```
/* Copyright 2006 by Chris Sanchirico*/

* What this file does
* ======================
/* It takes an individual level data set from PSID for all years and creates a family level data set for all years with all necessary
variables.  Details on any given variable are available from the PSID data center.  Type the variable name into the search facility.*/

* Open raw file and rename
* ====================
use "C:\Documents and Settings\csanchir\My Documents\Data Sources\PSID & CDS\PSID Work hours and Potential Income 79 89 00 03\All
years\All years individual with family with BRR variables raw.dta", clear
save "C:\Documents and Settings\csanchir\My Documents\Data Sources\PSID & CDS\PSID Work hours and Potential Income 79 89 00 03\All
years\All years obs on family with BRR variables.dta", replace

*Drop observations other than for heads
*====================================
       g year = (er30313 < .)*79 + ( er30642 < .)*89 + ( er33601 < .)*00
       g hd79 = (er30314 >= 1 & er30314 <= 20 & er30315 == 1)  /* sequence number between 1-20 (in family unit at time of interview)
and relation to head = 1*/
       g hd89 = (er30643 >= 1 & er30643 <= 20 & er30644 == 10) /* sequence number between 1-20 (in family unit at time of interview)
and relation to head = 10*/
       g hd00 = (er33602 >= 1 & er33602 <= 20 & er33603 == 10) /* sequence number between 1-20 (in family unit at time of interview)
and relation to head = 10*/

       drop if hd79 ~= 1 & hd89 ~= 1 & hd00 ~= 1

*Create variables
*=============

       * Family Identifier
       * ---------------------

       g famid79 =  er30313
       g famid89 =  er30642
       g famid00 =  er33601

       * Survey design
       * -------------------
              /*Early PSID documentation (e.g., Morgan & Smith (1968) contains some discussion of PSID's initial sample design.  The
initial combination of two samples was later reweighted for differential attrition and also poststratification to conform to the
census statistics.  Further subsamples were added for Latinos and immigrants and families were dropped to save costs, etc...*/

              g famweight79 = v7451
              g famweight89 = v18943
              g famweight00 = fcwt01 /* PSID had "dropped" its Latino subsample by 2000*/
              g immigrant00 = (er30001 >= 3001 & er30001 <= 3511) /*An immigrant subsample (er30001 = 3001 to 3511) was added by PSID
in 1997 and 1999.  It is not included here, however, because PSID makes no provision for BRR for the immigrant subsample.*/
              g famweight = famweight79
                     replace famweight = famweight89 if year == 89
                     replace famweight = famweight00 if year == 00
              /* The following variables are for use in calculating standard errors via balanced repeated replication.  See PSID
documentation regarding "SECU's."  Stratum and PSU identifiers are in the individual file, but examination reveals that, sensibly,
there a one unique pair for each family.*/
```

```
        g stratum = ER31996
        g psu = ER31997
        drop if stratum > 32 /* Latino subsample dropped */
        drop if stratum == 0 /* immigrant sample is excluded from PSID BRR design*/


        * Specify all-year survey design
        * ======================
        svyset psu [pw=famweight], strata(stratum)


    * Actual average hourly earnings for heads and wives--2000 dollars
    *-------------------------------------------------------------------------------------------

        * PSID has average hourly earnings data only for heads and wives.

        * Inflation adjustment
        *----------------------------
        *$ amounts stated in 2000 $'s using chain-type PCE price index from NIPA, as also used in Blau & Kahn (2006).
