

PANEL STUDY OF INCOME DYNAMICS:
1997 PUBLIC RELEASE I FAMILY FILE

TABLE OF CONTENTS

I. INTRODUCTION

A. WHAT'S NEW FOR 1997

1. NEW IMMIGRANT SAMPLE ADDITION
2. SLASHING CENSUS SAMPLE AND RESURRECTION OF BLACK KIDS
3. BACKGROUND QUESTIONS
4. CHILD SUPPORT QUESTIONS AND CHILD DEVELOPMENT SUPPLEMENT

II. CHARACTERISTICS

A. FILES AND FORMAT

1. THE DATA FILE
2. THE DATA DEFINITION STATEMENTS FILES
3. THE CODEBOOK AND MEANS FILES

B. VARIABLE NAMES, POSITIONS AND GENERATED VARIABLES

C. DOCUMENTATION AND CODES

1. DOCUMENTATION
2. 1997 QUESTIONNAIRE
3. CODE CATEGORIES

D. BRINGING FORWARD BACKGROUND INFORMATION FOR HEAD AND WIFE/"WIFE"

E. PROBLEM VARIABLES, MISSING VARIABLES

F. COMPARABILITY WITH 1993 FINAL-RELEASE FILE AND EARLIER
FINAL-RELEASE FILES

G. NOTES ON NON-QUESTIONNAIRE VARIABLES IN 1997

Release 6 Notes: January 2016: Added generated variables splitoff indicator (ER10005F), 1968 Family Identifier (ER10005G), Whether Re-instated Dropped Family (ER10005H), Collapsed Beale Code (ER12221F), Number of Individual records (ER10005I), DE1 Checkpoint (ER10562A) and F1 Checkpoint (ER11044A). In addition we corrected the top code amount for A20 (ER10036) and A21 (ER10037) to match the codebook values.

Release 5 Notes: December 2013: Added generated variables such as Family Composition Change (ER10004A), DUFU variables (ER12223F-ER12223R), Splitoff variables (ER12223D-ER12223E), Births to HD/WF/OFUM (ER12223S-ER12223V), Generated Marital Status (ER12223A), Change in Marital Status (ER12223B), Couple status of Head (ER12223C), Region where Head grew up (ER12221C), HD Geographic Mobility (ER12221D), Current Region (ER12221E). Interview length, mode of interview, language of interview, calls for iw (ER10005B-ER10005E). Brought forward the background section for old Heads and Wives in section KL.

I. INTRODUCTION

A. What's New for 1997

We retained a very few questions from the 1996 module about home financing. We still ask what type of loan the family has, whether it is the original loan or refinanced, and what the current interest rate is. At the least, these will be helpful in imputations.

Otherwise, the core content of the questionnaire remained the same as in 1996, with the addition of both incoming and outgoing child support questions, the re-asking of many background information items for heads and wives/"wives", and some new background questions relevant to the new immigrant sample.

Release 5 of the 1997 Public Release I Family File incorporates variables that had been available previously in supplemental files: Head and Wife/"Wife" occupation and industry codes, reason a job ended, industry codes for businesses owned, Head and Wife/"Wife" field of certificates earnings, Head's father's occupation, Head's first occupation, Head and Wife/"Wife" work hours and work weeks, Head, Wife/"Wife" annual income variables, needs standards, state code, completed education of Head and Wife/"Wife", and a recalculated 1997 core/immigrant family longitudinal weight (for details about the construction of this weight, go to the PSID website, PSIDONLINE.ISR.UMICH.EDU: select 'Data & Documentation', then select 'Sample Weights', then 'PSID Longitudinal Weights', and 'PSID Revised Longitudinal Weights 1993-2005').

In addition, a 1997 core/immigrant family cross-sectional weight has been calculated. For details about the construction of this new weight, go to the PSID internet site, PSIDONLINE.ISR.UMICH.EDU: select 'Data & Documentation', then select 'Sample Weights', then 'PSID Cross-sectional Weights', and 'Cross-sectional Analysis of PSID Data and Cross-sectional Weights 1997-2005'.

These sections were dropped in 1997: the Hispanicity question; 1995 mortgage questions A27a-I, A39a-e, and A39a-g; the 1995 Distress/Bankruptcy questions in Section G; ADLs/IADLs in Section H for Head, Wife/"Wife", and Other Family Unit Members aged 55 or older; Risk Aversion (1995 Section M).

