C. Computing the SIPP Sampling Weights

This appendix supplements the discussion in Chapter 8 (Using Sampling Weights on SIPP Files) with more detailed information about how the Core Wave File person-level weight WPFINWGT and the Longitudinal Weight File person-level weights LGTCYxWT (calendar year weight for year x) and LGTPNWTx (panel weight based on the sample universe in waves 1 through w(x)[1]) are computed[2], it is intended as a reference for users who require a comprehensive description of how the sampling weights are computed for the 2004 Panel[3].

Sections 1 and 2 of this appendix discuss the algorithms that are used to compute the final Core Wave File person-level weights WPFINWGT, with the first section discussing the Wave 1 weights and the second section discussing the Wave 2+ weights. The third section discusses the algorithm that computes the final Longitudinal Weight File person-level weights LGTCYxWT and LGTPNWTx.

Wave 1 Weights

The final weights used in deriving estimates consist of the product of four factors: the base weight, the duplication control factor, the household noninterview adjustment factor, and the second-stage adjustment factor.

Base Weight (BW)

The primary component of the sampling weight is the base weight. The base weight for any sampled person or sampled household is the reciprocal of the probability under the sample design of that person or household being selected. If there was full response and if there were no change in the sampled population before interviewing, then the summation of base weights for a particular subgroup (e.g., Hispanics in the Southwest) is an unbiased estimator of the total U.S.

---

[1] The 2004 Panel will have four panel weights when it is complete: LGTPNWT1 (waves 1-4), LGTPNWT2 (waves 1-7), LGTPNWT3 (waves 1-10), and LGTPNWT4 (waves 1-12).

[2] The remaining weights given in Table 8-2 (WHFNWGT, WFFINWGT, and WSFINWGT) are derived directly from the basic person-level weight WPFINWGT. This derivation is discussed in the "How Weights are Constructed" subsection of Chapter 8.

[3] A detailed description about how sampling weights were computed for panels prior to 2004 can be found in *SIPP Users' Guide*, 3rd Ed. [U.S. Census Bureau, 2001]

population within that subgroup. In simplified terms, a base weight of 1,000 assigned to a sampled person means that the sampled person "represents" 1,000 persons in the U.S. population. The base weight for a household and the base weight for a person within a household is the same, since every person within a sampled household is automatically selected (i.e., selected with a conditional probability of 1 given household selection).

Duplication Control Factors (DCF)

The duplication control factor, an integer value between 1 and 4 inclusive, is applied to the base weights of specified households to account for subsampling done in clusters of housing units selected at the last stage of sample selection. These clusters typically contain an unmanageable number of housing units. When this occurs, a sampling fraction, $1/N$, is determined by selecting a value of N such that the number of sample households in the cluster is reduced to a manageable size. After this is done, a duplication control factor of N or 4, whichever is smaller, is included as a weighting factor for sampled housing units in the cluster.

Household Noninterview Adjustment Factors (NAF)

The noninterview adjustment factor is intended to adjust for the presence of Type A noninterview households (households that are not interviewed because the occupants were temporarily absent, no one was home, the occupants refused participation, or the occupants could not be located). Noninterview adjustment factors are computed for each of a set of noninterview cells. These cells are based on 512 cells generated from all possible cross-classifications of the following household characteristics:

- Within-PSU oversampling strata: poverty stratum and nonpoverty stratum;

- Census region;

- Race of reference person: black or nonblack;

- Tenure: owner or renter;

- Residence status: MSA urban, MSA nonurban, NonMSA Census place, or NonMSA not Census place; and

- Household size: one, two, three, or four or more persons.

Any cells with less than 30 interviewed households or with noninterview adjustment factors exceeding 2.0 are collapsed with a neighboring cell. To define cells as neighboring, the Census Bureau uses a sort order and scale values based on estimates of the 1979 poverty rate within the cell. The total number of noninterview cells is less than or equal to 512 after cell collapsing. For

the 2004 Panel, no cells are collapsed over the cross-cells defined by race of reference person, tenure, within-PSU oversampling strata, and Census region.

Within each final noninterview cell *c*, the formula for the noninterview adjustment factor (NAF*c*) is

$$\text{NAF}_c = \frac{\text{sum of BW} * \text{DCF over all sampled households in cell } c}{\text{sum of BW} * \text{DCF over all interviewed households in cell } c} \tag{C-1}$$

The factor is applied to the weight of each interviewed household in the cell; with these noninterview-adjusted weights, the interviewed households in each cell can be seen to "represent" themselves and also the Type A noninterviewed households in the cell.

