construct survival curves by year, method and income class */
sort year method incg  age
gen surv=1 if age==40
by year method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by year method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+39
keep year method incg lifeexp
duplicates drop

/* Calculate difference between 20th and 80th percentile */
keep if incg==20 | incg==80
reshape wide lifeexp, i(year method) j(incg)
gen Dif_20_80=lifeexp80-lifeexp20

keep year method Dif*
duplicates drop

qui reg Dif_20_80 c.year if method=="NIM"
local bNIM=_b[c.year]
local cNIM=_b[_cons]
local rbNIM="`:di %04.2f _b[c.year]'"
local rclNIM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuNIM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

qui reg Dif_20_80 c.year if method=="MS"
local bM=_b[c.year]
local cM=_b[_cons]
local rbM="`:di %04.2f _b[c.year]'"
local rclM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

twoway (scatter Dif_20_80 year if method=="NIM", mcolor(black) msize(medsmall)) ///
       (scatter Dif_20_80 year if method=="MS", mcolor(red) msize(medsmall)) ///
       (function y=`cNIM'+`bNIM'*x , range(1983 2013) lcolor(black)) ///
       (function y=`cM'+`bM'*x , range(1983 2013) lcolor(red)), ///
graphregion(color(white)) xlabel(1985(5)2015) ylabel(0(2)8, nogrid) xtitle(Year)  ytitle("Difference LE, P20-P80") ///
legend(order(1 2 ) label(1 "Without mobility") label(2 "With mobility") rows(2) ring(0) pos(11) region(color(white))
rowgap(*0.5)) ///
yscale(titlegap(*1.5)) xscale(titlegap(*6)) xsize(8.4) ysize(6) ///
text(7 2012.8 "Annual change:", bcolor(white) box) ///
text(5.7 2013.8 "`rbNIM'" "[`rclNIM'-`rcuNIM']", bcolor(white) box) ///
text(1.5 2013.8 "`rbM'" "[`rclM'-`rcuM']", bcolor(white) box color(red)) ///
t2title("Females, change over time", position(12) ring(1) size(large)) title("D", ring(1) position(11) size(vhuge)
color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1_CrossSecByYearP20P80_Women.gph ,
replace

/*** Combine Figures ***/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  PaperFigures_1_CrossSectionGradient_Men.gph PaperFigures_1_CrossSectionGradient_Women.gph ///
               PaperFigures_1_CrossSecByYearP20P80_Men.gph PaperFigures_1_CrossSecByYearP20P80_Women.gph, ///
               rows(2) graphregion(color(white) margin(l-10)) iscale(*0.65) imargin(l+20 t-5) xsize(17.8) ysize(12)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_1.eps , replace

/////////////////////////////////////////
/***** Figure 2: Validation results ****/
/////////////////////////////////////////
```

```
/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1943,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-3
keep if inrange(year,minyear,maxyear)

* Balanced panel with non-missing income from age 40 to 60 *
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

/// Estimation ///
CohortLifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2_Actual")

LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60) method(M1 M2 MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2")

/// Create figures ///

/* Survival Curves */

 * Panel A - Bottom 5*
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_2_DeathRates

* Merge with actual mortality rates *
merge 1:1 method age incg using Main_2_Actual, nogen

* Construct survival curves by method and income class *
sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)

twoway (scatter surv age if incg==1 & method=="Act", mcolor(gs7) msymbol(circle_hollow) mlwidth(*1.5)) ///
       (scatter surv age if incg==1 & method=="NIM", mcolor(black) msymbol(circle)) ///
       (scatter surv age if incg==1 & method=="MS", mcolor(red) msymbol(+) msize(medium) mlwidth(*1.3)), ///
graphregion(color(white)) xlabel(40(5)60) ylabel(0.5(0.1)1, nogrid) xtitle(Age) ytitle(Probability of survival) ///
legend(label(1 "Actual") label(2 "Without mobility") label(3 "With mobility")  rows(3) region(color(white))
rowgap(*0.5) ring(0) pos(1)) ///
t2title("Survival curve, bottom 5 %", position(12) ring(1) size(large)) title("A", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*4)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2_SurvivalCurves_Bottom5.gph ,  replace

* Panel B - Top 5 *

twoway (scatter surv age if incg==20 & method=="Act", mcolor(gs7) msymbol(circle_hollow) mlwidth(*1.5)) ///
       (scatter surv age if incg==20 & method=="NIM", mcolor(black) msymbol(circle)) ///
       (scatter surv age if incg==20 & method=="MS", mcolor(red) msymbol(+) msize(medium) mlwidth(*1.3)), ///
graphregion(color(white)) xlabel(40(5)60) ylabel(0.92(0.02)1, nogrid) xtitle(Age) ytitle(Probability of survival) ///
legend(label(1 "Actual") label(2 "Without mobility") label(3 "With mobility")  rows(3) region(color(white))
rowgap(*0.5) ring(0) pos(1)) ///
t2title("Survival curve, top 5 %", position(12) ring(1) size(large)) title("B", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*4)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2_SurvivalCurves_Top5.gph ,  replace

/* Mobility and mortality predictions */

/* Panel C - Mobility */
```

```
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_2_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using Main_2_Actual, nogen

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==60
drop age
reshape wide avincg, i(incg) j(method) string

* MS *
twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ) ///
, ///
graphregion(color(white)) ylabel(,nogrid) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility")
label(2 "Actual") rows(3) pos(11) ring(0) region(color(white)) rowgap(*0.5)) ///
ytitle(Income percentile at age 60) xtitle(Income percentile at age 40) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mobility", position(12) ring(1) size(large)) title("C", ring(1) position(11) size(vhuge) color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2_MobilityGradient.gph ,  replace

/* Panel D - Mortality */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_2_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using Main_2_Actual, nogen

* Construct survival curves by method and income class *
sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==60
gen dead=1-surv
drop age surv

reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100

twoway (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg , color(red*0.75)  barwidth(5)) ///
, graphregion(color(white)) ylabel(-25(25)75, nogrid) xtitle(Income percentile at age 40) ytitle("Prediction error
(%), mortality from age 40 to 60") ///
legend( label(1 "Without mobility") label(2 "With mobility")  rows(2) region(color(white)) rowgap(*0.5) ring(0)
pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mortality", position(12) ring(1) size(large)) title("D", ring(1) position(11) size(vhuge) color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2_MortalityError.gph ,  replace

/*** Combine Figures ****/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  PaperFigures_2_SurvivalCurves_Bottom5.gph PaperFigures_2_SurvivalCurves_Top5.gph ///
               PaperFigures_2_MobilityGradient.gph PaperFigures_2_MortalityError.gph, ///
               rows(2) graphregion(color(white) margin(l-10)) iscale(*0.65) imargin(l+20 t-5) xsize(17.8) ysize(12)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PaperFigures_2.eps , replace


///////////////////////////////////////////////////////////////
/*** Table 3: Cross-section estimates from different ages ***/
///////////////////////////////////////////////////////////////

clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
```

```
keep if gender=="men"

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

/// Estimation ///
foreach a in 40 50 60 70 80 {
LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(`a') to(100) preserve method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_3_CrossStartAge_`a'")
}

/// Create figures ///

/* Calculate life expectancy for each start age */
foreach a in 40 50 60 70 80 {

quietly {
clear
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_3_CrossStartAge_`a'_DeathRates

/* Construct survival curves by method and income class */
sort  method incg  age
gen surv=1 if age==`a'
by  method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>`a'
by  method incg: egen lifeexp=sum(surv)

replace lifeexp=lifeexp+`a'-1
keep method incg lifeexp
duplicates drop

keep if incg==20 | incg==80

reshape wide lifeexp , i(method) j(incg)

gen difle=lifeexp80-lifeexp20
gen startage=`a'

keep method startage difle
duplicates drop
save Main_3_CrossStartAge_`a'_Lifeexp, replace
}
}

clear
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_3_CrossStartAge_40_Lifeexp

foreach a in 50 60 70 80 {

append using Main_3_CrossStartAge_`a'_Lifeexp
}

/* Create table */
mat X=J(2,5,.)
local j=1
foreach a in 40 50 60 70 80 {
    local i=1
    foreach m in NIM MS {
    qui sum difle if method=="`m'" & startage==`a'
    mat X[`i',`j']=r(mean)
    local i=`i'+1
    }
local j=`j'+1
}

frmttable using  H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_3_CrossStartAge.tex, replace  ///
ctitle("","Difference in life expectancy p20-p80 at age:","","","",""\"", "40","50","60","70","80"  )   ///
rtitle("Without income mobility" \ "With income mobility") ///
sdec(1)   nocenter  coljust(l c) statmat(X) ///
vlstyle(a) colwidth(16 10 10 10 10) multicol(1,2,5;) hlines(1 0 1 0  1)

frmttable using  H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_3_CrossStartAge.rtf, replace  ///
ctitle("","Difference in life expectancy p20-p80 at age:","","","",""\"", "40","50","60","70","80"  )   ///
rtitle("Without income mobility" \ "With income mobility") ///
sdec( 1)   nocenter  coljust(l c) statmat(X) ///
vlstyle(a) colwidth(16 10 10 10 10) multicol(1,2,5;) hlines(1 0 1 0  1)

/***********************************************************/
/***            3.  SI figures and tables               ***/
/***********************************************************/

