1997 Panel Study of Income Dynamics
Analysis Weights for Sample Families and Individuals

Steven G. Heeringa, Judith H. Connor
June 1, 1999

The Panel Study of Income Dynamics (PSID) underwent several important design changes in 1997. Leading these changes was a roughly 1/3 reduction in the number of PSID Core[1] families that will be eligible for continuous longitudinal data collection. A second important change to the 1997 PSID was the addition of a nationally representative sample of immigrant households and individuals that would not be eligible for PSID under the original 1968 sample recruitment and sample family "following rules". The 1997 data collection year also began the transition to every second year data collection for PSID. Finally, the 1997 PSID data collection included a special supplemental study of children age 0-12 in PSID Core and Immigrant Supplement families.

The reduction in the size of the core panel of PSID families and the introduction of the supplemental sample of post-1968 immigrant families requires compensating adjustments to the weights that are supplied for use in descriptive analysis of the public use data sets. This technical document describes the development of these analysis weight variables and their use in analysis of 1997 PSID data. Special analysis weights for the 1997 Child Supplement data are described in a separate document.

1. 1997 PSID Reduced Core Sample of Families and Individuals.

The PSID is based on an innovative design for a dynamic longitudinal sampling of U.S. families (Hill, 1992)[2]. The original national probability sample of approximately 4800 families was first selected and interviewed for the PSID panel in 1968. Each year from 1969 through 1996, original 1968 PSID sample families and any split-off families (i.e. separations and divorces, children leaving home) were recontacted for the annual PSID interview. The PSID panel lost sample families each year due to attrition (nonresponse and mortality) but many new families were added as a result of the dynamic recruitment process. By 1996, many original families had split multiple times and many of these were beginning to split again as the second generation of children and grandchildren moved out to establish independent families. The net result of the offsetting processes of attrition and dynamic sampling produced continued positive growth in the numbers of PSID families to be interviewed each year. Continued year-to-year growth in PSID family sample size placed increasing strain on the program's data collection funding. By the mid-1990's it was clear that cost-saving changes were needed. At the same

1 The term PSID "Core" is used here to refer to the longitudinal panel of U.S. families that was first sampled and interviewed in the 1968 PSID data collection as well as sample families that split-off from the original 1968 families in later years.
2 Hill, M. (1992) The *Panel Study of Income Dynamics – A User's Guide*. Sage Publications.

time there was a growing consensus on the PSID Board and in the user community that whatever changes were made, PSID should remain a nationally representative sample of the U.S. population. The PSID Board reviewed a number of options and decided to take three steps in 1997: 1) reduce the sample size of the core panel; 2) add the Immigrant Supplement and 3) increase the interview periodicity from one to two years. The reduction in core panel size and the shift to a two-year data collection cycle were seen as necessary steps to bring the PSID into long-term equilibrium with projected funding levels for this longitudinal survey program. The addition of the Immigrant Supplement was a step toward improving PSID representation of all U.S. families including those formed by immigrants who moved to the United States after 1968.

The PSID Board at its March 22, 1996 meeting reviewed five options for reducing the size of the 1997 PSID Core panel for 1997 data collection and made several recommendations to the PSID program staff. Specifically, the Board ruled out subsampling options that would break the intergenerational ties in the core panel (e.g. retain young and middle-age working families and subsample their retired parents). Therefore, only subsampling alternatives that retain (or set aside) entire linkages to a 1968 PSID sample family are considered. The subsampling required to identify the reduced 1997 Core panel sample of families is performed at the level of the 1968 household ID. A useful analogy is to view the subsampling at the level of the root or trunk of the complete 1968 PSID family Atrees≅ with the retention of complete trees-- the root and trunk and all the branches. The Board was also clear in recommending a subsampling option that maximized the retention of 1968 PSID family trees that originate with the 1968 SRC national sample of U.S. families. Upon further review the Board decided on a sampling plan for the 1997 Core panel reduction that retained SRC 1968 sample family trees with certainty and subselected 1968 Survey of Economic Opportunity (SEO) black sample family trees with probability proportionate their 1968 family weight. No non-black family trees from the SEO component of the 1968 PSID sample were retained.

The selection of the Immigrant Supplement is independent of the subsampling of 1968 PSID family trees. A description of the sample design for the 1997 PSID Immigrant Supplement is provided in a separate document.

2. Overview of the 1997 PSID Weight Calculation Procedure.

The approach used to develop the 1997 PSID household and individual weights involved three steps:

Step 1: Beginning with the 1996 Core longitudinal analysis weights for families and individuals, adjustments to these weights were calculated to compensate for changes in sample selection probabilities that result from the subsampling of 1968 black SEO household trees. A base sample selection weight for the Immigrant Supplement families and individuals was also developed.
Step 2: 1997 PSID Core and Immigrant Supplement sample weights were scaled to a common

population reference so that these two samples may be combined for pooled weighted analysis of household and individual data that is representative of the full U.S. populations. The chosen scaling is to express individual weights in 1000s of population units—families for the family weight and persons for the individual weight.