Wave 1 Second Stage Calibration Adjustments (SSCA)

For the second-stage calibration adjustments, the Census Bureau uses independent population controls provided by the Population Division of the Census Bureau. The Population Division produces SIPP controls for age, sex, race, Hispanic origin, and state. Because the Population Division does not produce family-type controls, the Census Bureau uses tallies of Current Population Survey (CPS) weights for SIPP family-type controls. The CPS weights are calibrated to match population controls provided by the Population Division of the Census Bureau and then a CPS "March type"[4] adjustment is done to equalize the weights of husbands and wives. Because SIPP family-type controls are derived from CPS weights, they are in fact CPS sample estimates.

The primary step in the calibration (or raking) process is the attachment of second-stage calibration adjustment factors to the pre-second-stage weights (BW*DCF*NAF) within particular cells (e.g., male Hispanic 14-year-olds) so that the resulting adjusted weights (BW*DCF*NAF*SSCA) aggregate to the independent population estimates within the cell. The summation of the pre-second-stage weights within any cell is an unbiased estimate (assuming the nonresponse adjustment successfully adjusts for all effects of nonresponse) of the population total for that cell (e.g., the summation of BW*DCF*NAF over all male Hispanic 14-year-olds in the panel is an unbiased estimate of the total number of male Hispanic 14-years-olds in the U.S. population).

The adjusted weights (BW*DCF*NAF*SSCA) give estimates then for these cells that are equal to the independent estimates. This adjustment generally improves the overall precision of all estimates of these cells or any other related survey characteristics that are prevalent in these cells.

The population cells for which adjustments are made to independent estimates are given in Figure C-1 (see pages C-7–C-14). The cells include (as can be seen in figure C-1) age, race, sex,

---

[4] The "March type" adjustment refers to a similar adjustment from the Current Population Survey, Annual Social and Economic Supplement (CPS-ASEC) which conducts interviews mostly in March.

Hispanic origin, state, family relationship, and household type. As noted earlier, the independently derived estimates for these cells are based on estimates from the Population Division of the Census Bureau and on CPS March-type estimates for family type. (The CPS family-type estimates are not the usual CPS monthly estimates. The estimates are specially computed for this purpose by summing the CPS weights within a given cell for all sample units in the relevant CPS sample [also, there are some extra steps to ensure equal numbers of husbands' and wives' and to ensure consistency of family type controls with other age and sex controls from Population Division]).

Outline of the Second Stage Calibration Algorithm

The second stage calibration algorithm uses as its inputs the pre-second-stage weights BW*DCF*NAF computed for each sampled person represented on a completed questionnaire in a SIPP panel[5]. These weights are run through a series of adjustments, which result in a final weight (FNLWGT). This final weight can be written as FNLWGT=SSCA*BW*DCF*NAF, with SSCA (the second stage calibration adjustment) equal to the ratio of the pre-second-stage weight and the final weight after the calibration process is completed.

This algorithm can be segmented into steps[6] as described below:

> Perform the Cell Collapsing Step
> Let $r=0$
> Repeat {
> > Perform the Calibration Step
> > Perform the Spouses' Weight Equalization Step
> > Let $r=r+1$
> > } Until the adjusted weights are within tolerance or $r=40$

The cell collapsing step is run separately for each second stage dimension (see Table C-1). The next section discusses details of the cell collapsing step. After the cell collapsing step, the calibration and spouse equalization steps are run until the adjusted weights are within tolerance, up to 40 times. The adjusted weights are within tolerance if they aggregate to within 500 persons of the control total, in every second-stage adjustment cell (see Figure C-1), after a spouse equalization step. The second section discusses details of the calibration step, and the final section describes the spouses' weight equalization step.

---

[5] Children do not answer any SIPP questionnaires, but any children who are indicated as dependents by a sampled household receive weights in this process.

[6] Separate runs of the calibration algorithm are made for each reference month and each rotation group (a total of 16 calibration runs for each panel wave).

Cell Collapsing

The initial raking (calibration) step is preceded by a cell collapsing step. This step is designed to prevent extreme alterations in the person weights (which will increase variability of the estimators) in any of the raking steps. Each second-stage cell is checked in its sample size: if the sample size is less than 35, then the cell is collapsed with a neighboring cell. The second-stage cells are also checked by computing the ratio adjustment for each of these cells. If the adjustment is less than 0.67 or greater than 4.0, then the cell is collapsed with a neighboring cell.

Ratio adjustment factors are computed for each set of second-stage cells in Figure C-1 before any of the raking steps are performed. The ratio adjustment factor for each cell is equal to the control total divided by the summation of the person weights (as they are at that point in the algorithm) for all sample persons in the cell. If the computed ratio adjustment factor for any cell is less than 0.67 or greater than 4.0, or the sample size for any cell is less than 35, then the cell is collapsed with a neighboring cell within the same dimension. The same process is carried out separately for each dimension. All collapsing of this kind is completed before the first raking step is executed.