/////////////////////////////////////////
/*** SI 1 - Other models of mobility ***/
/////////////////////////////////////////

/// Estimation ///
```

```
* Estimated in validation exercise (Figure 2) *

/// Create Figures ///

* Panel A - Mobility rates *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_2_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using Main_2_Actual , keepusing(avincg) nogen

reshape wide avincg , i(incg age) j(method) string
gen avincgNIM=incg
foreach m in M1 M2 MS NIM {
gen se_`m'=(avincg`m'-avincgAct)^2
bysort age: egen Sse_`m'=sum(se_`m')
}

keep age Sse_M1 Sse_M2 Sse_MS  Sse_NIM
duplicates drop

foreach m in M1 M2 MS NIM {
gen rse_`m'=sqrt(Sse_`m'/100)
qui sum rse_`m' if age==60
global m60`m'="`:di %3.00f r(mean)'"
global m60n`m'=r(mean)
}

twoway ///
(scatter rse_M1 age, mcolor(cyan*1.5)) ///
(scatter rse_M2 age, mcolor(green)) ///
(scatter rse_MS age, mcolor(red)) ///
(scatter rse_NIM age, mcolor(black)) ///
, graphregion(color(white)) ylabel(0(2)14, nogrid) xlabel(40(5)61) ytitle("RMSE - Average income percentile")
xtitle(Age) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
legend(order(4 1 2 3 ) label(4 "No income mobility") label(1 "First-order Markov")  label(2 "Second-order Markov")
label(3 "Mover-stayer") rows(4) region(color(white)) ring(0) pos(11)) ///
text(${m60nM1} 60.8 "${m60M1}", bcolor(white) color(cyan*1.5) size(medium)) text(${m60nM2} 60.8 "${m60M2}",
bcolor(white) color(green) size(medium)) ///
text(${m60nMS} 60.8 "${m60MS}", bcolor(white) color(red) size(medium)) text(${m60nNIM} 60.8 "${m60NIM}",
bcolor(white) color(black) size(medium)) ///
xsize(8)  ysize(6) t2title("Mobility fit", position(12) ring(1) size(large)) title("A", ring(1) position(11)
size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_1_TransErrorByAge.gph,  replace

* Panel B - Mortality rates *
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_2_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using Main_2_Actual, keepusing(deathrate) nogen

* Construct survival curves *
sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

gen mort=1-surv
keep mort incg age method
reshape wide mort , i(incg age) j(method) string

foreach m in M1 M2 MS NIM {
gen se_`m'=(mort`m'-mortAct)^2
bysort age: egen Sse_`m'=sum(se_`m')
}

keep age Sse_M1 Sse_M2 Sse_MS Sse_NIM
duplicates drop

foreach m in M1 M2 MS NIM {
gen rse_`m'=sqrt(Sse_`m'/100)
qui sum rse_`m' if age==60
global m60`m'="`:di %04.2f r(mean)'"
global m60n`m'=r(mean)
}

twoway ///
(scatter rse_M1 age, mcolor(cyan*1.5)) ///
(scatter rse_M2 age, mcolor(green)) ///
(scatter rse_MS age, mcolor(red)) ///
(scatter rse_NIM age, mcolor(black)) ///
, graphregion(color(white)) ytitle("RMSE - Mortality from age 40") ylabel(0(0.01)0.04, nogrid) xlabel(40(5)61)
xtitle(Age) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
legend(order(4 1 2 3 ) label(4 "No income mobility") label(1 "First-order Markov")  label(2 "Second-order Markov")
label(3 "Mover-stayer") rows(4) region(color(white)) ring(0) pos(11)) ///
```

```
text(${m60nM1} 61.1 "${m60M1}", bcolor(white) color(cyan*1.5) size(medium)) text(${m60nM2} 61.1 "${m60M2}",
bcolor(white) color(green) size(medium)) ///
text(${m60nMS} 61.1 "${m60MS}", bcolor(white) color(red) size(medium)) text(${m60nNIM} 61.1 "${m60NIM}",
bcolor(white) color(black) size(medium)) ///
xsize(8)  ysize(6) t2title("Mortality fit", position(12) ring(1) size(large)) title("B", ring(1) position(11)
size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_1_MortalityErrorByAge.gph ,  replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_1_TransErrorByAge.gph SIFigures_1_MortalityErrorByAge.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_1.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_1.eps , replace

//////////////////////////////////////////
/*** SI 2 - Income over several years ***/
//////////////////////////////////////////

/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1986 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1946,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-6
keep if inrange(year,minyear,maxyear)

* Balanced panel with non-missing income from age 40 to 60 *
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))

sort pnr year
cap drop inc
by pnr: gen inc=hhinc[_n-2] if pnr==pnr[_n-2] & year==year[_n-2]+2

keep if inrange(age,36,60)

/* Estimate mover-stayer model */
mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
LifeExp hhinc Dall if inrange(age,40,60), idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60)
method(MS) preserve save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\PanelEst1986sample")

/* Define Mata programs */

* Ranking program *
cap mata: mata drop RankNIM()
mata
function RankNIM(XUR,d2m,W) {
X=XUR
Mis=W:*sum(X[1..d2m,])-rowsum(X[1..d2m,])
rowW=editmissing(X[1..d2m,]:/rowsum(X[1..d2m,]),0)

     for (o=1; o<d2m; o++) {
     no=o+1

     sumx=sum(X[no,])
     if (sumx>Mis[o,1]) {
     X[o,]=X[o,]+(rowW[no,]:*Mis[o,1]):*(Mis[o,1]>=0)+(rowW[o,]:*Mis[o,1]):*(Mis[o,1]<0)
     X[no,]=X[no,]-(rowW[no,]:*Mis[o,1]):*(Mis[o,1]>=0)-(rowW[o,]:*Mis[o,1]):*(Mis[o,1]<0)
     }
     if (sumx<=Mis[o,1]) {
     X[o,]=X[o,]+(rowW[no,]:*sumx):*(Mis[o,1]>=0)
     X[no,]=J(1,d2m+1,0)
     rest=Mis[o,1]-sumx
         for (p=o+2; p<=d2m; p++) {
         sumx=sum(X[p,])
         lrest=rest
         X[o,]=X[o,]+(rowW[p,]:*rest):*(rest<sumx)+(rowW[p,]:*sumx):*(rest>=sumx)
         X[p,]=(X[p,]-rowW[p,]:*rest):*(rest<sumx)+J(1,d2m+1,0):*(rest>=sumx)
```

```
         rest=0*(lrest<sumx)+(lrest-sumx)*(lrest>=sumx)
         if (rest==0) break;
         }
     }
     Mis=W:*sum(X[1..d2m,])-rowsum(X[1..d2m,])
     rowW=editmissing(X[1..d2m,]:/rowsum(X[1..d2m,]),0)
     }
return(X)
}
end
* Estimation - No income mobility *
cap mata: mata drop NIM()
mata
function NIM(a,d2,DRNIM) {
d2m=d2-1
KN=I(d2):/d2m
KN[d2,d2]=0
LKN=KN
Dat=st_data(.,.)
    for (i=1; i<=a; i++) {
    l=(i-1)*d2m+1
    u=i*d2m
    KN=RankNIM(KN,d2m,J(d2m,1,1/d2m))
    T=I(d2)-diag(Dat[l..u,2]\0)
    T[d2,]=(Dat[l..u,2]',1)
    LKN=KN
    KN=T*KN
    DRNIM[i,]=(colsum(LKN[1..d2m,1..d2m])-colsum(KN[1..d2m,1..d2m])):/colsum(LKN[1..d2m,1..d2m])
    }
}
end

cap postclose res
postfile res yeargroup str10 method incg age deathrate avincg using
"H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_2_DifferentYearIntervals", replace

forvalues yg=1/5 {
preserve
qui expand 2 if Dall==1 , gen(LDall)
qui replace year=year+1 if LDall==1
qui replace age=age+1   if LDall==1

keep if inrange(age,41-`yg',60)
gen newage=floor((age-(41-`yg'))/`yg')*`yg'+40

collapse (mean) inc   (min) birthyear (max) LDall, by(pnr newage) fast

qui sum newage
local minage=r(min)
rename newage age

/* Create ranks */
set seed 112
qui gen double random=runiform()
replace random=(random-0.5)/_N

sort pnr age
by pnr: gen laginc=inc[_n-1] if pnr==pnr[_n-1]
by pnr: gen lrandom=random[_n-1] if pnr==pnr[_n-1]
replace random=lrandom
drop if mi(laginc)
drop lrandom

* Lagged income *
cap drop nonmiss
gen nonmiss=(!mi(laginc) & !mi(age)  & !mi(birthyear))

sort birthyear age nonmiss laginc random
cap drop lagincrank
by birthyear age nonmiss: gen double lagincrank=_n
by birthyear age nonmiss: gen double maxincrank=_N
replace lagincrank=ceil((((lagincrank-0.5)/maxincrank)*100 +random))
replace lagincrank=. if nonmiss==0
drop nonmiss maxincrank

* Current income *
cap drop nonmiss
gen nonmiss=(LDall==0 & !mi(inc) & !mi(age)  & !mi(birthyear))

sort birthyear age nonmiss inc random
cap drop incrank
by birthyear age nonmiss: gen double incrank=_n
by birthyear age nonmiss: gen double maxincrank=_N
replace incrank=ceil((((incrank-0.5)/maxincrank)*100 +random))
replace incrank=. if nonmiss==0
drop nonmiss maxincrank

sort pnr age
```

```
by pnr: gen firstincrank=lagincrank[1]

qui levelsof age, local(ages)
qui tab age
global anum=r(r)