        *2000 = 100; 1989 = 76.972; 1979 = 47.059

               g hdavghrlyearn79 = v7417*(100/47.059)
               g wfavghrlyearn79 = v7418*(100/47.059)
               g hdavghrlyearn89 = v18887*(100/76.972)
               g wfavghrlyearn89 = v18888*(100/76.972)
               g hdavghrlyearn00 = hdwge01
               g wfavghrlyearn00 = wfwge01

        * Top-coding
        *----------------
         *1) infrequent in the data;
         *2) Following convention, I multiply top-coded values by 1.5

               replace hdavghrlyearn79 = 1.5*hdavghrlyearn79 if (hdavghrlyearn79 >= 99.99)
               replace wfavghrlyearn79 = 1.5*wfavghrlyearn79 if (wfavghrlyearn79 >= 99.99)
               replace hdavghrlyearn89 = 1.5*hdavghrlyearn89 if (hdavghrlyearn89 >= 99.99)
               replace wfavghrlyearn89 = 1.5*wfavghrlyearn89 if (wfavghrlyearn89 >= 99.99)
                       /*In 2000, top-coding for heads was at 500, which is non binding.  For wives in 2000 it was apparently
also at 500, but the code book does not say so. Only two wives have precisely 500.  I assume these are actually 500.*/

        * Error codes
        *-----------------
        *1) these only appear in 2000
        *2) I convert them to missing values

        replace hdavghrlyearn00 = . if hdavghrlyearn00 ==9998 | hdavghrlyearn00 ==9999
        replace wfavghrlyearn00 = . if wfavghrlyearn00 ==9998 | wfavghrlyearn00 ==9999


    *Work hours and actual earnings
    *----------------------------------------
        /* PSID assigned missing data for work hours.  I use their assignments.*/
        /* wfannwrkhrs__ = 0 if there is no wife in the family unit, or if the wife did not work*/
```

```
g hdannwrkhrs79 = v6934
g wfannwrkhrs79 = v6946
g famannwrkhrs79 = hdannwrkhrs79 + wfannwrkhrs79
g hdfulltime79 = (hdannwrkhrs79 >= 2000)
g wffulltime79 = (wfannwrkhrs79 >= 2000)

g hdactinc79 = hdannwrkhrs79*hdavghrlyearn79 /* PSID defines average hourly earnings to be labor income divided by hours, so this variable merely restores average hourly earnings to labor income*/
g wfactinc79 = wfannwrkhrs79*wfavghrlyearn79
g famactinc79 = wfactinc79+hdactinc79


g hdannwrkhrs89 = v17744
g wfannwrkhrs89 = v17774
g famannwrkhrs89 = hdannwrkhrs89 + wfannwrkhrs89
g hdfulltime89 = (hdannwrkhrs89 >= 2000)
g wffulltime89 = (wfannwrkhrs89 >= 2000)

g hdactinc89 = hdannwrkhrs89*hdavghrlyearn89 /* PSID defines average hourly earnings to be labor income divided by hours, so this variable merely restores average hourly earnings to labor income*/
g wfactinc89 = wfannwrkhrs89*wfavghrlyearn89
g famactinc89 = wfactinc89+hdactinc89


g hdannwrkhrs00 = hdtot01
g wfannwrkhrs00 = wftot01
g famannwrkhrs00 = hdannwrkhrs00 + wfannwrkhrs00
g hdfulltime00 = (hdannwrkhrs00 >= 2000)
g wffulltime00 = (wfannwrkhrs00 >= 2000)

g hdactinc00 = hdannwrkhrs00*hdavghrlyearn00 /* PSID defines average hourly earnings to be labor income divided by hours, so this variable merely restores average hourly earnings to labor income*/
g wfactinc00 = wfannwrkhrs00*wfavghrlyearn00
g famactinc00 = wfactinc00+hdactinc00


*Demographic variables for separate prediction of head and wife average hourly earnings
*-------------------------------------------------------------------------------------------------------------------
* Age
/*error codes set to missing data. Top-coding ignored. Age = 0, if no wife in family unit*/

g hdage79 = v7067
      replace hdage79 = . if hdage79 == 99
g wfage79 = v7069
      replace wfage79 = . if wfage79 == 99
g hdagesq79 = hdage79^2
g wfagesq79 = wfage79^2

g hdage89 = v18049
      replace hdage89 = . if hdage89 == 99
g wfage89 = v18051
      replace wfage89 = . if wfage89 == 99
g hdagesq89 = hdage89^2
g wfagesq89 = wfage89^2
```

```
g hdage00 = er17013
      replace hdage00 = . if hdage00 == 999
g wfage00 =  er17015
      replace wfage00 = . if wfage00 == 999
g hdagesq00 = hdage00^2
g wfagesq00 = wfage00^2


* Gender
      /* Heads may be male or female; wives (cf. "husband of head") are always female.*/
      g hdfemale79 = v7068-1 /*v7068: 1=male; 2=female*/
      g wffemale79 = 1      /* This definition is (purposefully) superfluous*/
      g hdfemale89 = v18050-1 /*v18050: 1=male; 2=female*/
      g wffemale89 = 1      /* This definition is (purposefully) superfluous*/
      g hdfemale00 = er17014-1  /*er17014: 1=male; 2=female*/