When a second stage cell is designated as requiring collapsing during the cell collapsing step, the neighboring cell is chosen through a predetermined mechanism. Neighboring cells are found on the basis of the scale value (which is given for the 2004 Panel in Figure C-1) and the collapsing constraints (which are given in Table C-1 below). The cell satisfying the collapsing constraints with the scale value closest to that of the cell that requires collapsing becomes the neighboring cell used in collapsing.

Table C-1. Second Stage Dimensions and Collapsing Constraints

Dimension	Description	Collapsing Constraints
1	Hispanic Origin, Sex, and Age (16 Cells)	• Do not collapse children (age <15) • Collapse adults (age $\geq$ 15) only within sex and Hispanic origin
2	State and Hispanic Origin (18 Cells)	• Do not collapse between states
3	Family Type and Sex (14 Cells)	• Do not collapse children • Do not collapse adults between sexes
4	State and Race (36 Cells)	• Do not collapse across states
5	State, Sex, and Age (159 Cells)	• Do not collapse across states • Within a state, do not collapse across sex unless all categories of age are collapsed for both males and females
6	Race, Sex, and Age (108 Cells)	• Collapse only within the same sex and race • Do not collapse children with adults

Occasionally, second stage cells that would be collapsed otherwise are not collapsed due to the collapsing constraints. The most common occurrence of this occurs in Dimension 5 where 9 states and the District of Columbia are each represented by a single cell.

The cell-collapsing procedure requires more than one iteration if, after collapsing to the nearest neighbor, the collapsed cells are still too small or show extreme ratio adjustments. New scale values are computed for the collapsed cells, and are used to designate neighboring cells for any further collapsing that is necessary.

Calibration

The most important step in the algorithm is the raking step. The raking step consists of successive ratio adjustments to six sets (dimensions) of second-stage cells (see Figure C-1), with separate control totals. Each ratio adjustment takes all of the person weights (as they are at that point in the algorithm) within particular second-stage cells and multiplies them by a common ratio adjustment factor. The adjustment factor for each cell is equal to the population control for that cell divided by the total of the current person weights for all sample persons in the cell. After each ratio adjustment, the summation of the adjusted person weights within the second-stage cell is equal to the control total for that cell.

At the end of the raking process (also called iterative proportional fitting), each person weight (as it is at that point in the algorithm) has been adjusted so that all person weights aggregate to the appropriate control totals for the cells of each dimension. The adjusted person weights have the property of aggregating within the second-stage cells to each control total while remaining as "close as possible" (in terms of a particular algebraic distance function) to the person weight values at the beginning of the raking step. Thus, the new person weights are consistent with each of the six sets of independent control totals and have been altered as little as possible from the person weights before the step.

Spouse Weight Equalization

This step equalizes husbands' and wives' weights (so that spouses in one family have equal weights). Each spouse receives the average of the husband's weight and the wife's weight.

Figure C-1. Second Stage Cells

Second Stage Cells for Dimension 1: Hispanic Origin, Sex, and Age

Hispanic Origin	Sex	Age (years)	SCALE VALUE
Hispanic	Male	Under 15	1
		15 to 24	25
		25 to 44	27
		45 or more	30
	Female	Under 15	1
		15 to 24	35
		25 to 44	36
		45 or more	38

Hispanic Origin	Sex	Age (years)	SCALE VALUE
Non-Hispanic	Male	Under 15	1
		15 to 24	125
		25 to 44	127
		45 or more	130
	Female	Under 15	1
		15 to 24	135
		25 to 44	136
		45 or more	138

Second Stage Cells for Dimension 2: State and Hispanic Origin

State	Hispanic Origin	SCALE VALUE
New Mexico	Hispanic	1
	Non-Hispanic	2
New Jersey	Hispanic	1
	Non-Hispanic	2
Arizona	Hispanic	1
	Non-Hispanic	2
Illinois	Hispanic	1
	Non-Hispanic	2
Florida	Hispanic	1
	Non-Hispanic	2

State	Hispanic Origin	SCALE VALUE
New York	Hispanic	1
	Non-Hispanic	2
Texas	Hispanic	1
	Non-Hispanic	2
California	Hispanic	1
	Non-Hispanic	2
All Other States	Hispanic	1
	Non-Hispanic	2