/* Actual */
forvalues p=1/100 {
  foreach a of local ages {
    qui sum LDall if firstincrank==`p' & age==`a'
    local d=r(mean)
    qui sum incrank if LDall==0 & firstincrank==`p' & age==`a'

    post res  (`yg') ("Act") (`p') (`a') (`d') (r(mean))
    }
}

/* No income Mobility */
collapse (mean) LDall , by(age lagincrank) fast
sort age lagincrank
drop age
order lagincrank LDall

mata DRNIM=J(${anum},100,.)
mata NIM(${anum},101,DRNIM)

/* Save data */
clear
getmata (q1 q2 q3 q4 q5 q6 q7 q8 q9 q10 q11 q12 q13 q14 q15 q16 q17 q18 q19 q20 q21 q22 q23 q24 q25 q26 q27 q28 q29
q30 q31 q32 q33 q34 q35 q36 q37 q38 q39 q40 q41 q42 q43 q44 q45 q46 q47 q48 q49 q50 q51 q52 q53 q54 q55 q56 q57 q58
q59 q60 q61 q62 q63 q64 q65 q66 q67 q68 q69 q70 q71 q72 q73 q74 q75 q76 q77 q78 q79 q80 q81 q82 q83 q84 q85 q86 q87
q88 q89 q90 q91 q92 q93 q94 q95 q96 q97 q98 q99 q100)=DRNIM
gen age=_n

forvalues p=1/100 {
local i=1
  foreach a of local ages {
    qui sum q`p' if age==`i'

    post res  (`yg') ("NIM") (`p') (`a') (r(mean)) (.)
    local i=`i'+1
    }
}
restore
}
postclose res

/// Create figures ///
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use PanelEst1986sample_MobilityRates

merge 1:1 method incg age using PanelEst1986sample_DeathRates
gen yeargroup=1
append using FinalFigures_DifferentYearIntervals_Final
replace incg=inc if method!="MS"

bysort method yeargroup: egen maxage=max(age)
bysort method yeargroup: egen minage=min(age)
expand 2 if age==minage, gen(dup40)
replace age=40 if dup40==1
replace avincg=incg if dup40==1
drop if dup40==0 & age==40

* Construct survival curves *
sort method yeargroup incg age
cap drop surv
by method yeargroup incg: gen surv=1 if dup40==1
by method yeargroup incg: replace surv=surv[_n-1]-surv[_n-1]*deathrate if !mi(deathrate) & dup40==0 & method!="MS"
by method yeargroup incg: replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & dup40==0 &
method=="MS"

* Extrapolate age 60 value from last change *
sort method yeargroup incg age
by method yeargroup incg: gen Nobs=_N
by method yeargroup incg: gen obs=_n
expand 2 if age==maxage, gen(dupli)
replace age=60 if dupli==1

sort method yeargroup incg age
by method yeargroup incg: gen
impavincg=avincg[Nobs]+((avincg[Nobs]-avincg[Nobs-1])/(age[Nobs]-age[Nobs-1]))*(60-age[Nobs])
replace avincg=impavincg if age==60 & dupli==1
by method yeargroup incg: gen impsurv=surv[Nobs]+((surv[Nobs]-surv[Nobs-1])/(age[Nobs]-age[Nobs-1]))*(60-age[Nobs])
replace surv=impsurv if age==60 & dupli==1
gen mort=1-surv

keep if age==60 & dupli==1
```

```
keep method yeargroup incg avincg mort

forvalues i=1/5 {
    foreach m in NIM Act {
    replace method="`m'`i'" if yeargroup==`i' & method=="`m'"
    }
}
drop yeargroup

replace incg=ceil(incg/5)
collapse (mean) avincg mort, by(method incg)
replace incg=incg*5-2

reshape wide avincg mort, i(incg) j(method) string

* Panel A - Mobility *
twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct1 incg, color(navy))  ///
       (connect avincgAct2 incg, color(maroon)) ///
       (connect avincgAct3 incg, color(forest_green) ) ///
       (connect avincgAct4 incg, color(dkorange) ) ///
       (connect avincgAct5 incg, color(teal) ), ///
graphregion(color(white)) ytitle(income percentile at age 60) xtitle(Income percentile at age 40) ///
legend(order(2 3 4 5 6) label(2 "1") label(3 "2") label(4 "3") label(5 "4") label(6 "5") title("Years in income
avg." ,size(medsmall) color(black)) bmargin(vsmall) rowgap(*0.8) region(color(white)) colfirst rows(6) ring(0)
pos(11)) ///
yscale(titlegap(*2)) xscale(titlegap(*6)) ylabel(,nogrid) xsize(8) ysize(6)  ///
t2title("Mobility fit", position(12) ring(1) size(large)) title("A", ring(1) position(11) size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_2_DifferentYearIntervals_Mobility.gph ,
replace

* Panel B - Mortality *
gen percerrorMS=((mortMS-mortAct1)/mortAct1)*100
forvalues i=1/5 {
gen percerrorNIM`i'=((mortNIM`i'-mortAct`i')/mortAct`i')*100
}

twoway ///
(bar percerrorNIM1 incg, barwidth(5)) ///
(bar percerrorNIM2 incg, barwidth(5)) ///
(bar percerrorNIM3 incg, barwidth(5)) ///
(bar percerrorNIM4 incg, barwidth(5)) ///
(bar percerrorNIM5 incg, barwidth(5)) ///
(bar percerrorMS   incg, barwidth(5) color(red*0.85)), ///
graphregion(color(white)) xtitle(Income percentile at age 40) ylabel(-25(25)75, nogrid)  ytitle("Percentage error,
mortality from age 40 to 60") ///
yscale(titlegap(*3)) xscale(titlegap(*6)) ///
legend(label(1 "1") label(2 "2") label(3 "3") label(4 "4") label(5 "5") label(6 "Mover-stayer") title("Years in
income avg." ,size(medsmall) color(black)) region(color(white)) rows(6) ring(0) pos(1))  ///
 xsize(8)  ysize(6) t2title("Mortality fit", position(12) ring(1) size(large)) title("B", ring(1) position(11)
 size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_2_DifferentYearIntervals_Mortality.gph ,
replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_2_DifferentYearIntervals_Mobility.gph SIFigures_2_DifferentYearIntervals_Mortality.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_2.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_2.eps , replace

/////////////////////////////////////////////
/*** SI 3 - Life expectancy by education ***/
/////////////////////////////////////////////

/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

* Restrict to men and years 2000(2003)-2013 *
keep if gender=="men"
keep if inrange(year,2000,2013)

/* Estimate population-wide death rates from age 83 */
preserve
keep if inrange(year,2003,2013) & inrange(age,83,100)

collapse (mean) Dall, by(age) fast

save "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_3_LEByEduc_FullPopDeathRates", replace
```

```
restore

keep if inrange(age,37,82)

* Create education groups *
sort pnr year
by pnr: gen leduc=educ[_n-1] if pnr==pnr[_n-1] & year==year[_n-1]+1
xtile educg=leduc , nquantiles(3)

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

/* Overall */
LifeExp hhinc Dall if !mi(leduc), idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(82) method(MS NIM)
preserve save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_3_LEByEduc_Overall")

/* By Education */
LifeExp hhinc Dall if !mi(leduc), idvar(pnr) agevar(age) yearvar(year) byvar(educg) quantiles(100) from(40) to(82)
method(MS NIM) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_3_LEByEduc")

/// Create figures ///

/* Overall */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_3_LEByEduc_Overall_DeathRates

expand 2, gen(dupli)
replace deathrate=. if dupli!=0
replace age=age+43 if dupli==1
drop if age>100
drop dupli

* Merge with population deathrate *
merge m:1 age using SI_3_LEByEduc_FullPopDeathRates
replace deathrate=Dall if age>=83

* Calculate life expectancy
sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39

keep method incg lifeexp
duplicates drop

* Estimate gradient *
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
global DifNIM="`:di %04.1g r(mean)-`p20''"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
global DifMS="`:di %04.1g r(mean)-`p20''"

/* By education */
clear
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_3_LEByEduc_DeathRates

expand 2, gen(dupli)
replace deathrate=. if dupli!=0
replace age=age+43 if dupli==1
drop if age>100
drop dupli

* Merge with overall deathrate *
merge m:1 age using SI_3_LEByEduc_FullPopDeathRates
replace deathrate=Dall if age>=83

* Calculate life expectancy *
sort educg method incg age
gen surv=1 if age==40
by educg method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by educg method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39

keep educg method incg lifeexp
duplicates drop

* Estimate gradient *
forvalues e=1/3 {
qui sum lifeexp if incg==20 & method=="NIM" & educg==`e'
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM" & educg==`e'
```

```
global DifNIM`e'="`:di %04.1g r(mean)-`p20''"
}

forvalues e=1/3 {
qui sum lifeexp if incg==20 & method=="MS" & educg==`e'
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS" & educg==`e'
global DifMS`e'="`:di %04.1g r(mean)-`p20''"
}