1. New Immigrant Sample Addition

The PSID's original sample was selected in 1968, but since that time the profile of the U.S. population has changed because of immigration. In order to represent the current population more accurately, we mounted a large screening effort for 1997 interviewing. We managed to locate and persuade 441 immigrant families to be interviewed. These families are included on the files along with the core PSID families.

Immigrant families and individuals were assigned a unique range of values so that they can easily be distinguished from the other PSID samples. The 1997 Interview (ID) Number is in the range 10001-10441, and the 1968 Interview (ID) Number is in the range 3001-3441. Person Numbers for immigrants were assigned in the usual manner, with all those present in the first (1997) wave receiving values from 1-20. Institutionals received values of 20-29, and movers-out were given nonsample Person Numbers (170 or greater).

All new heads and wives/"wives", both immigrant and core sample, were asked in what year they came to the U.S. to stay. These questions served as screening questions for the immigrant sample and were useful in development of 1997 sample weights for both the core and the immigrants. These variables are located on the 1997 portion of the cross-year individual file.

2. Slashing the Census Sample and Resurrection of Dropped Sample

In order to keep data collection costs down, we were forced to cut the PSID sample for 1997. Several scenarios were discussed, but in the end, the Census (SEO) subsample was selected for reduction by two thirds. The rule was that all current Census sample families sharing the same original 1968 Interview (ID) Number would either be kept or cut.

At the eleventh hour, we received funding that enabled us to reinstate some of the dropped families because of the intense interest in children and child development for the 1997 wave. The families to be reinstated were headed by a Black individual and contained at least one child aged 12 or under in 1996. If other related families (those sharing the same 1968 ID Number) had been dropped, they were not reinstated unless they, too, met the rule about race of head and presence of a child in the proper age group.

3. Background Questions

Because of the new immigrant sample for 1997, many of the PSID's traditional background items (from questionnaire Sections K and L) were re-asked of all heads and wives/"wives", and some new items were added.

We collected information about where parents were born, where they grew up, their education, and their occupations. We asked race, ethnicity and occupation of the first full-time, regular job for head and wife/"wife". The information about where parents were born is completely new, but we have always asked where the head's parents grew up, their education, and the father's occupation. These items had not been regularly asked of wives/"wives", however.

We did not re-ask heads and wives/"wives" who retained their head/wife/"wife" status from 1996 about the number of their siblings and whether those siblings were living or deceased, nor did we repeat the education series and religious preference questions for them.

Heads and wives/"wives" from the new immigrant sample were asked questions about their immigration experience (questionnaire Section M, ER11901-ER11982 for heads and ER11983-ER12064 for wives/"wives").

4. Child Support Questions and Child Development Supplement

An extensive set of questions was asked about both incoming and outgoing child support, custody and visitation. These items were asked for all children in the family under age 18, and parallel information was collected for all parents in the family who had children living elsewhere. These data are not available as part of the PSID main public release I files (and will not be part of the main final release files), but instead are destined to become a separate, supplemental file.

The Child Development Supplement is an intensive study of 3,500 PSID children aged 12 or under. It includes assessments of the children's cognitive, behavioral and health status; caregiver time inputs; time use in school; and other measures of such things as school and neighborhood resources and the home learning environment.

Data are available there for linking with PSID main data.

II. CHARACTERISTICS

A. Files and Format

The .zip package for the 1997 family consists of the 1997 public release I family data, program statement files, a codebook, and a report of variable means. Their contents are described in subsequent, more detailed paragraphs.

File name	Contents
FAM1997ER.TXT	1997 raw family data
FAM1997ER.SAS	SAS statements for family variables
FAM1997ER.SPS	SPSS statements for family variables
FAM1997ER.DO	STATA statements for family variables
FAM1997ER_CODEBOOK_public.PDF	Codebook for family variables

1. The data file

The data are in raw ASCII form. Refer to the corresponding .SAS, .SPS, or .DO file for record format layout information, variable names, and variable labels.

The 1997 public release I family data file contains one record for each family interviewed in 1997 and includes all family-level variables collected in 1997. The 1997 public release I family variables are in the range ER10001-ER12224.

