Replication data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts

Konstantin Kashin, Gary King, and Samir Soneji

March 2015

This document provides information needed to replicate the results in Kashin, King and Soneji (In press, 2015). All the figures in the online version of the paper that are in color are generated by the code in `analysis.Rnw`. This is a Sweave file that can be compiled using the `knitr` package in R. `knitr` runs the `R` code, outputs the figures into a `figures` subdirectory, and creates the `analysis.tex` file. This file can then be compiled using `pdflatex`. Grayscale analogs of the color figures can be generated by the code in `analysis_grayscale.Rnw`. This workflow is in the `Makefile` and can be run by typing `make` in the command line. To compile just the color or grayscale figures, type `make color` or `make grayscale`, respectively. In addition to having LaTeX and the base version of `R` installed, you need to have the following `R` packages: `knitr`, `ggplot2`, `reshape2`, and `gridExtra`.

All data needed for replication is available in the `data` subdirectory. In addition to the original data, we also provide all data in a tabular `dta` format. The `helpers` subdirectory contains an auxiliary script called by `analysis.Rnw` and `analysis_grayscale.Rnw`.

Data

Human Mortality Database Life Tables

`hmd_fltper_1x1.txt` and `hmd_mltper_1x1.txt` contain the raw period life tables for females and males, respectively, for the United States. This is obtained from the Human Mortality Database (HMD).

`lt.RData` contains a dataframe called `lt` that has life tables from HMD across 39 different countries. The variables in the dataframe are standard life table variables from HMD, with the exception of `log.qx`, the log of the conditional probability of death that we calculated ourselves.

Source:
Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on June 1, 2014).

SSA's Forecasts of Cause-Specific Mortality

`ssa_mx_2009_array.RData` contains a 4-dimensional array called `ssa.mx.2009.array` that has observed and forecast cause-specific mortality rates from the SSA. The four dimensions are: age group, year, sex, and cause. Mortality for 1979-2005 is observed and post-2005 values are forecasts.

Source:
Data obtained in personal correspondence with Felicitie Bell on May 15, 2009.

SSA's Performance in Forecasting Life Expectancy

`ssa_ex.RData` contains a single data frame, called `ex.ssa`. The variables in the data frame are:

`TR:` year of Trustees Report

`forecast.year:` year forecast

`sex:` sex (Male or Female)

`age:` age (0 or 65)

`forecast:` SSA's life expectancy forecast for intermediate cost scenario ("best guess" scenario)

`high:` SSA's life expectancy forecast for high cost scenario

`low:` SSA's life expectancy forecast for low cost scenario

`hmd_observed:` Observed life expectancy (HMD)

`ssa_observed:` Observed life expectancy (SSA)

`hmd_residual:` residual based on HMD observed values (`forecast-hmd_observed`)

`ssa_residual:` residual based on SSA observed values (`forecast-ssa_observed`)

`lower:` lower bound of SSA "confidence interval"

`upper` : upper bound of SSA "confidence interval"

`ci_width:` width of SSA's "confidence interval" (`upper-lower`)

`hmd_coverage`: coverage of SSA's "confidence interval" based on HMD observed value (`TRUE` if HMD observed value falls inside the SSA "confidence interval")

`ssa_coverage`: coverage of SSA's "confidence interval" based on SSA observed value (`TRUE` if SSA observed value falls inside the SSA "confidence interval")

Sources: We collected all life expectancy forecasts published in the annual Trustees Reports 1982–2010. In reports prior to 2001, SSA published life expectancy at birth and at age 65 forecasts for males and females projected in 5-year intervals for a total of 75 years into the future. Post-2001, supplementary single-year tables are included online. Our sources are Table 11 of Trustees Reports 1982-1991, Table II.D.2 of Trustees Reports 1992-2000, and Table V.A3 of the supplemental single-year tables of Trustees Reports 2001-2010.

SSA's Performance in Forecasting Balance and Cost Rate of the Trust Funds

`ssa_balance.RData` contains a single data frame, called `tf.balance.pred`.

The variables in `tf.balance.pred` (where `[metric]` is a placeholder for `income`, `cost`, or `balance`) are:

`TR`: year of Trustees Report

`forecast.year`: year forecast

`income.rate`: forecast income rate for intermediate cost scenario ("best guess" scenario)

`cost.rate :` forecast cost rate for intermediate cost scenario ("best guess" scenario)

`balance`: forecast balance for intermediate cost scenario ("best guess" scenario)

`[metric].lower :` lower bound of forecast scenarios

`[metric].upper :` upper bound of forecast scenarios

`[metric].I`: forecast for low-cost scenario / alternative I

`[metric].III`: forecast for high-cost scenario / alternative III

`[metric].IIA`: forecast for second intermediate scenario / alternative II-A

`[metric].truth`: observed value

Notes and sources: For 1978–2012, we take the observed cost rate and balance from Table IV.B1 and the observed trust fund ratio from Table IV.B3 of the 2013 Trustees Report (`http://j.mp/2013tables`). We calculate residuals as the difference between SSA's "best guess" projection (intermediate-cost scenario / alternative II) and the historic statistics from SSA. Note that for 1981–1990, Trustees Reports have two intermediate-cost scenarios: alternative II-A and II-B. For these years, we follow

subsequent Trustees Reports and use II-B as the "best guess" projection. To calculate the uncertainty interval, we use the minimum and maximum values of projected life expectancy across the three scenarios (four for 1981-1990).