* Panel A - Without mobility *
twoway (scatter lifeexp incg if method=="NIM" & educg==1, msymbol(smcircle) msize(medium)) ///
       (scatter lifeexp incg if method=="NIM" & educg==2, msymbol(smcircle) msize(medium)) ///
       (scatter lifeexp incg if method=="NIM" & educg==3, msymbol(smcircle) msize(medium)) ///
 , graphregion(color(white)) xlabel(0(10)100) ylabel(65(5)85, nogrid) xtitle(Income percentile at age 40)
 ytitle(Life expectancy at age 40)  ///
 legend(order(3 2 1) label(1 "0-12") label(2 "13-14") label(3 "15-21")  rows(3)  region(color(none)) ring(0) pos(11)
 title("Years of education:",size(medsmall) color(black))) ///
  text(73.2 85 "Difference in life years, p20-p80:")  text(68.5 100 "0-12:   ${DifNIM1}" , justification(right)
  placement(west)) ///
  text(70 100 "13-14:   ${DifNIM2}", justification(right) placement(west)) text(71.5 100 "15-21:   ${DifNIM3}",
  justification(right) placement(west)) ///
    text(67 100 "Overall:   ${DifNIM}", justification(right) placement(west)) ///
 yscale(titlegap(*3.5)) xscale(titlegap(*6)) t2title("Without mobility", position(12) ring(1) size(large))
 title("A", ring(1) position(11) size(vhuge) color(black)) ///
xsize(8) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_3_CrossEduc_NoIncMobility.gph , replace

* Panel B - With mobility *
twoway (scatter lifeexp incg if method=="MS" & educg==1, msymbol(smcircle) msize(medium)) ///
       (scatter lifeexp incg if method=="MS" & educg==2, msymbol(smcircle) msize(medium)) ///
       (scatter lifeexp incg if method=="MS" & educg==3, msymbol(smcircle) msize(medium)) ///
 , graphregion(color(white)) xlabel(0(10)100) ylabel(65(5)85, nogrid) xtitle(Income percentile at age 40)
 ytitle(Life expectancy at age 40)  ///
 legend(order(3 2 1) symxsize(large) symysize(large) label(1 "0-12") label(2 "13-14") label(3 "15-21") rows(3)
 region(color(none)) ring(0)  pos(11) title("Years of education:",size(medsmall) color(black))) ///
   text(73.2 85 "Difference in life years, p20-p80:")  text(68.5 100 "0-12:   ${DifMS1}" , justification(right)
   placement(west)) ///
  text(70 100 "13-14:   ${DifMS2}", justification(right) placement(west)) text(71.5 100 "15-21:   ${DifMS3}",
  justification(1right) placement(west)) ///
  text(67 100 "Overall:   ${DifMS}", justification(right) placement(west)) ///
 yscale(titlegap(*3.5)) xscale(titlegap(*6)) t2title("With mobility", position(12) ring(1) size(large)) title("B",
 ring(1) position(11) size(vhuge) color(black)) ///
xsize(8) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_3_CrossEduc_MS.gph , replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_3_CrossEduc_NoIncMobility.gph SIFigures_3_CrossEduc_MS.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_3.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_3.eps , replace

///////////////////////////////////////////////////////
/*** SI 4 - Sensitivity to number of income groups ***/
///////////////////////////////////////////////////////

/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1943,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-3
keep if inrange(year,minyear,maxyear)

/* Sample Restrictions */

* Balanced panel with non-missing income from age 40 to 60 *
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
```

```
!mi(dyear))))

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

foreach s in 5 10 20 100 {
CohortLifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(`s') from(40) to(60)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_4_SensIncomeGroupSize_Actual_`s'")

LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(`s') from(40) to(60) preserve method(MS)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_4_SensIncomeGroupSize_`s'")
}

/// Create figures ///
foreach g in 5 10 20 100 {
if `g'==5 {
local captit "A"
}
if `g'==10 {
local captit "B"
}
if `g'==20 {
local captit "C"
}
if `g'==100 {
local captit "D"
}

clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_4_SensIncomeGroupSize_`g'_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using SI_4_SensIncomeGroupSize_Actual_`g', nogen

replace incg=ceil(incg/(`g'/5))
collapse (mean) avincg , by(method age incg)
replace avincg=(avincg-0.5)/(`g'/5)+0.5

keep if age==60

twoway (function y=x, range(1 5) lcolor(black) )  ///
       (connect avincg incg if method=="Act", color(gs7) lwidth(*1.3))  ///
       (connect avincg incg if method=="MS", color(red) lwidth(*1.3)) ///
, ///
graphregion(color(white)) ylabel(,nogrid) legend(order(1 2 3) label(3 "Mover-stayer") label(1 "No income mobility")
label(2 "Actual") symxsize(medium) rows(1) pos(6) size(vsmall) region(color(white))) ///
ytitle(Income quintile at age 60) xtitle(Income quintile at age 40) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Based on `g' income classes", position(12) ring(1) size(large)) title("`captit'", ring(1) position(11)
size(vhuge) color(black)) ///
xsize(8.4) ysize(6)  nodraw

graph save fig`g'.gph, replace
}
/** Combine figures **/
run "H:\workdata\704838\Benjamin\Stataprogrammer\grc1legben.do"
bgrc1leg  fig5.gph fig10.gph fig20.gph fig100.gph ,  bxsize(17.8) bysize(12)  iscale(*0.65) ///
graphregion(color(white) margin(l-10)) imargin(l+20 t-5)

graph export SIFigures_4.png, replace width(2000)
graph export SIFigures_4.eps, replace

/////////////////////////////////
/*** SI 5 - External validity ***/
/////////////////////////////////

/// Estimation ///

/** Panel A - Predicting Mobility on US survey (PSID) data **/

/* Rename raw PSID file */
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Build\RenamePSID"

/* Create ID variables */
gen id = 1000*ER30001 + ER30002
gen id_pers = ER30002
gen id_fam = ER30001

/* Drop unnecessary variables */
drop ER*
drop V*
drop non_*

/* Reshape data */
reshape long age_ income_ sexOfHead_ relation_  seq_, i(id) j(year)

/* Sample selection */
```

```
* Remove non-head family members *
keep if (rel == 1 & year <= 1982) | (rel == 10 & year > 1982)

* Remove missing income observations in 1994/95 latino sample (about 4,000 obs) *
drop if inc>=9999999 & inrange(year,1994,1995)

* Remove individuals observed after or within 2 years of their death *
keep if (year <= yearOfDeath-2) | (yearOfDeath == 0)

* Remove those with sequence number > 20 *
keep if seq <= 20 | seq == .

* Create consistent age variable from the latest observation *
sort id year
by id: gen max_age = age[_N]
by id: gen max_year = year[_N]
by id: replace age = max_age - (max_year - year)
drop yearOfDeath seq max_age max_year

* Use working age sample and period with survey each year  *
keep if inrange(age,29,60)
keep if inrange(year,1960,1997)

* Non-missing lagged, current and lead income *
replace age=. if age>200
sort id year
by id: gen startage=age[1]
by id: gen laginc=inc[_n-1]
by id: gen leadinc=inc[_n+1]

drop if (mi(inc) | mi(laginc) | mi(leadinc))

keep if inrange(age,startage,startage+10+1)

* Balanced sample over 11 years *
cap drop Nobs
by id: gen Nobs=_N
by id: egen maxyear=max(year)
by id: egen minyear=min(year)
keep if Nobs>=11
keep if inrange(age,30,60)

drop Nobs maxyear minyear startage
by id: gen countyear=_n

/* Take out year and age effects */
foreach var in inc laginc leadinc {
qui  reg `var' i.age i.year
cap drop res
qui predict res, r
qui replace `var'=res
}

/* Generate ranks */
set seed 112
cap drop random
gen random=runiform()

* Current *
sort countyear inc random
by countyear: gen incrank=_n
by countyear: gen maxincrank=_N
replace incrank=ceil(((incrank-0.5)/maxincrank)*100)
drop maxincrank

* Lagged *
sort countyear laginc random
by countyear: gen lagincrank=_n
by countyear: gen maxincrank=_N
replace lagincrank=ceil(((lagincrank-0.5)/maxincrank)*100)
drop maxincrank

* Lead *
sort countyear leadinc random
by countyear: gen leadincrank=_n
by countyear: gen maxincrank=_N
replace leadincrank=ceil(((leadincrank-0.5)/maxincrank)*100)
drop maxincrank random

/* Determine share of stayers */
gen ori_incrank=incrank
gen ori_leadincrank=leadincrank
replace lagincrank=ceil(lagincrank/20)
replace incrank=ceil(incrank/20)
replace leadincrank=ceil(leadincrank/20)

gen match=(incrank==leadincrank)
```

```
gen match100=(ori_incrank==ori_leadincrank)

sum match100
local m100=r(mean)
sum match
global ratio=`m100'/r(mean)

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

gen inc=0
gen Dall=0
CohortLifeExp inc Dall, idvar(id) agevar(countyear) yearvar(countyear) quantiles(5) from(1) to(11)
inputranks(incrank) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_PSID_Actual")

LifeExp inc Dall, idvar(id) agevar(countyear) yearvar(countyear) quantiles(5) from(1) to(11) method(M1 M2 MS NIM)
inputranks(lagincrank incrank leadincrank) ratio(${ratio}) outputtrans(Qt Wt Wat)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_PSID")