Step 3: The scaled population weights from Step 2 were poststratified to 1997 CPS household and individual totals (in 1000s) for major demographic groups and geographic subclasses of the survey population.

The three-step procedure for the 1997 PSID family weight calculation is described in detail in Sections 3-5 below. Section 6 describes the parallel computation of the 1997 PSID individual weights.

3. Adjusting 1996 PSID Family weights for 1997 Changes in Selection Probabilities.

The 1996 PSID family weights which account for the original joint selection probabilities of the SRC and SEO household samples and for panel attrition due to nonresponse and mortality are the starting point for the construction of the 1997 PSID family weights. To calculate weights for the 1997 PSID, all sample families are assigned to one of six computation strata. Table 1 defines these six computation strata and the stratum-specific weighting adjustments that were required.

Stratum 1: This 1997 weight calculation stratum includes all families linked to the SRC component of the 1968 PSID sample that did not reside in geographic areas that were eligible for the SEO low-income sample. The calculation of the original sample selection weights for these families did not include a probability of also being selected to the SEO sample. Since all SRC families are retained in the 1997 core reduction, no adjustment to the base weight of families in this stratum is required.

Strata 2, 3, and 4: The pre-1997 PSID family weights for calculation strata 2, 3, and 4 families incorporate a joint probability of being selected for both the SEO and the SRC components of the 1968 sample. (The 1968 PSID household resided in a geographic area that was eligible for the special SEO low income sample). The subsampling of black SEO families and the exclusion of non-black SEO families for the 1997 PSID alters these original selection probabilties. The selection factor component of the 1996 family weight must be recalculated to decrease the SEO probability factor for black families in Stratum 2 and 4 and to eliminate the SEO probability factor for non-black families in Stratum 3.

Stratum 5: 1997 PSID Immigrant Supplement sample families were initially selected with equal probability. During the field period, sample replicates of area segments with higher expected prevalence of immigrant households were oversampled to increase the efficiency of the household screening. Each of the interviewed families in the 1997 Immigrant Supplement was

assigned an initial base weight value that reflected the probability of selection and screening for the area segment in which they resided. These base weights for immigrant supplement households were scaled to U.S. population values (See Section 4 below).

Stratum 6: Stratum 6 includes families that were included as a special supplement for the 1997 PSID Child Development Sample. These families were black low-income families from the SEO sample, which had children age 0-12. These families were not retained in the subsampling of the PSID core but were included as a special supplement to the 1997 Child Development Study. These families are not considered part of the 1997 PSID Core longitudinal sample and are assigned a 1997 PSID family weight value of zero. These families are assigned a population weight in the special series of analysis weights that has been prepared for the Child Supplement data sets.

In addition to families in Stratum 6, some families in Strata 1, 2, 3, and 4 could also have 1997 PSID family weights equal to zero. Stratum 1-4 families with a family weight of zero have large amounts of missing data in the longitudinal series of annual PSID observations. Most of these families were reintroduced to the PSID panel through special recontact efforts conducted in 1993 and 1994. Work is now in progress to develop a cross-sectional family weight that will enable analysts to include these families in cross-sectional analysis of the PSID data.

Table 1: 1997 PSID Weight Calculation Strata

Stratum	Adjustment to 1996 PSID Family Weight
1. SRC sample. No overlap with the SEO sample domain.	No adjustment; 1997 PSID preliminary family weight is equal to the 1996 PSID family weight.
2. Black SRC sample overlapped with the SEO sample domain.	The 1997 PSID preliminary family weight = the 1996 PSID family weight * the maximum of [6.23/1968 PSID family weight or 1.0] .
3. Non-black SRC sample overlapped with the SEO sample domain.	The 1997 PSID preliminary family weight = the 1996 PSID family weight * (15272/(1968 PSID family weight*400)).
4. Black SEO sample. Overlapped with the SRC sample domain.	The 1997 PSID preliminary family weight = the 1996 PSID family weight * the maximum of [6.23/1968 PSID family weight or 1.0].
5. 1997 PSID Immigrant Sample	The 1997 PSID preliminary family weight = 16.198* Immigrant household weight (See Section 4 below).
6. Supplemental sample for low-income black families with children age 0-12.	The 1997 PSID preliminary family weight = 0.