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 3: Family Type and Sex

Sex	Family Type			Age (years)	SCALE VALUE
Male	Child			Under 15	1
	Persons in Households that Contain a Primary Family or Subfamily	Husband of Primary Family		15 or more	20
		Male Householder, No Spouse Present			30
		Other Household Members	Husband of Subfamily		22
			Not a Husband		40
	Persons Not in Households Containing a Primary Family or Subfamily	Householder			32
		Not a Householder or Person in Group Quarters			42
Female	Child			Under 15	100
	Persons in Households that Contain a Primary Family or Subfamily	Wife of Primary Family		15 or more	120
		Female Householder, No Spouse Present			130
		Other Household Members	Wife of Subfamily		122
			Not a Wife		140
	Persons Not in Households Containing a Primary Family or Subfamily	Householder			132
		Not a Householder or Person in Group Quarters			142

(figure continues)

Figure C-1. Second Stage Cells (continued)

Second Stage Cells for Dimension 4: State and Race

State	Race	SCALE VALUE	State	Race	SCALE VALUE
Alabama	Not Black Alone	1	New Jersey	Not Black Alone	1
Alabama	Black Alone	2	New Jersey	Black Alone	2
California	Not Black Alone	1	New York	Not Black Alone	1
California	Black Alone	2	New York	Black Alone	2
Florida	Not Black Alone	1	North Carolina	Not Black Alone	1
Florida	Black Alone	2	North Carolina	Black Alone	2
Georgia	Not Black Alone	1	Ohio	Not Black Alone	1
Georgia	Black Alone	2	Ohio	Black Alone	2
Illinois	Not Black Alone	1	Pennsylvania	Not Black Alone	1
Illinois	Black Alone	2	Pennsylvania	Black Alone	2
Louisiana	Not Black Alone	1	South Carolina	Not Black Alone	1
Louisiana	Black Alone	2	South Carolina	Black Alone	2
Maryland	Not Black Alone	1	Texas	Not Black Alone	1
Maryland	Black Alone	2	Texas	Black Alone	2
Michigan	Not Black Alone	1	Virginia	Not Black Alone	1
Michigan	Black Alone	2	Virginia	Black Alone	2
Mississippi	Not Black Alone	1	All Other States	Not Black Alone	1
Mississippi	Black Alone	2	All Other States	Black Alone	2

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 5: State, Sex, and Age

State	Sex	Age (years)	SCALE VALUE	State	Sex	Age (years)	SCALE VALUE
Alabama		Under 15	1	Hawaii			1
		15 to 44	10	Idaho	Male		1
		45 or more	11		Female		2
Alaska			1	Illinois	Male	Under 15	1
Arizona		Under 15	1			15 to 44	10
		15 to 44	10			45 or more	11
		45 or more	11		Female	Under 15	31
Arkansas	Male		1			15 to 44	40
	Female		2			45 or more	41
California	Male	Under 15	1	Indiana	Male	Under 15	1
		15 to 44	10			15 to 44	10
		45 or more	11			45 or more	11
	Female	Under 15	31		Female	Under 15	31
		15 to 44	40			15 to 44	40
		45 or more	41			45 or more	41
Colorado		Under 15	1	Iowa	Male		1
		15 to 44	10		Female		2
		45 or more	11	Kansas	Male		1
Connecticut		Under 15	1		Female		2
		15 to 44	10	Kentucky		Under 15	1
		45 or more	11			15 to 44	10
Delaware			1			45 or more	11
District of Columbia			1	Louisiana		Under 15	1
Florida	Male	Under 15	1			15 to 44	10
		15 to 44	10			45 or more	11
		45 or more	11	Maine			1
	Female	Under 15	31	Maryland		Under 15	1
		15 to 44	40			15 to 44	10
		45 or more	41			45 or more	11
Georgia	Male	Under 15	1	Massachusetts		Under 15	1
		15 to 44	10			15 to 44	10
		45 or more	11			45 or more	11
	Female	Under 15	31				
		15 to 44	40				
		45 or more	41				

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 5: State, Sex, and Age *(continued)*

State	Sex	Age (years)	SCALE VALUE
Michigan	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
Minnesota		Under 15	1
		15 to 44	10
		45 or more	11
Mississippi		Under 15	1
		15 to 44	10
		45 or more	11
Missouri		Under 15	1
		15 to 44	10
		45 or more	11
Montana			1
Nebraska	Male		1
	Female		2
Nevada	Male		1
	Female		2
New Hampshire			1
New Jersey	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
New Mexico	Male		1
	Female		2
New York	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41

State	Sex	Age (years)	SCALE VALUE
North Carolina	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
North Dakota			1
Ohio	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
Oklahoma		Under 15	1
		15 to 44	10
		45 or more	11
Oregon		Under 15	1
		15 to 44	10
		45 or more	11
Pennsylvania	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
Rhode Island			1
South Carolina		Under 15	1
		15 to 44	10
		45 or more	11
South Dakota			1
Tennessee		Under 15	1
		15 to 44	10
		45 or more	11