2. The data definition statement files

FAM1997ER.SAS, FAM1997ER.SPS, and FAM1997ER.DO contain SAS, SPSS, and STATA data definition statements that provide variable names, locations, and variable labels. These statements are NOT intended to represent complete and full programs for the respective statistical program packages to run extracts, analysis, etc. You must provide all other SAS, SPSS, or STATA statements needed to complete a program.

3. The codebook and means files

The .zip package includes the 1996 Public Release Family File Codebook, FAM1996ER_CODEBOOK.PDF. Unweighted frequencies appear to the left of the codeframes.

No annotation of differences in codes for the public release I files has been done, although this should not be a problem for most variables. The primary code scheme difference is the time unit codes for dollar amounts; a uniform code for all such was used from 1996 forward, but the 1994 and 1995 waves are idiosyncratic. Check questionnaires thoroughly!

B. Variable Names, Positions and Generated Variables

All variable names for both 1997 public release I family file are prefaced with "ER" rather than "V" to assist both analysts and study staff in determining whether reference is to the public release I file.

Variables not included in the public release I file for which component items are available include:

- * imputed and annualized mortgage and rent payments,
- * imputed and annualized food costs,
- * poverty thresholds,
- * imputed and annualized work hours,
- * imputed and annualized unemployment, etc., hours, and
- * education of Head and Wife/"Wife".

Most background items have not been asked about all Heads and Wives/"Wives" each and every year. We ask the questions for new Heads and new Wives/"Wives" only. During processing, we have traditionally "brought forward" background information from previous waves for Heads or Wives/"Wives" who are the same persons as in the prior year. In every wave, each set of background variables is preceded by a variable indicating whether data needed to be brought forward.

C. Documentation and Codes

1. Documentation

In addition to this file, README.TXT, the .zip package includes the 1996 Public Release I Family File Codebook, FAM1996ER_CODEBOOK.PDF. Unweighted frequencies appear to the left of the codeframes.

2. 1997 Questionnaire

We have not produced the traditional codebooks for the public release I files. Our website includes an HTML depiction of the computer-assisted interviewing application for 1997. This replaces the PDF-format questionnaires we have provided for the past several years. Use the SAS and SPSS data definition statements to match variables with questions in the CAI application. The HTM application contains codes for most data items.

3. Code Categories

The HTM version of the 1997 interviewing application provides most codes for the family data. In general, codes follow our traditional structure, although "don't know" responses are now largely distinguished from other missing data responses. Generally, code 8 (or 98 or 998, etc.) represents "don't know" and code 9 (or 99 or 999, etc.) represents other missing data or a refusal. Inappropriate questions are padded with zeroes. If a variable contains a code value that is neither included in the application nor one of the "zero", "eight" or "nine" codes just mentioned, assume missing data for that value.

D. Bringing forward Background Information for Head and Wife/"Wife"

In Release 5, we have brought the Head and Wife/"Wife" background information forward.

E. Problem Variables, Missing Variables

1. Family File Problem Variables

In general, we are very pleased with our public release I data quality for 1997. We have conquered past problems with data extraction that had heretofore caused unfortunate gaps in the data items.

Since for the most part these data have not yet been edited, outrageous answers given by respondents or entered by interviewers remain. However, the user may be assured that the public release I data are a faithful representation of collected information.

Some family variables contain wildly improbable or incorrect values:

ER10032 A16 ACTUAL 3 ROOMS has values of 50 and 97.

ER10084 B2 YEAR RETIRED (HD-R). Almost all cases that should have been asked this question have values of 9999. This was due to an error in the data collection instrument, so unfortunately the information is simply missing for 1997.

ER10566 D2 YEAR RETIRED (WF-R). See ER10084 above.

Values for variables referring to nonworking heads' third and fourth extra jobs (ER10519-ER10562) contain only zeroes because in this wave no such head had more than two extra jobs. Similarly, working wives/"wives" reported no fourth extra job (ER10785-ER10806), and nonworking wives/"wives" reported no second through fourth extra jobs (ER10979-ER11044).

All ending wage rates and time units for second main jobs of nonworking heads/wives/"wives" are included in ER10419-ER10420 and ER10901-ER10902, not ER10417-ER10418 and ER10899-ER10900. These latter two sets of variables should not have been included in the data at all and contain zeroes for everyone.