4. Scaling the Preliminary 1997 PSID Weights to 1997 U.S. Population Totals.

1997 PSID weights for families and individuals are scaled to represent the corresponding United States population estimates (in 1000s). The first step in this scaling process is to apportion the total population weight between the domain of families and individuals who were eligible for the 1968 PSID Core panel and those who were only eligible for inclusion under the 1997 Immigrant Supplement. Based on data from the 1997 CPS, an estimated 7.5 percent of U.S. households have immigrated to the United States after 1968. Therefore the Immigrant Supplement sample represents about 7,500,000 of the 100,000,000 U.S. households and the 1968-based PSID sample represents the remaining 92,500,000 households. The 1997 Immigrant Supplement sample includes 441 respondent families. Although 441 immigrant households were interviewed, the sum of the initial base weights for these households is 463. To adjust the preliminary PSID weights to U.S. totals in thousands (based on the 1997 CPS), the Immigrant Sample household base weights were multiplied by a factor equal to 7,500/463 = 16.198.

For the second domain, the sum of the recalculated 1997 Preliminary family weights for Core sample families is 132,298. To scale these family weights to the desired U.S. population estimate of 92,500 (1000s), the 1997 Preliminary weight for each family in the 1968-based core sample was multiplied by a factor of 92,500/132,998. After this scaling step, the sum of the 1997 PSID family weights equals the estimated count of U.S. households of 100,000 (1000s).

5. Poststratification of 1997 PSID Family weighted Counts to 1997 CPS Estimates.

Prior to 1997, PSID analysis weights do not include poststratification adjustments to control PSID weighted estimates to other sources of U.S. population data. The decision to not include poststratification adjustments in the pre-1997 PSID weight variables was governed in large part by the fact that the panel represented only U.S. families that were eligible for the 1968 sample selection or as a result of PSID "following rules". With the addition of the 1997 Immigrant Supplement, the PSID became nationally representative of the entire U.S. population including post-1968 immigrants who had not been represented previously. The extensive recalculation of the basic analysis weights that was necessitated by the 1997 sample design changes presented an added motive for benchmarking key marginal distributions of the PSID sample to current CPS estimates for U.S. households. A decision was made to poststratify 1997 PSID weighted estimates to CPS totals for major demographic groups and geographic regions.

Weighted tabulations of PSID household characteristics were computed using the preliminary 1997 weights described in Table 1 above. These preliminary weighted estimates were compared to weighted estimates from the March 1997 CPS. Household characteristics that were compared to CPS estimates included age, race and education of the household head,

a four-category descriptor of household type (single parent, two parent, single person, and other) and geographic characteristics including metropolitan status (MSA or Non-MSA) and Census Region (Northeast, Midwest, South, or West).

Table 2 shows the result of these original comparisons between the PSID preliminary weighted estimates and the CPS estimates. Some characteristics included in this comparison turn out to not be useful choices as poststratification variables due to inconsistency or possible ambiguity in the PSID and CPS definitions (i.e. gender of household head, household type)[3]. For those characteristics where the survey definitions are expected to be consistent in both surveys (race of head, MSA status, Census Region) the PSID weighted estimates correspond reasonably well to the 1997 CPS proportions. However, the PSID percent of black heads of families is higher than the corresponding CPS percentage (weighted estimates are 15.5 percent of families for PSID and 12.0 percent for CPS). Likewise, the weighted proportions of PSID families from the Midwest region and MSA domain are greater than the proportions computed from the 1997 CPS household data. As noted above, no external benchmarking corrections were applied to the PSID analysis weights during the period 1968 to 1996. Over this period, PSID analysis weights were adjusted for attrition due to nonresponse and mortality. One possible explanation for the small differences that are observed in the 1997 comparison of PSID and CPS weighted distributions may be small but cumulative overcorrections or undercorrections that have occurred for some characteristics in this adjustment process. For example, household survey response rates in the Midwest region are often higher than those obtained in other Census regions. Nonresponse adjustments that do not systematically control for region and the regional variation in factors associated with nonresponse will tend to bias weighted sample representation toward regions with higher response rates.

To benchmark weighted 1997 PSID sample estimates to CPS estimates for key demographic and geographic domains, the 1997 PSID family weights were poststratified so that final weighted estimates match 1997 CPS proportions for cells formed by crossing race of head (Black/Non-Black) by MSA status (MSA/Non-MSA) by Census Region (Northeast / Midwest/ South / West). Table 2 also shows the comparison of CPS and PSID weighted estimates for household proportions after poststratification by race of head, MSA/non-MSA status and Census Region.

[3] The definition of household head for PSID differs from the CPS definition of a reference person. In the PSID, the household head is always male, unless there is no husband at that household or the husband is too ill to answer questions. The CPS uses a reference person who may be either male or female. Therefore, the percentages for sex of head of families are very different for PSID and CPS.