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 5: State, Sex, and Age *(continued)*

State	Sex	Age (years)	SCALE VALUE
Texas	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41
Utah	Male		1
	Female		2
Vermont			1
Virginia	Male	Under 15	1
		15 to 44	10
		45 or more	11
	Female	Under 15	31
		15 to 44	40
		45 or more	41

State	Sex	Age (years)	SCALE VALUE
Washington		Under 15	1
		15 to 44	10
		45 or more	11
West Virginia	Male		1
	Female		2
Wisconsin		Under 15	1
		15 to 44	10
		45 or more	11
Wyoming			1

Second Stage Cells for Dimension 6: Race, Sex, and Age

Race	Sex	Age (years)	SCALE VALUE
White Alone	Male	Under 1	601
		1	603
		2	607
		3	608
		4	611
		5	612
		6	615
		7	616
		8	620
		9	622
		10 to 11	627
		12 to 13	632
		14	633
		15	15

Race	Sex	Age (years)	SCALE VALUE
White Alone	Female	Under 1	701
		1	703
		2	707
		3	708
		4	711
		5	712
		6	715
		7	716
		8	720
		9	722
		10 to 11	727
		12 to 13	732
		14	733
		15	115

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 6: Race, Sex, and Age *(continued)*

Race	Sex	Age (years)	SCALE VALUE	Race	Sex	Age (years)	SCALE VALUE
White Alone	Male	16 to 17	16	White Alone	Female	16 to 17	116
		18 to 19	18			18 to 19	118
		20 to 21	27			20 to 21	127
		22 to 24	29			22 to 24	129
		25 to 29	47			25 to 29	147
		30 to 34	49			30 to 34	149
		35 to 39	57			35 to 39	157
		40 to 44	59			40 to 44	159
		45 to 49	63			45 to 49	163
		50 to 54	65			50 to 54	165
		55 to 59	73			55 to 59	173
		60 to 62	75			60 to 62	175
		63 to 64	76			63 to 64	176
		65 to 69	93			65 to 69	193
		70 to 74	95			70 to 74	195
		75 to 79	103			75 to 79	203
		80 to 84	104			80 to 84	204
		85 or more	106			85 or more	206
Black Alone	Male	Under 5	821	Black Alone	Female	Under 5	921
		5 to 9	823			5 to 9	923
		10 to 14	825			10 to 14	925
		15 to 19	315			15 to 19	415
		20 to 24	325			20 to 24	425
		25 to 29	327			25 to 29	427
		30 to 34	345			30 to 34	445
		35 to 39	347			35 to 39	447
		40 to 44	349			40 to 44	449
		45 to 49	355			45 to 49	455
		50 to 54	357			50 to 54	457
		55 to 64	359			55 to 64	459
		65 or more	365			65 or more	465
Residual Race	Male	Under 5	1021	Residual Race	Female	Under 5	1121
		5 to 9	1023			5 to 9	1123
		10 to 14	1025			10 to 14	1125
		15 to 24	1520			15 to 24	1620

(figure continues)

Figure C-1. Second Stage Cells *(continued)*

Second Stage Cells for Dimension 6: Race, Sex, and Age *(continued)*

Race	Sex	Age (years)	SCALE VALUE	Race	Sex	Age (years)	SCALE VALUE
Residual Race	Male	25 to 34	1525	Residual Race	Female	25 to 34	1625
		35 to 44	1540			35 to 44	1640
		45 to 54	1545			45 to 54	1645
		55 to 64	1550			55 to 64	1650
		65 or more	1565			65 or more	1665

Wave 2+ Weights

The later wave cross-sectional weight is computed separately for each reference month of each wave. This Wave 2+ WPFINWGT has the following factors for people in households whose residents have not changed from Wave 1: an initial weight (IW), a later wave noninterview adjustment (LWNIA), and a second stage calibration adjustment (SSCA). The initial weight is generally equal to the pre-second-stage weight for the Wave 1 household weight (with some exceptions). For households that have had people move into or out of the household after Wave 1, there is an adjustment to the initial weight called the mover's weight (MW). For these people, the cross-sectional weight has as factors the mover's weight, the later wave noninterview adjustment, and the second-stage calibration adjustment. In summary, people in households that do not need mover's adjustments receive the cross-sectional weight WPFINWGT = IW*LWNIA*SSCA, and persons in households that do require a mover's adjustment receive the Wave 2+ final weight WPFINWGT=MW*LWNIA*SSCA.