Sources for SSA projections published in the annual Trustees Reports (TR) from 1978 to 2012:

Cost rate: Table 26 of TR 1978, Table 27 of TR 1980-1980, Table 28 of TR 1982, Table 29 of TR 1982-1983, Table 30 of TR 1984-1985, Table 28 of TR 1986, Table 26 of TR 1987-1991, Table II.F.13 of TR 1992-2000, Table IV.B1 of the supplemental single-year tables of TR 2001-2012.

Balance: For Trustees Reports from 1986, balance projections are available from the same tables as the cost rate projections. In 1978-1982 TR, the Trustees project the same scheduled tax rate across the cost scenarios (available in Table 25 for TR 1978, Table 26 for TR 1979-1980, Table 28 for TR 1981, and Table 29 for TR 1982). We subtract the cost rates across the scenarios from the scheduled income rate to obtain the range of balance projections. For 1983-1984 TR, the income tax rate varies slightly across the cost scenarios and is only published for the two intermediate projections (Table 27 for TR 1983 Table 28 for TR 1984). We are thus unable to evaluate coverage of SSA's uncertainty intervals for projections made in these two reports.

Policy Proposal Scoring

We collected all policy proposals submitted by Congress and select public policy organizations (`http://www.ssa.gov/oact/solvency/`). The Social Security Administration Office of the Chief Actuary estimated the effect of each proposal on Social Security solvency (e.g., the cost rate and the Trust Fund balance). We denote the estimated effect of a policy proposal as its 'score'. The Social Security Administration scores each proposal using the same methodology used to produce the annual Trustees Report. We focus on scores for the cost rate and Trust Fund balance [1] ten years post policy enactment and the [2] average score over 75 years post policy enactment. We then calculate for each score the proportion of observed forecast errors since 2000 that were as large or larger than the specific score. The proportion represents the inferred significance level.

The policy proposal scoring is available in `proposal_scoring.csv`. The variables in the table are:

`month:` month of policy scoring memorandum prepared by SSA

`day:` day of policy scoring memorandum prepared by SSA

`year:` year of policy scoring memorandum prepared by SSA

`sponsor:` developer(s) of proposal.

`variable:` measure of financial effect on the Trust Funds; either "change in actuarial balance" or "change in cost rate"

`change.10:` score 10 years post-policy enactment

change.max.10: maximum score over the first 10 years post-policy enactment

change.summarized.rate: average score over 75 years post-policy enactment

Sources: All policy proposals submitted by Congress and select public policy organizations as of April 1, 2015. Available at http://www.ssa.gov/oact/solvency/.

Ultimate Rates of Decline

urod.csv contains the ultimate rates of decline (UROD) of mortality.

year: year of Trustees Report

urod: ultimate rate of decline of mortality

type: type of ultimate rate of decline (TR=Trustees Report, TAR=Technical Advisory Report)

urod_arrows.csv is a helper file that is used to correctly position the arrows in Figure 5 of analysis.pdf. The starting coordinate of the arrow is specified by the variables (fromx,fromy). The end coordinate is given by (tox,toy).The type of arrow is given by type.

Sources: We gather the intermediate-cost scenario ("best guess") URODs assumed by 1991-2013 Trustees Reports and the URODs recommended by the four Technical Panels commissioned by the Social Security Advisory Board (quadrennially starting with 1999). For the 1999 TR, we average the male and female UROD as reported in the 2003 Technical Panel Report (http://j.mp/SSATech03). Page 71 of the 2001 TR gives the UROD assumed by 2000 and 2001 TR. Table II.C1 of TR 2002-2009 give the assumed UROD. UROD for 2010 TR is found on page 80 of the report. For 2011, we average the male and female UROD - 0.75% and 0.73%, respectively - as reported in Table 2.3 of *Long-Range Demographic Assumptions for the 2013 Trustees Report* (http://j.mp/OCACT12). For 2012 and 2013 TR, the UROD are found in Table 2.2 of the respective *Long-Range Demographic Assumptions* (http://j.mp/OCACT12; http://j.mp/OCACT13).

UROD recommendations made by the 1999 and 2003 Technical Panel are available in Table 3 of the 2003 Technical Panel Report (http://j.mp/SSATech03). The recommended UROD by the 2007 Technical Panel is available in Table 1 of their report (http://j.mp/SSATech07). Although the 2011 Technical Panel doesn't make its recommendations in terms of UROD, *The Long-Range Demographic Assumptions for the 2013 Trustees Report* converts the 2011 Technical Panel's life expectancy recommendation into an UROD.

References

Kashin, Konstantin, Gary King and Samir Soneji. In press, 2015. "Explaining Systematic Bias and Nontransparency in U.S. Social Security Administration Forecasts." *Political Analysis* . 1