      g wffemale00 = 1      /* This definition is (purposefully) superfluous*/

*Race and ethnicity
      g hdblack79 = (v7447 == 2)
            /*  For 1979 PSID gives "race" only for family, not head and wife separately; and even this from
questions asked last in 1972, perhaps of predessor families.*/
      g wfblack79 = (v7447 == 2)
      g hdhisp79 = (v7447 == 3)  /*For 1979 PSID labels this "Spanish-American"*/
      g wfhisp79 = (v7447 == 3)
      g hdblack89 = (v18814 == 2 | v18815 == 2)
      g wfblack89 = (v18749 == 2 | v18750 == 2)
            /*  There are two mentions of race.   A individual is here considered black if to either mention the
answer is black.*/
      g hdhisp89 = (v18814 == 5 | v18815 == 5)
      g wfhisp89 = (v18749 == 5 | v18750 == 5)
            /* See note on black.  The label in 1989 is "Latino origin or descent."*/
      g hdblack00 = (er19989 == 2 | er19990 == 2 | er19991 == 2 | er19992 == 2)
      g wfblack00 = (er19897 == 2 | er19898 == 2 | er19899 == 2 | er19900 == 2)
            /*There are several mentions of race.   A individual is here considered black if to any mention the
answer is black.*/
      g hdhisp00 = (er19989 == 5 | er19990 == 5 | er19991 == 5 | er19992 == 5)
      g wfhisp00 = (er19897 == 5 | er19898 == 5 | er19899 == 5 | er19900 == 5)
            /*see note on black.  The label in 2000 is "Latino origin or descent."*/

*Education
      /* All education variables are cumulative: e.g., hdhs79 = 1 if head completed at least high school, and possibly
college etc....*/

      g hdeduc79 = v7433
            replace hdeduc79  = . if (hdeduc79 == 9)
      g hdhs79 = (hdeduc79 >= 4)
            note hdhs79: =1 if head completed at least high school
      g hdcoll79 = (hdeduc79 >= 7)
            note hdcoll79: =1 if head completed at least college
      g hdadvprof79 = (hdeduc79 >= 8)

      g wfeduc79 = v7434 /*In contrast to hdeduc79, 0 here means no wife in family unit, as opposed to illiterate*/
            replace wfeduc79  = . if (wfeduc79 == 9)
```

```
g wfhs79 = (wfeduc79 >= 4)
       note wfhs79: =1 if wife completed at least highschool
g wfcoll79 = (wfeduc79 >= 7)
       note wfcoll79: =1 if wife completed at least college
g wfadvprof79 = (wfeduc79 >= 8)

/* For 1989 education last asked of preexisting heads and wives in 1985 and then carried forward.  Asked of new
head or wives when that status was attained.*/
g hdeduc89 = v18898
       replace hdeduc89  = . if (hdeduc89 == 9)
g hdhs89 = (hdeduc89 >= 4)
       note hdhs89: =1 if head completed at least highschool
g hdcoll89 = (hdeduc89 >= 7)
       note hdcoll89: =1 if head completed at least college
g hdadvprof89 = (hdeduc89 >= 8)

g wfeduc89 = v18899
       replace wfeduc89  = . if (wfeduc89 == 9)
g wfhs89 = (wfeduc89 >= 4)
       note wfhs89: =1 if wife completed at least highschool
g wfcoll89 = (wfeduc89 >= 7)
       note wfcoll89: =1 if wife completed at least college
g wfadvprof89 = (wfeduc89 >= 8)

/* For 2000 education last asked of preexisting heads and wives in 1995 and then carried forward.  Asked of new
head or wives when that status was attained. This variable not completely comparable to education variables used in earlier years.*/
g hdeduc00 = upedu01h
       replace hdeduc00  = . if (hdeduc00 == 99)
g hdhs00 = (hdeduc00 >= 12)
g hdcoll00 = (hdeduc00 >= 16)
g hdadvprof00 = (hdeduc00 >= 17)

g wfeduc00 = upedu01w
       replace wfeduc00  = . if (wfeduc00 == 99)
g wfhs00 = (wfeduc00 >= 12)
g wfcoll00 = (wfeduc00 >= 16)
g wfadvprof00 = (wfeduc00 >= 17)

*Geography
       /* In 1979, as opposed to 1989 and 2000, there are no Beale-Ross Rural Urban codes in PSID; this measure did not
exist in 1979.  For consistency across years, and at the expense of a more valid coding convention in 1989 and 2000, in all years I
code an individual as urban if the largest urban area ("SMSA") in the primary sampling unit (county, county group, or SMSA) has at
least 100,000 people.*/

gen region79 = V7419
       replace region79 = . if region79 == 9
       note region79: 1=northeast; 2=northcentral; 3=south; 4=west; 5=Alaska, Hawaii; 6 = foreign country; 9 =
NA
gen urban79 = (V6906 <= 2)
       note urban79: urban79 = 1 if largest urban area ("SMSA") in PSU (county, county group, or SMSA) has at
least 100,000 people.  Otherwise urban79 = 0.