/** Panel B - Rank-Rank Correlation **/
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

keep if gender=="men"

/* Use Chetty et al. definition of household income */
cap drop hhinc
egen hhinc=rowtotal(inc incy)

/* Only use observations between 1999 - 2013 */
keep if inrange(year,1999,2013)

sort pnr year
cap drop inc
by pnr: gen inc=hhinc[_n-2] if pnr==pnr[_n-2] & year==year[_n-2]+2
drop if mi(inc) | mi(year) | mi(age)

/* Income ranks */
set seed 112
qui gen double random=runiform()
qui replace random=(random-0.5)/_N

sort age year inc random
by age year: gen double incrank=_n
by age year: gen double maxincrank=_N
qui replace incrank=ceil((((incrank-0.5)/maxincrank)*100 +random))
drop maxincrank

/* Create income percentile lags */
sort pnr year
forvalues l=0/10 {
by pnr: gen lagincrank`l'=incrank[_n-`l'] if pnr==pnr[_n-`l'] & year==year[_n-`l']+`l'
}
keep if inrange(age,40,60)
keep pnr year incrank lagincrank*

cap postclose res
postfile res lag cor using H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_RankRankCorrelation , replace
forvalues l=0/10 {

qui corr lagincrank`l' incrank

post res (`l') (r(rho))
}
postclose res

/** Panel C - US income gradient with mobility **/
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

keep if gender=="men"
keep if inrange(year,1998,2013)

/* Drop individuals with disability pension or zero income */
sort pnr year
gen disabdrop=((qpensny>0 & !mi(qpensny)) & age<60 )
gen incdrop=(inc<=0 | hhinc<=0)
by pnr: gen maxdisabdrop=(disabdrop[_n-2]==1 )
by pnr: gen maxincdrop=(incdrop[_n-2]==1 )
gen drop=(maxincdrop==1 | maxdisabdrop==1)
```

```
/* Use Chetty et al. definition of household income */
cap drop hhinc
egen hhinc=rowtotal(inc incy)

* Load US Death rates *
preserve
clear
use "H:\workdata\704838\Benjamin\DeathIncomeGradient\USdata\mskd_national_mortratesBY_gnd_hhincpctile_age_year"

keep if gnd=="M" & inrange(yod,2001,2013)
collapse (mean) mortrate (rawsum) count [w=count], by(pctile age_at_d)

keep age_at_d pctile mortrate
reshape wide mortrate, i(age_a) j(pctile)
keep mortrate*

mata RD=st_data(.,.)
restore

run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(75) method(MS NIM)
inputdeath(RD) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_USGradient")

/** Panel D - DK income gradient with US restrictions **/
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

keep if gender=="men"
keep if inrange(year,1998,2013)

/* Drop individuals with disability pension or zero income */
sort pnr year
gen disabdrop=((qpensny>0 & !mi(qpensny)) & age<60 )
gen incdrop=(inc<=0 | hhinc<=0)
by pnr: gen maxdisabdrop=(disabdrop[_n-2]==1 )
by pnr: gen maxincdrop=(incdrop[_n-2]==1 )

gen drop=(maxincdrop==1 | maxdisabdrop==1)

/* Find out who to drop from sample after retirement */
set seed 112
qui gen double random=runiform()
replace random=(random-0.5)/_N

* Create ranks to identify people to drop *
sort pnr year
by pnr: gen laginc=hhinc[_n-2] if pnr==pnr[_n-2] & year==year[_n-2]+2

gen nonmiss=(!mi(laginc) & !mi(year) & !mi(age))
sort age year nonmiss laginc random
by age year nonmiss: gen double incrank=_n
by age year nonmiss: gen double maxincrank=_N
replace incrank=(((incrank-0.5)/maxincrank)*100 +random)
replace incrank=. if nonmiss==0

gen incperc=ceil(incrank)
bysort year incperc: egen sumdrop=sum(drop==1 & age==59)
bysort year incperc: egen sumage=sum(age==59)
gen percdrop=sumdrop/sumage

gen placewithinpercentile=incrank-(incperc-1)
replace drop=(placewithinpercentile<percdrop) if age>=60

drop if drop==1
drop placewithinpercentile percdrop incperc incrank sumdrop sumage drop nonmiss maxincrank

run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(100) method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_DKGradient")

/// Create figures ///

/* Panel A - Predicting mobility on US survey data */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_5_PSID_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method countyear incg using SI_5_PSID_Actual
keep if countyear==11

twoway (function y=x, range(1 5) lcolor(black) )  ///
       (connect avincg incg if method=="Act", color(gs7) lwidth(*1.3))  ///
       (connect avincg incg if method=="MS", color(red) lwidth(*1.3)), ///
graphregion(color(white)) ylabel(,nogrid) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility")
```

```
label(2 "Actual") symxsize(medium) rows(3) pos(11) ring(0) rowgap(*0.5) region(color(white))) ///
ytitle(Income quintile after ten years) xtitle(Initial income quintile) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mobility prediction on US survey data", position(12) ring(1) size(large)) title("A", ring(1) position(11)
size(vhuge) color(black)) ///
xsize(8.4) ysize(6)

graph save "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5_PSID.gph", replace

/* Panel B - Rank-Rank correlation in US/DK */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\USdata"
use RankRankUS

rename incomecorr_M cor_US

merge 1:1 lag using "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_5_RankRankCorrelation"
keep if inrange(lag,0,10)

/* Men */
twoway (connect cor lag , color(red)) ///
        (connect cor_US lag , color(navy)), ///
        graphregion(color(white)) ylabel(,nogrid) xtitle(Income lag) ytitle(Rank-rank correlation) ///
        legend(label(1 "DK: Men") label(2 "US: Men")  rows(2) pos(9) ring(0) region(color(none))) ///
        ylabel(0(0.2)1) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
        t2title("Income mobility Denmark vs. US", position(12) ring(1) size(large)) title("B", ring(1) position(11)
        size(vhuge) color(black)) ///

graph save "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5_RankRankCorrelation.gph", replace

/* Panel C -  US income gradient in life expectancy */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_5_USGradient_DeathRates

expand 2, gen(dupli)
replace deathrate=. if dupli!=0
replace age=age+36 if dupli==1
drop dupli
drop if age>100

gen logdeathrate=log(deathrate)
gen logestdeathrate=.
foreach m in MS NIM  {
    forvalues q=1/100 {
    qui reg logdeathrate c.age if incg==`q' & method=="`m'"  & age<=75

    cap drop pred
    predict pred, xb
    qui replace logestdeathrate=pred if  incg==`q' & method=="`m'"
    }
}
gen estdeathrate=exp(logestdeathrate)

* Create final deathrate *
gen finaldeathrate=deathrate if age<=75
replace finaldeathrate=estdeathrate if age>75

merge m:1 age using "H:\workdata\704838\Benjamin\DeathIncomeGradient\USdata\CDC_SSA_Mort"
replace finaldeathrate=deathrate_all if inrange(age,90,100)

sort  method incg age
gen surv=1 if age==40
by  method incg: replace surv=surv[_n-1]-surv[_n-1]*finaldeathrate[_n-1] if !mi(finaldeathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39

keep  method incg lifeexp
duplicates drop

/* Estimate gradient */
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
global DifR="`:di %04.1g r(mean)-`p20''"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
global DifM="`:di %04.1g r(mean)-`p20''"

twoway (scatter lifeexp incg if method=="NIM" , mcolor(black) msymbol(smcircle) msize(medsmall)) ///
       (scatter lifeexp incg if method=="MS" , mcolor(red) msymbol(smcircle)  msize(medsmall)) ///
 , graphregion(color(white)) xlabel(0(10)100) ylabel(70(5)85, nogrid) xtitle(Income percentile at age 40)
 ytitle(Life expectancy at age 40)  ///
 legend(order(1 2) label(1 "Without mobility")  label(2 "With mobility")  region(color(none)) rows(2) pos(11)
 ring(0) rowgap(*0.5)) ///
 yscale(titlegap(*1.5)) xscale(titlegap(*6)) ///
 text(76.2 70 "Difference in life years, p20-p80:")  text(75 91 "Without mobility:     ${DifR}" ,
```

```
 justification(right) placement(west)) text(74 91 "With mobility:           ${DifM}", justification(right)
 placement(west)) ///
 xsize(8.4) ysize(6) t2title("US life expectancy", position(12) ring(1) size(large)) title("C", ring(1) position(11)
 size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5_USCrossSection.gph ,  replace

/* Panel D -  DK income gradient in life expectancy without DI */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_5_DKGradient_DeathRates

sort  method incg age
gen surv=1 if age==40
by  method incg: replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by  method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39

keep  method incg lifeexp
duplicates drop

/* Estimate gradient */
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
global DifR="`:di %04.1g r(mean)-`p20''"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
global DifM="`:di %04.1g r(mean)-`p20''"

twoway (scatter lifeexp incg if method=="NIM" , mcolor(black) msymbol(smcircle) msize(medsmall)) ///
       (scatter lifeexp incg if method=="MS" , mcolor(red) msymbol(smcircle)  msize(medsmall)) ///
 , graphregion(color(white)) xlabel(0(10)100) ylabel(70(5)85, nogrid) xtitle(Income percentile at age 40)
 ytitle(Life expectancy at age 40)  ///
 legend(order(1 2) label(1 "Without mobility")  label(2 "With mobility")  region(color(none)) rows(2) pos(11)
 ring(0) rowgap(*0.5)) ///
 yscale(titlegap(*1.5)) xscale(titlegap(*6)) ///
 text(76.2 70 "Difference in life years, p20-p80:")  text(75 91 "Without mobility:     ${DifR}" ,
 justification(right) placement(west)) text(74 91 "With mobility:          ${DifM}", justification(right)
 placement(west)) ///
 xsize(8.4) ysize(6) t2title("DK life expectancy excl. disabled", position(12) ring(1) size(large)) title("D",
 ring(1) position(11) size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5_DKCrossSection.gph ,  replace

