Getting started: Put inputdata and outputdata in a single folder. Make a dofiles and outputtables folder, and put polvol_20130305 (that's an underscore) in dofiles. All data management and analysis take place in STATA (we used mostly 10 SE). In main.do, put the folder's name in $direct (instead of polvol) and put the pathway to that folder in $startdir.

Subdirectories:

- polvol_20130305 contains the final set of our daily do-files. Main.do has notes on what each do-file does. Should be enclosed in a folder named dofiles that is in the same folder as inputdata, outputdata, and outputtables.
- inputdata contains our raw input files. If all lines in main.do are uncommented and the $startdir and $direct globals are edited to find inputdata, these files will be transformed into the files currently in outputdata, from which all the tables can be recreated. We obtained the 1% Census and ACS files from IPUMS (July 2010 versions). The economic conditions data come from various sources detailed in Section 3.2 in the paper.
- outputdata contains the final versions of our basic analytical data files. That is to say, all of the do-files that begin with "tab" call on one of these files and each of these files is required to run one of the tables in the paper. These files are created from the IPUMS datasets and raw state and economic data (which can be found in inputdata) using the code in the do-files that precede the "tab" files in main.do (i.e., "load", "make", "merge", etc.)
- final outputtables contains the final versions of each table that were reformatted into the tables in the body and appendix. These tables are produced by running the do-files that begin with "tab" on the files in outputdata; which will make new versions in outputtables_20130305. The text files created by STATA are then pasted into the corresponding pages in Excel (tabletemplates_20111113.xls) to make the tables in the form they appear in the paper.

Key variables for Tables 3 & 4, found in statecohortswconditionsfipsyes.dta

- *varloginc*: cross-sectional variance of residual log income for individuals in a birth-state cohort (*A* is their five year age bin, *C* is their five year birth cohort where *C*=1 was born 1881-1885, *year* is the Census or ACS year, and *fips* is their birth state). Each observation in the data is an observation on a birthstate-cohort X year (*fips* X *C* X *year* or *fips* X *A* X *year*)
- *[adjcumsum/currrecent] neggrowth, dgdp, dsgdp2,* and *empshare*: years of negative state growth, state GDP growth and demeaned state GDP growth squared (run together), and the employment prediction variable described in Section 3.2. The prefix *adjcumsum* indicates the sum from the beginning of work life to five years ago. The prefix *currrecent* indicates the sum over the past five years of work life.
- *[betaadjcumsum/betacurrrecent] recsty, nipadgdp_neg, recmonth, nipadgdp, and nipadgdp_nsq2*: Recession year (see Section 3.2), annual GDP growth was negative, share of year in recession, and national GDP growth and demeaned national GDP growth squared (run together). The prefix *betaadjcumsum* indicates that the variable is multiplied by the state's beta from Table A.2 and then summed from the beginning of work life to five years ago. The prefix *betacurrrecent* indicates the variable is multiplied by the state's beta from Table A.2 and then summed over the past five years.