gen region89 = V18889
```

```
                        note region89: 1=northeast; 2=northcentral; 3=south; 4=west; 5=Alaska, Hawaii; 6 = foreign country; 9 =
NA
                gen urban89 = (V17706 <= 2)

                gen region00 =  ER20376
                        note region00: 1=northeast; 2=northcentral; 3=south; 4=west; 5=Alaska, Hawaii; 6 = foreign country; 9 =
NA
                gen urban00 = (ER20378 <= 2)
                        note urban00: urban00 = 1 if the largest urban area ("SMSA") in the primary sampling unit (county,
county group, or SMSA) has at least 100,000 people; otherwise urban00 = 0.

        *Additional variables for Heckman selection equation
        *-------------------------------------------------------------------

                g numchld79 =  v7070 /* actual number of children in family unit aged 0-17*/
                g numchld89 =  v18052
                g numchld00 =  er17016

                g couple79 = (v7069 ~= 0)
                        /* = 1 if age of wife non zero, which indicates presence of wife.  Open question: This dummy seems to have the
value zero for female-headed couples.  For such couples there is no PSID wife.  PSID FAQ's say that the males in such couples are
usually disabled.*/
                g couple89 = (v18051 ~=0)
                g couple00 = (er17015 ~= 0)

* Generate region and numchld dummies
* ==============================

* 1979
* ------

g region1_79 = 1 if region79 == 1
replace region1_79 = 0 if region79 ~= 1 & region79 < .
g region2_79 = 1 if region79 == 2
replace region2_79 = 0 if region79 ~= 2 & region79 < .
g region3_79 = 1 if region79 == 3
replace region3_79 = 0 if region79 ~= 3 & region79 < .
g region4_79 = 1 if region79 == 4
replace region4_79 = 0 if region79 ~= 4 & region79 < .
g region5_79 = 1 if region79 == 5
replace region5_79 = 0 if region79 ~= 5 & region79 < .
g region6_79 = 1 if region79 == 6
replace region6_79 = 0 if region79 ~= 6 & region79 < .

g numchld0_79 = 1 if numchld79 >= 0
replace numchld0_79 = 0 if numchld79 < 0 & numchld79 < .
g numchld1_79 = 1 if numchld79 >= 1
replace numchld1_79 = 0 if numchld79 < 1 & numchld79 < .
g numchld2_79 = 1 if numchld79 >= 2
replace numchld2_79 = 0 if numchld79 < 2 & numchld79 < .
g numchld3_79 = 1 if numchld79 >= 3
replace numchld3_79 = 0 if numchld79 < 3 & numchld79 < .
g numchld4_79 = 1 if numchld79 >= 4
replace numchld4_79 = 0 if numchld79 < 4 & numchld79 < .
```

```
g numchld5_79 = 1 if numchld79 >= 5
replace numchld5_79 = 0 if numchld79 < 5 & numchld79 < .
g numchld6_79 = 1 if numchld79 >= 6
replace numchld6_79 = 0 if numchld79 < 6 & numchld79 < .
g numchld7_79 = 1 if numchld79 >= 7
replace numchld7_79 = 0 if numchld79 < 7 & numchld79 < .
g numchld8_79 = 1 if numchld79 >= 8
replace numchld8_79 = 0 if numchld79 < 8 & numchld79 < .

* 1989
* ------

g region1_89 = 1 if region89 == 1
replace region1_89 = 0 if region89 ~= 1 & region89 < .
g region2_89 = 1 if region89 == 2
replace region2_89 = 0 if region89 ~= 2 & region89 < .
g region3_89 = 1 if region89 == 3
replace region3_89 = 0 if region89 ~= 3 & region89 < .
g region4_89 = 1 if region89 == 4
replace region4_89 = 0 if region89 ~= 4 & region89 < .
g region5_89 = 1 if region89 == 5
replace region5_89 = 0 if region89 ~= 5 & region89 < .
g region6_89 = 1 if region89 == 6
replace region6_89 = 0 if region89 ~= 6 & region89 < .