/*** Combine figures ***/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_5_PSID.gph SIFigures_5_RankRankCorrelation.gph ///
                SIFigures_5_USCrossSection.gph SIFigures_5_DKCrossSection.gph, ///
               rows(2) graphregion(color(white) margin(l-10)) iscale(*0.65) imargin(l+20 t-5) xsize(17.8) ysize(12)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_5.eps , replace

/////////////////////////////////////////
/*** SI 6 - Out-of-sample prediction ***/
/////////////////////////////////////////

/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1943,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-3
keep if inrange(year,minyear,maxyear)

/* Balanced panel with non-missing income from age 40 to 60 */
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))
```

```
mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

/* Split sample into two equally sized random samples */
set seed 200
gen random=runiform()
bysort pnr: egen prand=mean(rand)
replace random=(prand>0.5)

CohortLifeExp hhinc Dall if random==0, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_6_OutofSample_Actual")

LifeExp hhinc Dall  if random==1, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60) method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_6_OutofSample")

/// Create figures ///

/* Panel A - Mobility */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_6_OutofSample_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using SI_6_OutofSample_Actual, nogen

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==60
drop age
reshape wide avincg, i(incg) j(method) string

twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ), ///
graphregion(color(white)) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility") label(2 "Actual")
rows(3) pos(11) ring(0) region(color(white)) rowgap(*0.5)) ///
ytitle(Income percentile at age 60) xtitle(Income percentile at age 40) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
 xsize(8.4) ysize(6) t2title("Mobility", position(12) ring(1) size(large)) title("A", ring(1) position(11)
 size(vhuge) color(black)) ///
 xscale(titlegap(*2)) yscale(titlegap(*4)) ylabel(,nogrid)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_6_OutofSample_Mobility.gph ,  replace

/* Panel B - Mortality */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_6_OutofSample_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using SI_6_OutofSample_Actual, nogen

sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==60
gen dead=1-surv
drop surv age
reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100

twoway (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg , color(red*0.75)  barwidth(5)) ///
 , graphregion(color(white)) /*xlabel(0 5 10 15 20)*/ ylabel(-25(25)75, nogrid) xtitle(Income percentile at age 40)
 ytitle("Prediction error (%), mortality from age 40 to 60") ///
 legend( label(1 "Without mobility") label(2 "With mobility")  rows(2) region(color(white)) rowgap(*0.5) ring(0)
 pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
 xsize(8.4) ysize(6) t2title("Mortality", position(12) ring(1) size(large)) title("B", ring(1) position(11)
 size(vhuge) color(black)) ///
 xscale(titlegap(*2)) yscale(titlegap(*4))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_6_OutofSample_Mortality.gph , replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_6_OutofSample_Mobility.gph SIFigures_6_OutofSample_Mortality.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)
```

```
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_6.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_6.eps , replace

///////////////////////////////////////////////////
/*** SI 7 - Predicting mobility at other ages ***/
///////////////////////////////////////////////////

/***************************/
/**** From age 50 to 70 ****/
/***************************/

/// Estimation ///

/* Panel A and B - From age 50 to 70 */
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 50 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1933,1943)
gen maxyear=birthyear+50+20
gen minyear=birthyear+50-3
keep if inrange(year,minyear,maxyear)

* Balanced panel with non-missing income from age 50 to 70 *
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

CohortLifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(50) to(70)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_7_DifAge_50_Actual")

LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(50) to(70) method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_7_DifAge_50")

/* Panel C and D - From age 60 to 80 */
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 60 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1923,1933)
gen maxyear=birthyear+60+20
gen minyear=birthyear+60-3
keep if inrange(year,minyear,maxyear)

* Balanced panel with non-missing income from age 60 to 80 *
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>minyear)
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

CohortLifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(60) to(80)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_7_DifAge_60_Actual")

LifeExp hhinc Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(60) to(80) method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_7_DifAge_60")
```

```
/// Create figures ///

/* Panel A - Mobility from age 50*/
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_7_DifAge_50_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using SI_7_DifAge_50_Actual, nogen

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==70
drop age
reshape wide avincg, i(incg) j(method) string

* MS *
twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ) ///
, ///
graphregion(color(white)) ylabel(,nogrid) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility")
label(2 "Actual") rows(3) pos(11) ring(0) region(color(white)) rowgap(*0.5)) ///
ytitle(Income percentile at age 70) xtitle(Income percentile at age 50) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mobility from age 50 to 70", position(12) ring(1) size(large)) title("A", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7_DifAge_50_MobilityGradient.gph ,  replace

/* Panel B - Mortality from age 50 */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_7_DifAge_50_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using SI_7_DifAge_50_Actual, nogen

* Construct survival curves by method and income class *
sort method incg  age
gen surv=1 if age==50
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>50

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==70
gen dead=1-surv
drop age surv

reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100

twoway (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg , color(red*0.75)  barwidth(5)) ///
, graphregion(color(white)) ylabel(-25(25)75, nogrid) xtitle(Income percentile at age 50) ytitle("Prediction error
(%), mortality from age 50 to 70") ///
legend( label(1 "Without mobility") label(2 "With mobility")  rows(2) region(color(white)) rowgap(*0.5) ring(0)
pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mortality from age 50 to 70", position(12) ring(1) size(large)) title("B", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7_DifAge_50_MortalityError.gph ,  replace

/* Panel C - Mobility from age 60 */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_7_DifAge_60_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using SI_7_DifAge_60_Actual, nogen

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==80
drop age
reshape wide avincg, i(incg) j(method) string

* MS *
```

```
twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ) ///
, ///
graphregion(color(white)) ylabel(,nogrid) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility")
label(2 "Actual") rows(3) pos(11) ring(0) region(color(white)) rowgap(*0.5)) ///
ytitle(Income percentile at age 80) xtitle(Income percentile at age 60) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mobility from age 60 to 80", position(12) ring(1) size(large)) title("C", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7_DifAge_60_MobilityGradient.gph ,  replace

/* Panel D - Mortality from age 60 */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_7_DifAge_60_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using SI_7_DifAge_60_Actual, nogen

* Construct survival curves by method and income class *
sort method incg  age
gen surv=1 if age==60
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>60

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==80
gen dead=1-surv
drop age surv

reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100

twoway (bar MSdif incg if abs(MSdif)>abs(NIMdif), color(red*0.75)  barwidth(5)) ///
       (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg if abs(MSdif)<=abs(NIMdif), color(red*0.75)  barwidth(5)) ///
, graphregion(color(white)) ylabel(-25(25)75, nogrid) xtitle(Income percentile at age 60) ytitle("Prediction error
(%), mortality from age 60 to 80") ///
legend(order(2 3) label(2 "Without mobility") label(3 "With mobility")  rows(2) region(color(white)) rowgap(*0.5)
ring(0) pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
t2title("Mortality from age 60 to 80", position(12) ring(1) size(large)) title("D", ring(1) position(11) size(vhuge)
color(black)) ///
xscale(titlegap(*2)) yscale(titlegap(*4)) xsize(8.4) ysize(6)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7_DifAge_60_MortalityError.gph ,  replace

cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_7_DifAge_50_MobilityGradient.gph SIFigures_7_DifAge_50_MortalityError.gph ///
               SIFigures_7_DifAge_60_MobilityGradient.gph SIFigures_7_DifAge_60_MortalityError.gph, ///
               rows(2) graphregion(color(white) margin(l-10)) iscale(*0.65) imargin(l+20 t-5) xsize(17.8) ysize(12)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_7.eps , replace

//////////////////////////////////////////////////////////////////////////////
/*** SI 8 - Sensitivity to income definition and income after retirement ***/
//////////////////////////////////////////////////////////////////////////////

/// Estimation ///

/*** Panel A and B - Sensitivity to income definition ***/
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

/* Restrict to people aged 40 between 1983 and 1993 */
gen birthyear=year-age
keep if inrange(birthyear,1943,1953)
gen maxyear=birthyear+40+20
gen minyear=birthyear+40-3
keep if inrange(year,minyear,maxyear)

/* Balanced panel with non-missing income from age 40 to 60 */
by pnr: gen increst=hhearn[_n-1]
drop if (mi(increst) & year>minyear)
```

```
drop increst

sort pnr year
by pnr: gen obs=_N
drop if (obs!=dyear-minyear+1 & dyear<maxyear) | (obs!=maxyear-minyear+1 & (mi(dyear) | (dyear>=maxyear &
!mi(dyear))))

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

CohortLifeExp hhearn Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_8_SensIncomeDef_Actual")

LifeExp hhearn Dall, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60) method(MS NIM)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_8_SensIncomeDef")