Wave 2+ Initial Weights

The initial weight is essentially the pre-second-stage Wave 1 weight, that is, IW = BW*DCF*NAF. The second-stage calibration adjustment for the Wave 1 reference months is not included as a factor: the second-stage calibration adjustment is redone using control totals current for the later wave reference months. The initial weight allows the original sample person to represent unsampled persons in the population and persons in households who were not successfully interviewed in Wave 1. The initial weight does not generally change from wave to wave after Wave 1, unless circumstances arise that cause an alteration in the panel sample (such as a cut in the sample for budgetary or other reasons)[7].

---

[7] For example; 53% of sample units were cut from the SIPP 2004 panel beginning in October 2006.

Mover's Weights

Persons in any households that an original sample person enters during later waves, or any people who become part of a Wave 1 sample household during later waves, also becomes part of the sample for those waves. If the original sample person moves away from the household containing those people, the additional people immediately drop from the sample (their in-sample status in any given wave is entirely dependent on the presence of original sample persons in the household). Any of the additional people who were part of the SIPP population in Wave 1 (and therefore could have been sampled) and who become members of households with original sample persons are called associated sample persons. If any of these additional persons were not part of the SIPP population in Wave 1 (because they were out of the country, institutionalized, etc.), then they are called additional sample persons.

Any household that consists of persons who were in the SIPP universe who lived in separate households during the Wave 1 reference period (with at least one of the households sampled in Wave 1) is called an enhanced household. In most cases, an enhanced household consists of original sample persons from a Wave 1 sample household and associated sample persons from a household (or households) not sampled in Wave 1. In a few rare cases, an enhanced household will contain original sample persons from more than one Wave 1 sample household. Those households are rare because the probability of selection of any given household in SIPP is quite small, making the joint probability of a later wave merged household having two or more of its Wave 1 predecessor households selected in Wave 1 even smaller (but the situation does occur in the SIPP panels).

Enhanced households require an adjustment of the Wave 1 base weight for each person in the household. These persons in effect had multiple chances of being in the selected enhanced household: they could have been selected as original sample persons in the household they were in during Wave 1 (which then became an enhanced household), or they could become an associated sample person if their Wave 1 household was not selected but merged later with a sampled Wave 1 household. Their true probability of being included in the enhanced household is higher than their nominal Wave 1 probability of selection, and their assigned base weight should be the reciprocal of this true sample inclusion probability.

This true inclusion probability is not computed directly, for it requires the computation of joint probabilities of selection of multiple households, some of which were not in the original Wave 1 household sample. Instead, a "mover's weight" is assigned to each original and associated sample person in the enhanced household, which has as its expectation the inverse of the true sample inclusion probability. In other words, the movers' weights are unbiased weights, taking into account the complex realized sample design for enhanced households.

In the case when an enhanced household is formed from only one Wave 1 sample household (with associated persons added to it), the mover's weight for each person in the household (original, associated, or additional) is computed as follows for reference month t, enhanced household i:

$$W_{ti} = \frac{W_{1i} S_{1ti}}{S_{ti} - S_{tai}} \qquad \text{(C-2)}$$

where W_{1i} is the initial weight that is common to all original sample persons in the ith enhanced household, S_{1ti} is the number of original sample persons in the ith enhanced household in month t, S_{ti} is the size of the ith enhanced household in month t (all persons), and S_{tai} is the number of additional sample persons in the ith enhanced household in month t. The numerator of this expression is the sum of the initial weights over all original sample persons in the household during month t, and the denominator of this expression is the number of original and associated sample persons in the ith enhanced household in month t. For a discussion of why these are unbiased weights, see, for example, Kalton and Brick (1994).

When two Wave 1 sample households merge, the mover's weight for each sample person (original, associated, or additional) in the household is computed as follows:

$$W_{ti} = \frac{W_{1i} S_{1ti} + W'_{1i} S'_{1ti}}{S_{ti} - S_{tai}} \qquad \text{(C-3)}$$

The two terms in the numerator are for the first and second Wave 1 sample households. The movers' weights for more than two merged Wave 1 sample households are computed analogously.

Wave 2+ Later Wave Noninterview Adjustments

The initial weights have an adjustment for noncooperation in Wave 1; that is, the sample households with nonzero initial weights represent households for which an interview was not completed in Wave 1. There are, however, further losses of sample households in later waves (Wave 2+) for several reasons:

- The household refuses to cooperate in some or all of the later waves.

- The people in the household have moved and cannot be found.

- The household has moved, and has been found, but is too far away for a personal interview and cannot be reached by telephone.[8]

The weights of households for which later wave interviews are completed are adjusted to "represent" sample households (who cooperated in Wave 1) whose interviews are not completed for any of the above reasons. Those adjustments are computed by assigning each sample household with a nonzero initial weight to one of 109 later wave noninterview cells. The noninterview cells are based on the following household characteristics:

---

[8] The SIPP sample is designed so that most of the field work takes place within the SIPP PSUs, to reduce traveling costs. If a household moves too far away from the field areas, a telephone interview is attempted.