g numchld0_89 = 1 if numchld89 >= 0
replace numchld0_89 = 0 if numchld89 < 0 & numchld89 < .
g numchld1_89 = 1 if numchld89 >= 1
replace numchld1_89 = 0 if numchld89 < 1 & numchld89 < .
g numchld2_89 = 1 if numchld89 >= 2
replace numchld2_89 = 0 if numchld89 < 2 & numchld89 < .
g numchld3_89 = 1 if numchld89 >= 3
replace numchld3_89 = 0 if numchld89 < 3 & numchld89 < .
g numchld4_89 = 1 if numchld89 >= 4
replace numchld4_89 = 0 if numchld89 < 4 & numchld89 < .
g numchld5_89 = 1 if numchld89 >= 5
replace numchld5_89 = 0 if numchld89 < 5 & numchld89 < .
g numchld6_89 = 1 if numchld89 >= 6
replace numchld6_89 = 0 if numchld89 < 6 & numchld89 < .
g numchld7_89 = 1 if numchld89 >= 7
replace numchld7_89 = 0 if numchld89 < 7 & numchld89 < .
g numchld8_89 = 1 if numchld89 >= 8
replace numchld8_89 = 0 if numchld89 < 8 & numchld89 < .

* 2000
* ------

g region1_00 = 1 if region00 == 1
replace region1_00 = 0 if region00 ~= 1 & region00 < .
g region2_00 = 1 if region00 == 2
replace region2_00 = 0 if region00 ~= 2 & region00 < .
g region3_00 = 1 if region00 == 3
replace region3_00 = 0 if region00 ~= 3 & region00 < .
g region4_00 = 1 if region00 == 4
```

```
replace region4_00 = 0 if region00 ~= 4 & region00 < .
g region5_00 = 1 if region00 == 5
replace region5_00 = 0 if region00 ~= 5 & region00 < .
g region6_00 = 1 if region00 == 6
replace region6_00 = 0 if region00 ~= 6 & region00 < .

g numchld0_00 = 1 if numchld00 >= 0
replace numchld0_00 = 0 if numchld00 < 0 & numchld00 < .
g numchld1_00 = 1 if numchld00 >= 1
replace numchld1_00 = 0 if numchld00 < 1 & numchld00 < .
g numchld2_00 = 1 if numchld00 >= 2
replace numchld2_00 = 0 if numchld00 < 2 & numchld00 < .
g numchld3_00 = 1 if numchld00 >= 3
replace numchld3_00 = 0 if numchld00 < 3 & numchld00 < .
g numchld4_00 = 1 if numchld00 >= 4
replace numchld4_00 = 0 if numchld00 < 4 & numchld00 < .
g numchld5_00 = 1 if numchld00 >= 5
replace numchld5_00 = 0 if numchld00 < 5 & numchld00 < .
g numchld6_00 = 1 if numchld00 >= 6
replace numchld6_00 = 0 if numchld00 < 6 & numchld00 < .
g numchld7_00 = 1 if numchld00 >= 7
replace numchld7_00 = 0 if numchld00 < 7 & numchld00 < .
g numchld8_00 = 1 if numchld00 >= 8
replace numchld8_00 = 0 if numchld00 < 8 & numchld00 < .

* Family average hourly earnings
* ========================
               g famavghrlyearn79 = hdavghrlyearn79
                      replace famavghrlyearn79 = hdavghrlyearn79 + wfavghrlyearn79 if couple79 == 1
               g famavghrlyearn89 = hdavghrlyearn89
                      replace famavghrlyearn89 = hdavghrlyearn89 + wfavghrlyearn89 if couple89 == 1
               g famavghrlyearn00 = hdavghrlyearn00
                      replace famavghrlyearn00 = hdavghrlyearn00 + wfavghrlyearn00 if couple00 == 1

* Hadamard matrices for balanced repeated replication variance estimation
* ========================================================

/* Begin note:
1.  The following hadamard matrices are created using Kronecker matrix multiplication.  They conform to hadamards provided by Wolter
(1985) pp. 320-352.  Other hadamards could also have been used.  See Wolger (1985) pp. 111-116 for a discussion of Balanced Half-
Sample Replication and the role of Hadamards, and in particular their two important properties.