/** Panel C - Correlation between pre- and post-retirement income **/
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

keep if inrange(year,1997,2013) & gender=="men" & !mi(hhinc) & Dall==0

/* Identify retirement */
sort pnr year
gen pens=(pstill=="92" | pstill=="50") if !mi(pstill)
by pnr: gen retire=(pens[_n-1]==0 & pens==1) if !mi(pens) & !mi(pens[_n-1])
by pnr: gen sumretire=sum(retire)
replace retire=0 if sumretire>1 & !mi(sumretire)

/* Restrict sample to 3 years before, 1 year before and 1 year after retirement */
by pnr: gen pre3retire=retire[_n+3]==1 if year[_n+3]==year+3
by pnr: gen preretire=retire[_n+1]==1 if year[_n+1]==year+1
by pnr: gen postretire=retire[_n-1]==1 if year[_n-1]==year-1
keep if pre3retire==1 | preretire==1 | retire==1 | postretire==1

by pnr: gen obs=_N
keep if obs==4
drop obs

gen time=0   if pre3retire==1
replace time=1 if preretire==1
replace time=2 if retire==1
replace time=3 if postretire==1

/* Create income ranks */
set seed 112
cap drop random
qui gen double random=runiform()
replace random=(random-0.5)/_N

cap drop nonmiss
gen nonmiss=(!mi(year) & !mi(age))

sort age year time nonmiss hhinc random
cap drop incrank
by age year time nonmiss: gen double incrank=_n
by age year time nonmiss: gen double maxincrank=_N
replace incrank=ceil((((incrank-0.5)/maxincrank)*100 +random))
replace incrank=. if nonmiss==0
drop nonmiss maxincrank

keep pnr time gender incrank
reshape wide incrank , i(pnr) j(time)

/* Collapse and save data */
preserve
collapse (mean) incrank3 , by(incrank1) fast

cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

save SI_8C_RetirementInc, replace
restore

preserve
collapse (mean) incrank1 , by(incrank0) fast

cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

save SI_8C_RetirementInc_Placebo, replace
restore

/** Panel D - Using Gompertz from age 65 **/
* Based on death rates in figure 1 in main text *
```

```
/// Create figures ///

/* Panel A - Sensitivity to income definition - Mobility */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_8_SensIncomeDef_MobilityRates

* Merge with actual mobility rates*
merge 1:1 method age incg using SI_8_SensIncomeDef_Actual, nogen

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==60
drop age
reshape wide avincg, i(incg) j(method) string

twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ), ///
graphregion(color(white)) legend(order(2 1 3) label(3 "With mobility") label(1 "Without mobility") label(2 "Actual")
rows(3) pos(11) ring(0) region(color(white)) rowgap(*0.5)) ///
ytitle(Earnings percentile at age 60) xtitle(Earnings percentile at age 40) yscale(titlegap(*2))
xscale(titlegap(*6)) ///
 xsize(8.4) ysize(6) t2title("Mobility", position(12) ring(1) size(large)) title("A", ring(1) position(11)
 size(vhuge) color(black)) ///
 xscale(titlegap(*2)) yscale(titlegap(*4)) ylabel(,nogrid)

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8_SensEarn_Mobility.gph ,  replace

/* Panel B - Sensitivity to income definition - Mortality */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_8_SensIncomeDef_DeathRates

* Merge with actual mortality rates*
merge 1:1 method age incg using SI_8_SensIncomeDef_Actual, nogen

sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==60
gen dead=1-surv
drop surv age
reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100

twoway (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg , color(red*0.75)  barwidth(5)) ///
 , graphregion(color(white)) /*xlabel(0 5 10 15 20)*/ ylabel(-25(25)75, nogrid) xtitle(Earnings percentile at age
 40) ytitle("Prediction error (%), mortality from age 40 to 60") ///
 legend( label(1 "Without mobility") label(2 "With mobility")  rows(2) region(color(white)) rowgap(*0.5) ring(0)
 pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
 xsize(8.4) ysize(6) t2title("Mortality", position(12) ring(1) size(large)) title("B", ring(1) position(11)
 size(vhuge) color(black)) ///
 xscale(titlegap(*2)) yscale(titlegap(*4))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8_SensEarn_Mortality.gph , replace

/* Panel C - Correlation between pre- and post-retirement income */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_8C_RetirementInc

append using SI_8C_RetirementInc_Placebo

twoway (scatter incrank1 incrank0  , mcolor(black) msize(small)) ///
        (scatter incrank3 incrank1 , mcolor(red) msize(small)) ///
        (function y=x , range(0 100) lcolor(black)), ///
        graphregion(color(white))  ylabel(,nogrid) xtitle(Income percentile before retirement) ytitle(Income
        percentile after retirement) ///
        legend(order(2 1) label(1 "Placebo") label(2 "Estimate") region(color(white)) ring(0) pos(11) rows(2))
        yscale(titlegap(*2)) xscale(titlegap(*6)) ///
        xsize(8.4) ysize(6) t2title("Income before and after retirement", position(12) ring(1) size(large))
        title("C", ring(1) position(11) size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8_StableIncRetire.gph ,  replace

/* Panel D - Using Gompertz from age 65 */
```

```
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_men_DeathRates

gen logdeathrate=log(deathrate)
gen logestdeathrate=.
foreach m in MS NIM  {
    forvalues q=1/100 {
    qui reg logdeathrate c.age if incg==`q' & method=="`m'" & age<65

    cap drop pred
    predict pred, xb
    qui replace logestdeathrate=pred if  incg==`q' & method=="`m'"
    }
}
gen estdeathrate=exp(logestdeathrate)

gen finaldeathrate=deathrate if age<65
replace finaldeathrate=estdeathrate if age>=65
forvalues age=80/100 {
    foreach m in MS NIM {
    qui sum deathrate  if method=="`m'"
    qui replace finaldeathrate=r(mean) if age==`age' & method=="`m'"
    }
}

gen surv=1 if age==40
by method incg: replace surv=surv[_n-1]-surv[_n-1]*finaldeathrate[_n-1] if !mi(finaldeathrate[_n-1]) & age>40
by method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39
keep method incg lifeexp
duplicates drop

/* Estimate gradient */
qui sum lifeexp if incg==20 & method=="NIM"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="NIM"
global DifR="`:di %04.1g r(mean)-`p20''"

qui sum lifeexp if incg==20 & method=="MS"
local p20=r(mean)
qui sum lifeexp if incg==80 & method=="MS"
global DifM="`:di %04.1g r(mean)-`p20''"

twoway (scatter lifeexp incg if method=="NIM" , mcolor(black) msymbol(smcircle) msize(medsmall)) ///
       (scatter lifeexp incg if method=="MS" , mcolor(red) msymbol(smcircle) msize(medsmall)) ///
 , graphregion(color(white)) xlabel(0(10)100) ylabel(,nogrid) xtitle(Income percentile at age 40) ytitle(Life
 expectancy at age 40)  ///
 legend(order(1 2) label(1 "Without mobility") label(2 "With mobility")  region(color(white)) rows(2) pos(11)
 ring(0) rowgap(*0.5) ) ///
 yscale(titlegap(*3.5)) xscale(titlegap(*6)) text(71.2 76 "Difference in life years, p20-p80:", bcolor(white) box) ///
 text(70 91 "Without mobility:       ${DifR}" , box justification(right) placement(west) bcolor(white)) ///
 text(69 91 "With mobility:            ${DifM}", box bcolor(white) justification(right) placement(west)) xsize(8.4)
 ysize(6) ///
 t2title("Males, 1983-2013 with Gompertz approx. from age 65", position(12) ring(1) size(large)) title("D", ring(1)
 position(11) size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8_Gompertz.gph ,  replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_8_SensEarn_Mobility.gph SIFigures_8_SensEarn_Mortality.gph ///
               SIFigures_8_StableIncRetire.gph SIFigures_8_Gompertz.gph , ///
               rows(2) graphregion(color(white) margin(l-10)) iscale(*0.55) imargin(l+20 t-5) xsize(17.8) ysize(12)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_8.eps , replace

//////////////////////////////////////////
/*** SI 9 - Mobility in cross-section ***/
//////////////////////////////////////////

/// Estimation ///
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

/* Restrict to men */
keep if gender=="men"

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\CohortLifeExp"

gen birthyear=year-age
```

```
/* Estimation using cross-section in each cohort 1983-1993 */
forvalues y=1983/1993 {
preserve
keep if inrange(year,`y'-3,`y'+20)

* Balanced panel with non-missing income from age 40 to 60 *
sort pnr year
by pnr: gen increst=hhinc[_n-1]
drop if (mi(increst) & year>`y'-3)
drop increst

by pnr: gen obs=_N
drop if (obs!=dyear-(`y'-3)+1 & dyear<`y'+20) | (obs!=24 & (mi(dyear) | (dyear>=`y'+20 & !mi(dyear))))

CohortLifeExp hhinc Dall if birthyear==`y'-40, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60)
outputdeath(PD) uncond

LifeExp hhinc Dall if year==`y', idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40) to(60) method(MS)
inputdeath(PD) save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_9_CrossSec_`y'")
restore
}