Table 2. Comparison of 1997 PSID and CPS Weighted Estimates of Household Characteristics

		1997 PSID			1997CPS	Ratio
		[1]	[2]	[3]	[4]	[5]=[3]/[4]
		Un-weighted %	Preliminary Weighted %	Post-strat Weighted %	%	
Head of Household						
Race						
	Black	30.6	15.5	11.9	12.0	0.99
	Non-black	69.4	84.5	88.1	88.0	1.00
Sex						
	Male	70.1	69.3	70.0	59.5	1.18
	Female	29.9	30.7	30.0	40.5	0.74
Age						
	18-30	17.6	14.5	14.5	15.5	0.94
	31-44	41.7	34.7	34.3	32.2	1.07
	45-64	26.5	30.8	30.8	31.0	0.99
	65+	14.2	20.0	20.4	21.2	0.96
Education						
	Less than high school (<12)	22.3	20.4	20.6	18.4	1.12
	High school (12)	33.8	32.2	32.2	31.8	1.01
	College (13-16)	35.4	36.9	36.6	41.3	0.89
	Above college (17+)	8.6	10.5	10.6	8.5	1.25
Household Type						
	Single parent	19.5	13.6	12.9	11.0	1.17
	Two parents	32.5	26.9	26.8	26.0	1.03
	Single person	21.7	30.6	30.4	30.4	1.00
	Other	26.3	28.9	29.9	32.5	0.92
MSA vs. NON-MSA						
	MSA	72.5	72.8	66.2	66.2	1.00
	NON-MSA	27.5	27.2	33.8	33.8	1.00
Census Region						
	Northeast	14.9	19.5	19.6	19.5	1.01
	Midwest	25.5	26.5	23.7	23.7	1.00
	South	41.7	33.3	35.3	35.3	1.00
	West	17.9	20.7	21.4	21.4	1.00

6. 1997 PSID Individual Weights

The preceding sections describe the development of 1997 PSID analysis weights for analysis of household characteristics. The development of the 1997 weights for the analysis of PSID data on individuals closely parallels the three-step method used to construct the 1997 family weights.

Step 1 was repeated exactly as shown in Table 1 with the exception that the 1996 PSID individual weight is used in place of the 1996 PSID family weight. Newborns in the 1997 PSID sample were assigned a 1997 individual weight equal to the average of the individual weights of the all other individuals in the sample household. "Re-appears" (persons not interviewed in the 1996 but located and interviewed in 1997) do not have a value for the 1996 individual weight. A reference weight value equal to the individual weight from the latest year in which they were interviewed was assigned to these cases.

Step 2 in the individual weight calculation process also scaled the weighted totals for sample individuals in Core and Immigrant Supplement sample families to the 1997 estimate of total U.S. population (in 1000s).

The third and final step poststratified the individual weights to 1997 CPS totals in thousands for cells defined by Race (Black/Non-Black), metropolitan status (MSA/Non-MSA) and Census Region (Northeast/Midwest/South/West). Table 3 compares 1997 PSID and CPS weighted estimates of population proportions for a number of individual demographic characteristics. The two series of estimates of multinomial proportions for race, MSA/non-MSA and Census Region will match exactly due to the post-stratification. The estimated distributions by gender and age match quite well but are not explicitly controlled in the poststratification of the individual weights.

Table. 3. Comparison of 1997 PSID and CPS Weighted Estimates of Individual Characteristics

		1997 PSID			1997 CPS	Ratio
		[1] Unweighted	[2] Preliminary Weighted	[3] Post-strat.	[4] Weighted	[5]=[3]/[4]
		%	%	%	%	%
Age						
	0-17	34.5	25.1	24.8	26.7	0.93
	18-30	18.7	17.4	17.4	18.1	0.96
	31-44	24.0	23.3	23.2	22.8	1.02
	45-64	15.1	21.7	21.8	20.4	1.07
	65+	7.6	12.5	12.7	11.9	1.07
Sex						
	Male	48.0	47.8	47.8	49.0	0.98
	Female	52.0	52.2	52.2	51.0	1.02
Race						
	Black	33.5	15.5	12.8	12.8	1.00
	Non-Black	66.5	84.5	87.2	87.2	1.00
MSA						
	MSA	73.2	71.6	66.6	66.6	1.00
	Non-MSA	26.8	28.4	33.4	33.3	1.00
Region						
	Northeast	14.6	19.7	19.3	19.3	1.00
	Midwest	25.3	27.2	23.3	23.3	1.00
	South	42.2	32.7	35.0	35.0	1.00
	West	18.0	20.4	22.4	22.4	1.00