End note.*/

matrix h2 = (1,1\1,-1)
matrix h4 = h2#h2
matrix h8 = h2#h4
matrix h16 = h2#h8
matrix h32 = h2#h16
matrix h64 = h2#h32
matrix h128 = h2#h64

/* Begin note:
```

1. The following Hadamard matrix, which cannot be created by Kronecker matrix multiplication as above, is taken from the website of Professor Jennifer Seberry, http://www.uow.edu.au/~jennie/hadamard.html, and is there attributed to Vladimir Tonchev, Nagoya Math. J. 104 (1986), 163-174.
2. Notice the matrix input command, as opposed to just matrix. Matrix input allows the inputting of larger matrices, but does not evaluate expression for entries.
End note */

```
matrix   input   h36   =   (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1\1,1,-1,-1,-1,1,-1,-1,1,-1,-1,-1,1,-
1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1\1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-
1,1,-1,1,1\1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1\1,1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-
1,-1,1,-1,1,-1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,1\1,1,1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,1,-1,-
1,1,-1,1,-1,-1,1\1,-1,1,1,1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,1\1,1,-1,1,1,1,1,1,1,-1,-1,-
1,1,-1,-1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,1\1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,1,-1,-1,1,-1,1,1,1,-1,-
1,-1,1,1,-1,-1,1,-1,1\1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,-1,-1,1,-1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,1\1,-1,-1,-1,1,-
1,1,1,1,1,1,-1,-1,-1,1,-1,-1,1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1\1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,-1,-1,-1,1,-1,1,-
1,-1,1,-1,1,1,1,-1,-1,-1,1,1,-1,1\1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,-1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,1\1,-1,-
1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,-1,-1,-1,1,1\1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,1,-
1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,1\1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,1\1,-
1,-1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,1\1,-1,-1,-1,1,1,-1,-1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,-
1,-1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,1\1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-
1\1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,-1\1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-
1,1,1,1,-1,-1,1,-1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1\1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-
1,1,1,1,-1,-1,-1,-1\1,1,-1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,-1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1\1,1,1,-1,1,-1,-1,1,1,-
1,1,-1,-1,1,1,1,-1,-1,-1,1,1,-1,-1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1\1,1,1,1,-1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,1,1,-
1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1\1,-1,1,1,1,1,-1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,-1\1,-1,-1,1,1,1,1,-
1,1,-1,-1,1,-1,1,-1,-1,1,1,1,-1,1,1,1,-1,-1,-1,-1,-1,1,1,-1,1,1,1,1,-1,1,1,1,-1,1,1,-1\1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,1,1,1,1,1,-1,-1,-
1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-1,-1\1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,1,-1,1,1,1,-1\1,1,1,-1,-
1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,1,-1,1,1,-1\1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,1,1,1,-
1,1,1,1,-1,-1,-1,-1,-1,1,1,-1,1,1,1,1,-1,-1\1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,1,-
1\1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,-1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1,1,1,1,-1\1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-
1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1,1,-1\1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-
1,-1\1,-1,1,-1,1,-1,-1,1,1,1,-1,-1,-1,1,1,1,1,-1,1,1,-1,-1,1,1,1,1,-1,1,1,1,-1,1,1,1,1,-1,-1,-1,-1,-1,1,-1)

* Variables without year indices
* =======================
```

/* Begin note:

1. The following variables are generated to avoid case-wise deletion in Stata e-class programs that manipulate observations from more than one year. Such programs delete observations with missing values for any variable. Thus, if the program took 1979 and 1989 variables, for instance, no observations would be passed to the program. 1979 observations would be deleted for missing values for 1989 variables and vice versa.