/// Create figures ///

/* Panel A - Mobility */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_9_CrossSec_1983_MobilityRates

forvalues yr=1984/1993 {
append using SI_9_CrossSec_`yr'_MobilityRates
}

collapse (mean) avincg , by(method age incg)
replace method="MS_Cross" if method=="MS"

append using "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2_Actual"
append using "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2_MobilityRates"

replace incg=ceil(incg/5)
collapse (mean) avincg , by(method age incg)
replace incg=incg*5-2

keep if age==60
drop age
reshape wide avincg, i(incg) j(method) string

twoway ///
       (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincgAct incg, color(gs7))  ///
       (connect avincgMS incg, color(red) ) ///
       (connect avincgMS_Cross incg, color(green)), ///
graphregion(color(white)) legend(order(1 2 3 4) label(3 "Mover-stayer, panel") label(4 "Mover-stayer,
cross-section") label(1 "No income mobility")  label(2 "Actual") rows(4) pos(11) ring(0) region(color(none))) ///
ytitle(Income percentile at age 60) xtitle(Income percentile at age 40) ylabel(,nogrid) yscale(titlegap(*2))
xscale(titlegap(*6)) ///
xsize(8.7) ysize(6) t2title("Mobility", position(12) ring(1) size(large)) title("A", ring(1) position(11)
size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_9_Mobility.gph ,  replace

/* Panel B - Mortality */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_9_CrossSec_1983_DeathRates

forvalues yr=1984/1993 {
append using SI_9_CrossSec_`yr'_DeathRates
}
collapse (mean) deathrate , by(method age incg)
replace method="MS_Cross" if method=="MS"

* Merge with actual mortality rates*
append using "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2_Actual"
append using "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\Main_2_DeathRates"

* Construct survival curves *
sort method incg  age
gen surv=1 if age==40
by method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40

replace incg=ceil(incg/5)
collapse (mean) surv , by(method age incg)
replace incg=incg*5-2

keep if age==60
gen dead=1-surv
```

```
drop surv age
reshape wide dead, i(incg) j(method) string

gen NIMdif=((deadNIM-deadAct)/deadAct)*100
gen MSdif=((deadMS-deadAct)/deadAct)*100
gen MS_Crossdif=((deadMS_Cross-deadAct)/deadAct)*100

twoway (bar NIMdif incg , color(black*0.75) barwidth(5)) ///
       (bar MSdif incg , color(red*0.75)  barwidth(5)) ///
       (bar MS_Crossdif incg , color(green*0.75)  barwidth(5)) ///
 , graphregion(color(white)) /*xlabel(0 5 10 15 20)*/ ylabel(-25(25)75, nogrid) xtitle(Income percentile at age 40)
 ytitle("Prediction error (%), mortality from age 40 to 60") ///
 legend( label(1 "No income mobility") label(2 "Mover-stayer, panel") label(3 "Mover-stayer, cross-section") rows(3)
 region(color(white)) rowgap(*0.5) ring(0) pos(1)) yscale(titlegap(*2)) xscale(titlegap(*6)) ///
 xsize(8.4) ysize(6) t2title("Mortality", position(12) ring(1) size(large)) title("B", ring(1) position(11)
 size(vhuge) color(black)) ///
 xscale(titlegap(*2)) yscale(titlegap(*4))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_9_Mortality.gph , replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_9_Mobility.gph SIFigures_9_Mortality.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_9.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_9.eps , replace

//////////////////////////////////////////////
/*** SI 10 - Changes in mobility over time ***/
//////////////////////////////////////////////

/// Estimation ///

/* Panel A - Holding mobility constant */
clear all
tmpdir
local dir=r(tmpdir)
cd "`dir'"
use SocialMobilityData

mata mata clear
run "H:\workdata\704838\Benjamin\DeathIncomeGradient\Code\Final\StataProgram\LifeExp"

/* Estimate mover-stayer in 1983 */
LifeExp hhinc Dall if gender=="men" & year==1983, idvar(pnr) agevar(age) yearvar(year) quantiles(100) from(40)
to(100) method(MS) preserve outputtrans(Q1983 W1983 Wa1983)

/* Estimate mover-stayer during 1983-2013 with 1983-mobility rates */
LifeExp hhinc Dall if gender=="men", idvar(pnr) agevar(age) yearvar(year) byvar(year) quantiles(100) from(40)
to(100) method(MS) inputtrans(Q1983 W1983 Wa1983)
save("H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SI_10_ConstantMobility")

/** Panel B - Mobility rates in mover-stayer in different years **/

/* Calculated in figure 1 */

/// Create figures ///

/* Panel A - Holding mobility constant */
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use SI_10_ConstantMobility_DeathRates

replace method="MS_1983" if method=="MS"

append using Main_1_men_ByYear_DeathRates

/* Construct survival curves and calculate life expectancy */
sort year method incg  age
gen surv=1 if age==40
by year method incg : replace surv=surv[_n-1]-surv[_n-1]*deathrate[_n-1] if !mi(deathrate[_n-1]) & age>40
by year method incg: egen lifeexp=sum(surv)
replace lifeexp=lifeexp+39

keep year method incg lifeexp
duplicates drop

keep if incg==20 | incg==80
reshape wide lifeexp, i(year method) j(incg)
gen Dif_20_80=lifeexp80-lifeexp20

keep year method Dif*
duplicates drop

/* Plot trends */
qui reg Dif_20_80 c.year if method=="NIM"
```

```
local bNIM=_b[c.year]
local cNIM=_b[_cons]
local rbNIM="`:di %04.2f _b[c.year]'"
local rclNIM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuNIM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

qui reg Dif_20_80 c.year if method=="MS_1983"
local bM8=_b[c.year]
local cM8=_b[_cons]
local rbM8="`:di %04.2f _b[c.year]'"
local rclM8="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuM8="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

qui reg Dif_20_80 c.year if method=="MS"
local bM=_b[c.year]
local cM=_b[_cons]
local rbM="`:di %04.2f _b[c.year]'"
local rclM="`:di %04.2f _b[c.year]-1.96*_se[c.year]'"
local rcuM="`:di %04.2f _b[c.year]+1.96*_se[c.year]'"

twoway ///
 (scatter Dif_20_80 year if method=="NIM", mcolor(black)) (scatter Dif_20_80 year if method=="MS", mcolor(red))
 (scatter Dif_20_80 year if method=="MS_1983", mcolor(green))  ///
 (function y=`cNIM'+`bNIM'*x , range(1983 2013) lcolor(black))  (function y=`cM'+`bM'*x , range(1983 2013)
 lcolor(red)) (function y=`cM8'+`bM8'*x , range(1983 2013) lcolor(green)), ///
graphregion(color(white)) xlabel(1985(5)2015) ylabel(0(2)8, nogrid) xtitle(Year)  ytitle("Difference LE, P20-P80") ///
legend(order(1 2 3) label(1 "No income mobility") label(2 "Mover-stayer") label(3 "Mover-stayer, 1983-Mobility")
rows(3) ring(0) pos(11) region(color(none))) ///
yscale(titlegap(*1.5)) xscale(titlegap(*6)) text(8.2 2012.8 "Annual change:", bcolor(white) box)  ///
text(7.15 2013.8 "`rbNIM'" "[`rclNIM'-`rcuNIM']", color(black) bcolor(white) box) ///
text(2.6 2013.8 "`rbM'" "[`rclM'-`rcuM']", color(red) bcolor(white) box) ///
text(4.3 2013.8 "`rbM8'" "[`rclM8'-`rcuM8']", color(green) bcolor(white) box)  ///
xsize(8.4) ysize(6) t2title("Keeping mobility constant", position(12) ring(1) size(large)) title("A", ring(1)
position(11) size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_10_Gradient1983Mobility.gph ,  replace

/** Panel B - Mobility rates in mover-stayer in different years **/
clear all
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"
use Main_1_men_ByYear_MobilityRates

replace incg=ceil(incg/5)
collapse (mean) avincg , by(year age incg)
replace incg=incg*5-2

keep if age==60
keep if year==1983 | year==1993 | year==2003 | year==2013
reshape wide avincg, i(incg) j(year)

twoway (function y=x, range(0 100) lcolor(black) lwidth(*1.3))  ///
       (connect avincg1983  incg, color(gs2) ) ///
       (connect avincg1993 incg, color(gs5) ) ///
       (connect avincg2003 incg, color(gs8) ) ///
       (connect avincg2013 incg, color(gs11)), ///
graphregion(color(white)) legend(order(1 2 3 4 5)  label(2 "Mover-stayer 1983") label(3 "Mover-stayer 1993") label(4
"Mover-stayer 2003") label(5 "Mover-stayer 2013") label(1 "No income mobility")  rows(5) pos(11) ring(0)
region(color(none))) ///
ytitle(Income percentile at age 60) xtitle(Income percentile at age 40) yscale(titlegap(*2)) xscale(titlegap(*6))
ylabel(,nogrid) ///
xsize(8.4) ysize(6) t2title("Mobility over time", position(12) ring(1) size(large)) title("B", ring(1) position(11)
size(vhuge) color(black))

graph save H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_10_MobilityOverTime.gph ,  replace

/** Combine Figures **/
cd "H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs"

graph combine  SIFigures_10_Gradient1983Mobility.gph SIFigures_10_MobilityOverTime.gph, ///
               rows(1) graphregion(color(white) margin(l-10)) /*iscale(*0.85)*/ imargin(l+20) xsize(17.8) ysize(6)

graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_10.png , width(2000) replace
graph export H:\workdata\704838\Benjamin\DeathIncomeGradient\Figs\SIFigures_10.eps , replace
```