F. Comparability with 1993 Final-Release File and Earlier Final-Release Files

Beginning with the 1993 wave, the data were collected using CATI (Computer Assisted Telephone Interviewing). This meant that information about each question was collected electronically by the interviewer and, in effect, was coded at the time of data collection. Conversion to standardized units of measurement, formerly performed as part of our coding operation, has not yet been done. As a result, the data in the 1994-1997 public release I and the 1993 final-release family files much more directly resemble the answers to questionnaire questions than the 1992 and early years' data did. For example, instead of one variable indicating monthly rental expense, rent costs now exist as two variables: one for the dollar amount and one for the time unit. Typical responses are \$500 per month, \$100 per week, or don't know how much per month. In addition, the 1993 final-release file includes an annualized rent amount, free of missing data, and an accuracy variable indicating whether an imputation was necessary to arrive at the annualization. This is more in line with 1992 and earlier procedures. But the public release I files have not yet been processed through the imputation algorithm. Therefore, many of the 1994 through 1997 public release I variables are not directly equivalent to variables from 1992 and earlier waves.

Unlike data collected through 1993, the public release I family data have NOT been cleaned with our manual economic edit process, so you must convert these kinds of amounts into some sort of consistent unit for inter-case comparison and make decisions about handling missing data. In addition, we expect that values for quite a few cases will change when we do perform economic edit operations. For example, all time unit questions include an "other" code, as well as options for missing data; amounts associated with these "other" codes will be recoded from interviewers' notes when the data are cleaned.

You can create cross-year files with variables needed for your particular analysis by merging the necessary information from the appropriate family files. The needed family identification numbers appear both in the final-release and public release I cross-year individual files. Since they do not appear on any family data files from 1993 forward, you can obtain them from the Head's record in the 1968-1997 public release I cross-year individual file (see Section G, step 1, above). Detailed instructions for the process of creating the cross-year files are included in the 1993 family documentation volume and are also available at our Internet site; thus, they are not repeated here.

G. Notes on Non-Questionnaire Variables in 1997

The 1997 family file includes quite a number of variables that are not lifted directly from the questionnaire. Some are income summaries and are located at the end of the data file. These income amounts do not quite match the 1997 'Income Plus' files, however, as we recalculated for the public release I data and included new immigrant cases (which were excluded from the 'Income Plus' file).

Each of these income components has an associated accuracy code that indicates whether imputations were made in order to calculate totals. Accuracy variable codes are 0 for no imputation and 1 if an imputation was needed.

The state of residence is included on the 1997 public release I family file. This variable contains FIPS values. The code for FIPS states is located in Appendix 1 of the codebook for the 'Income Plus' files. See our Supplemental Files web page.

All sample persons in the 1997 family aged 12 or under were considered eligible for the CDS and are coded 1 on ER33418. All other cases contain values of 5 or 0. The next two paragraphs explain code 5.

The maximum number of CDS interviews per family was limited to two. Thus, in a family with three or more eligible children, two were randomly selected for interview. ER33419 contains values of 1 for those children who were selected, and values of 5 for those who were eligible but not selected. All other individuals have values of 0 here.

Note that around 80 children are coded 5 on ER33418, whether eligible. All these children have values of 1 for ER33419, whether selected. These individuals were thought to be sample members during main PSID interviewing

and were selected and interviewed for the CDS, but during the family composition editing process we discovered that they indeed were not sample. Their interviews were kept by the CDS, but ER33418 indicates their retroactively discovered ineligibility.

The third variable in this series, ER33420, contains codes indicating the reason for CDS nonresponse for those who were selected. We list the codes below.

- 1 Interview with sample child
 - 2 Interview with nonsample child
 - 3 Nonresponse--refused
 - 4 Nonresponse--lost
 - 5 Nonresponse--incapacitated, permanent condition
 - 6 Nonresponse--deceased
 - 7 Nonresponse--could not contact primary caregiver for actual interview
 - 8 Nonresponse--area too dangerous for interviewer
 - 9 Nonresponse--language barrier

 - 20 Nonresponse--study ended before interview could be completed

 - 96 Office Error--interviewer reports interview taken, but no primary caregiver interview logged
 - 97 Office Error--eligible child who should have been selected for interview
 - 98 Office Error--outside continental U.S., should have been called
 - 0 Inap.; ineligible and no interview taken; eligible but not selected
-