End note.*/

```
g famavghrlyearn = famavghrlyearn79
	replace famavghrlyearn = famavghrlyearn89 if year == 89
	replace famavghrlyearn = famavghrlyearn00 if year == 00
g hdavghrlyearn = hdavghrlyearn79
	replace hdavghrlyearn = hdavghrlyearn89 if year == 89
	replace hdavghrlyearn = hdavghrlyearn00 if year == 00
g hdage = hdage79
	replace hdage = hdage89 if year == 89
	replace hdage = hdage00 if year == 00
```

```
g hdagesq = hdagesq79
       replace hdagesq = hdagesq89 if year == 89
       replace hdagesq = hdagesq00 if year == 00
g hdfemale = hdfemale79
       replace hdfemale = hdfemale89 if year == 89
       replace hdfemale = hdfemale00 if year == 00
g hdblack = hdblack79
       replace hdblack = hdblack89 if year == 89
       replace hdblack = hdblack00 if year == 00
g hdhisp = hdhisp79
       replace hdhisp = hdhisp89 if year == 89
       replace hdhisp = hdhisp00 if year == 00
g hdhs = hdhs79
       replace hdhs = hdhs89 if year == 89
       replace hdhs = hdhs00 if year == 00
g hdcoll = hdcoll79
       replace hdcoll = hdcoll89 if year == 89
       replace hdcoll = hdcoll00 if year == 00
g hdadvprof = hdadvprof79
       replace hdadvprof = hdadvprof89 if year == 89
       replace hdadvprof = hdadvprof00 if year == 00
g region = region79
       replace region = region89 if year == 89
       replace region = region00 if year == 00
g urban = urban79
       replace urban = urban89 if year == 89
       replace urban = urban00 if year == 00
g hdfulltime = hdfulltime79
       replace hdfulltime = hdfulltime89 if year == 89
       replace hdfulltime = hdfulltime00 if year == 00
g numchld = numchld79
       replace numchld = numchld89 if year == 89
       replace numchld = numchld00 if year == 00
g couple = couple79
       replace couple = couple89 if year == 89
       replace couple = couple00 if year == 00
g wfavghrlyearn = wfavghrlyearn79
       replace wfavghrlyearn = wfavghrlyearn89 if year == 89
       replace wfavghrlyearn = wfavghrlyearn00 if year == 00
g wfage = wfage79
       replace wfage = wfage89 if year == 89
       replace wfage = wfage00 if year == 00
g wfagesq = wfagesq79
       replace wfagesq = wfagesq89 if year == 89
       replace wfagesq = wfagesq00 if year == 00
g wffemale = wffemale79
       replace wffemale = wffemale89 if year == 89
       replace wffemale = wffemale00 if year == 00
g wfblack = wfblack79
       replace wfblack = wfblack89 if year == 89
       replace wfblack = wfblack00 if year == 00
g wfhisp = wfhisp79
       replace wfhisp = wfhisp89 if year == 89
       replace wfhisp = wfhisp00 if year == 00
```

```
g wfhs = wfhs79
       replace wfhs = wfhs89 if year == 89
       replace wfhs = wfhs00 if year == 00
g wfcoll = wfcoll79
       replace wfcoll = wfcoll89 if year == 89
       replace wfcoll = wfcoll00 if year == 00
g wfadvprof = wfadvprof79
       replace wfadvprof = wfadvprof89 if year == 89
       replace wfadvprof = wfadvprof00 if year == 00
g wffulltime = wffulltime79
       replace wffulltime = wffulltime89 if year == 89
       replace wffulltime = wffulltime00 if year == 00
g hdannwrkhrs = hdannwrkhrs79
       replace hdannwrkhrs = hdannwrkhrs89 if year == 89
       replace hdannwrkhrs = hdannwrkhrs00 if year == 00
g wfannwrkhrs = wfannwrkhrs79
       replace wfannwrkhrs = wfannwrkhrs89 if year == 89
       replace wfannwrkhrs = wfannwrkhrs00 if year == 00
g famannwrkhrs = famannwrkhrs79
       replace famannwrkhrs = famannwrkhrs89 if year == 89
       replace famannwrkhrs = famannwrkhrs00 if year == 00
g famid = famid79
       replace famid = famid89 if year == 89
       replace famid = famid00 if year == 00
g hd = hd79
       replace hd = hd89 if year == 89
       replace hd = hd00 if year == 00


* Generate predicted average hourly earnings
* ==================================