1. Reference person is a non-Hispanic white person, or other (two categories).

2. Reference person is a female householder without a spouse and with her own children, a householder 65 years of age or older, or other (three categories).

3. Household income includes welfare payments (AFDC, WIC, Food Stamps, Medicaid, or other welfare), or not (two categories).

4. Household size is 1, 2, 3, or 4 or more persons (four categories).

5. Household has some bond-type financial assets, or not (two categories).

6. Household owns housing unit, is renter, or is living in a public housing project or receiving a rent subsidy from the government (three categories).

7. Number of imputations for household reference person is none, 1, or more than 1 (three categories).

8. Household income as a percentage of the household poverty threshold (with both averaged over 4 reference months): less than or equal to 175 percent, 176 through 450 percent, and more than 450 percent (three categories).

9. Census division (nine categories).

10. Reference person's education level is less than 8 years, 8 to 11 years, 12 to 15 years, or 16 or more years (four categories).

These categories have been found in empirical research to be consistently heterogeneous in later wave noninterview rates (i.e., the categories have divergent noninterview rates). The later wave noninterview adjustment for each noninterview cell is equal to the sum of the initial or mover's weights of all households that have had the later wave interview completed, divided by the sum of the initial or mover's weights of all Wave 1 sample households. (The mover's weight is used whenever a mover's weight is computed for the household.) These adjustments are made separately for each reference month of each later wave of the panel.

Before the final noninterview adjustment is computed for each wave, each noninterview cell is checked. Any noninterview cell with fewer than 30 interviewed households, or with a noninterview adjustment greater than 2, is collapsed with a neighboring cell. Cells are defined as neighboring on the basis of a set of scale values assigned to each noninterview cell. This procedure prevents extreme noninterview adjustments from being made (which will increase sampling variability). The final noninterview adjustment (LWNIA) for the cell, or collapsed cell, is assigned to each household within the cell.

Table C-2 presents the major groupings of noninterview cells (the noninterview cells within these major groupings have similar scale values, and would be collapsed together within these groupings before any collapsing was done across groupings).

**Table C-2. Major Groupings of Later Wave
Noninterview Cells**

Household Characteristics	Number of Nonresponse Cells
Hispanic or nonwhite	
Minimal Assets	15
Assets include Bonds	9
White Non-Hispanic	
Single female householder	1
Householder 65 and older	14
Other householder	
No welfare income	
One person in household	20
Two persons in household	14
Three persons in household	7
Four or more in household	19
Has welfare income	10
Total	109

Wave 2+ Second-Stage Calibration Adjustment (SSCA)

A second-stage calibration adjustment is carried out for each reference month in each later wave, for each rotation group of the panel separately. This adjustment uses the same algorithm as described for Wave 1 weights, with new CPS or CPS-derived control totals computed for each new reference month. The pre-second-stage weights in this case are IW*LWNIA, or MW*LWNIA if a mover's weight was computed for the household. The second-stage calibration adjustments reduce sampling variability by calibrating the final weights to agree with independent control totals. With the later wave cross-sectional weights, the second-stage calibration adjustments also have the effect of reducing biases from population undercoverage (arising from eligible people entering the U.S. population after the Wave 1 reference months).

Calendar Year and Panel Weights

The algorithm for generating the calendar year and panel weights is very similar to that used for computing Wave 2+ weights, with some differences. The most important differences are the following:

- A control date is associated with each calendar year and panel weight (rather than the weight being associated with a month, as with the Wave 1 and Wave 2+ weights).

- For a sample person to have a nonzero weight,

  o data must be present for the sequence of months defined for the weight (all months of year x for calendar year weights, LGTCYxWT, and all reference months of waves 1 to w(x) for panel weights, LGTPNWTx); or

  o data must be present for the first month defined for the weight and for each month thereafter until the sample person is known to have died or moved to an ineligible address (institution, military barracks or foreign living quarters).

Calendar Year and Panel Initial Weights

Longitudinal initial weights (LGTIW) are derived from cross-sectional pre-second-stage weights:

- For the first year calendar weight and all panel weights

  o LGTIW=IW (equals BW*DCF*NAF, the same quantity used for Wave 2+ initial weights.)

- For the second and later calendar year weights

  o LGTIW=IW*LWNIA for persons living in households that did not need a movers adjustment in the first month of the calendar year.

  o LGTIW=MW*LWNIA for persons living in households that had a movers adjustment in the first month of the calendar year.

In all cases, the longitudinal initial weight is the same quantity as the cross-sectional pre-second-stage weight for first month of the longitudinal period. The longitudinal initial weight allows each sample person who has interviews for the months for which they are eligible in the calendar year or panel to represent unsampled people in the population and people in households that were not successfully interviewed in the first month of the longitudinal period.