do "C:\Documents and Settings\csanchir\My Documents\Data Sources\PSID & CDS\PSID Work hours and Potential Income 79 89 00 03\All
years\Wage prediction only 79.do"
do "C:\Documents and Settings\csanchir\My Documents\Data Sources\PSID & CDS\PSID Work hours and Potential Income 79 89 00 03\All
years\Wage prediction only 89.do"
do "C:\Documents and Settings\csanchir\My Documents\Data Sources\PSID & CDS\PSID Work hours and Potential Income 79 89 00 03\All
years\Wage prediction only 00.do"

estprogram79 hdavghrlyearn79 hdage79 hdagesq79 hdfemale79 hdblack79 hdhisp79 hdhs79 hdcoll79 hdadvprof79 region79 urban79 famweight79
hdfulltime79  numchld79  couple79  wfavghrlyearn79  wfage79  wfagesq79  wffemale79  wfblack79  wfhisp79  wfhs79  wfcoll79  wfadvprof79
wffulltime79 hdannwrkhrs79 wfannwrkhrs79 famannwrkhrs79 famid79 hd79 year
estprogram89 hdavghrlyearn89 hdage89 hdagesq89 hdfemale89 hdblack89 hdhisp89 hdhs89 hdcoll89 hdadvprof89 region89 urban89 famweight89
hdfulltime89  numchld89  couple89  wfavghrlyearn89  wfage89  wfagesq89  wffemale89  wfblack89  wfhisp89  wfhs89  wfcoll89  wfadvprof89
wffulltime89 hdannwrkhrs89 wfannwrkhrs89 famannwrkhrs89 famid89 hd89 year
estprogram00 hdavghrlyearn00 hdage00 hdagesq00 hdfemale00 hdblack00 hdhisp00 hdhs00 hdcoll00 hdadvprof00 region00 urban00 famweight00
hdfulltime00  numchld00  couple00  wfavghrlyearn00  wfage00  wfagesq00  wffemale00  wfblack00  wfhisp00  wfhs00  wfcoll00  wfadvprof00
wffulltime00 hdannwrkhrs00 wfannwrkhrs00 famannwrkhrs00 famid00 hd00 year

g prehdavghrlyearn = prehdavghrlyearn79
       replace prehdavghrlyearn = prehdavghrlyearn89 if year == 89
       replace prehdavghrlyearn = prehdavghrlyearn00 if year == 00

g prewfavghrlyearn = prewfavghrlyearn79
       replace prewfavghrlyearn = prewfavghrlyearn89 if year == 89
```

```
        replace prewfavghrlyearn = prewfavghrlyearn00 if year == 00

g prefamavghrlyearn = prefamavghrlyearn79
        replace prefamavghrlyearn = prefamavghrlyearn89 if year == 89
        replace prefamavghrlyearn = prefamavghrlyearn00 if year == 00


* Generate predicted actual income
* ==========================
g hoursregressionsample79 = (hdage79 >= 25 & hdage79 <=55 & (wfage79 >=25 & wfage79 <=55 | couple79 == 0)) & couple79 == 1 & year ==
79
g hoursregressionsample89 = (hdage89 >= 25 & hdage89 <=55 & (wfage89 >=25 & wfage89 <=55 | couple89 == 0)) & couple89 == 1 & year ==
89
g hoursregressionsample00 = (hdage00 >= 25 & hdage00 <=55 & (wfage00 >=25 & wfage00 <=55 | couple00 == 0)) & couple00 == 1 & year ==
00

fracpoly regress famactinc79 prefamavghrlyearn79 numchld1_79 numchld2_79 numchld3_79 numchld4_79 numchld5_79 numchld6_79 numchld7_79
numchld8_79 if hoursregressionsample79 == 1 [pw=famweight79], degree(4)
predict fampreactinc79

fracpoly regress famactinc89 prefamavghrlyearn89 numchld1_89 numchld2_89 numchld3_89 numchld4_89 numchld5_89 numchld6_89 numchld7_89
numchld8_89 if hoursregressionsample89 == 1 [pw=famweight89], degree(4)
predict fampreactinc89

fracpoly regress famactinc00 prefamavghrlyearn00 numchld1_00 numchld2_00 numchld3_00 numchld4_00 numchld5_00 numchld6_00 numchld7_00
numchld8_00 if hoursregressionsample00 == 1 [pw=famweight00], degree(4)
predict fampreactinc00

g taxcohort79 = hoursregressionsample79 == 1 & numchld79 == 2
g taxcohort89 = hoursregressionsample89 == 1 & numchld89 == 2
g taxcohort00 = hoursregressionsample00 == 1 & numchld00 == 2

g fampotinc79 = prefamavghrlyearn79*2000  /*Family predicted average hourly earnings is the sum of same for head and wife. Thus, we
multiply by one set of full time hours, not two eventhough family has two earners.*/
g fampotinc89 = prefamavghrlyearn89*2000
g fampotinc00 = prefamavghrlyearn00*2000

g realizratio79 = fampreactinc79/fampotinc79
g realizratio89 = fampreactinc89/fampotinc89
g realizratio00 = fampreactinc00/fampotinc00


compress
```