Calendar Year and Panel Noninterview Adjustments

The noninterview adjustments for each calendar year and panel weight are computed by first assigning each sampled person with a nonzero initial weight to one of 149 noninterview cells. These noninterview cells are based on the following person-level characteristics:

1. Person is a non-Hispanic white person, or other (two categories).

2. Person was self-employed, or not (two categories).

3. Family income was a percentage of the family poverty threshold (with both averaged over four reference months): less than or equal to 175 percent, 176 through 450 percent, and more than 450 percent (three categories).

4. Person in household whose income includes welfare payments (SSI, AFDC, WIC, Food Stamps, Medicaid, or other welfare), person receiving unemployment compensation but not in household with welfare payments, or neither (three categories).

5. Person in household with some bond-type financial assets, or not (two categories).

6. Person's educational level is less than 12 years, 12 to 15 years inclusive, or 16 or more years (three categories).

7. Person was in labor force at least 1 month of wave, or not (two categories).

8. Census division of household (nine categories).

9. Number of imputations in household Wave 1 questionnaire is none, 1, or more than 1 (three categories).

10. Within PSU stratum code of household: poverty stratum or nonpoverty stratum (two categories).

These categories have been found in empirical research to be consistently heterogeneous in later wave noninterview rates. The noninterview adjustment for the noninterview cell (for the particular calendar year [panel] weight) is equal to the sum of the longitudinal initial weights of all sampled persons who live in households that were interviewed in the first month of the calendar year (panel),[9] divided by the sum of the longitudinal initial weights of all sampled persons who have interviews for every month of the calendar year (panel) in which they are eligible.

As with other noninterview adjustments discussed in this appendix, each noninterview cell is checked for small sample sizes and extreme noninterview adjustments. Any noninterview cell with fewer than 30 sampled persons with complete interview strings, or with a calendar year (panel) noninterview adjustment greater than 2, is collapsed with a neighboring cell for that calendar year and panel weight. If necessary, this process can be iterative: a cell may be collapsed into another cell, and then the combined cell may be collapsed further with other cells. A set of scale values determines how cells are collapsed when collapsing is necessary. Table C-3 presents the major groupings of noninterview cells (i.e., the noninterview cells with similar scale values). The noninterview cells within these groupings would be collapsed together among themselves before any collapsing would be done outside of these groupings.

---

[9] Persons who entered the sample after the first month of the calendar year (panel) period (by entering a sampled household) are excluded from these calculations (and receive calendar year [panel] weights of zero). Children who move without their parents (into nonsampled households) during the period are also excluded from these computations and receive calendar year (panel) weights of zero.

**Table C-3. Major Groupings of Calendar Year (Panel)
Noninterview Cells**

Person Characteristics	Number of Nonresponse Cells
Hispanic or Nonwhite	50
White Non-Hispanic	
Less than 12 years education	25
12 to 15 years education	
In labor force	32
Not in labor force	18
16 or more years education	24
Total	149

Calendar Year and Panel Second Stage Adjustments

The calendar year and panel weights that have been computed to this point (called the pre-second stage-weights) for each sampled person (with a complete set of interviews for their eligible months) are equal to LGTIW*LGTNIA. The formula for the final calendar year weights (LGTCY*x*WT) is LGTIW*LGTNIA*SSCA, where SSCA is the second-stage calibration adjustment. The final panel weight follows the same formula: LGTPNWT*x* = LGTIW*LGTNIA*SSCA, though LGTNIA and SSCA are computed differently here. The final weight is computed in both cases from the pre-second-stage longitudinal weights, LGTIW*LGTNIA, in accordance with the algorithm described below. As with the Wave 1 and Wave 2+ cross-sectional weights, the algorithm for second-stage adjustment for calendar year and panel weights can be segmented into a small number of major steps:

- Perform the Cell Collapsing Step
- Perform the Calibration Step

The primary difference between the calendar year (panel) second-stage adjustment procedure and the Wave 1 and 2+ second-stage adjustment procedures is that a spouses' weight equalization is not done for the calendar year (panel) weights; the cell collapsing and calibration steps are nearly identical[10].

The independent estimates for the control month are the same that were used for the Wave 1 and Wave 2+ weights. The second-stage cells for calendar year (panel) weights are given in Figure

---

[10] The maximum number of ratio adjustments allowed in the calendar year (panel) calibration step is higher than the maximum number of ratio adjustments allowed in a single Wave 1 or Wave 2+ calibration step.

C-1. The second-stage calibration algorithm is run separately for each rotation group, with the control totals for each rotation group equal to one-quarter of the CPS control totals.