Weighting the 1996 and 1997 American Community Surveys

Scot A. Dahl

U.S. Census Bureau
Washington, DC 20233

Presented to the American Community Survey Symposium, March 1998.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

1.0 Overview of the American Community Survey

The American Community Survey is an annual survey, under development to provide demographic information about communities and populations every year. This survey will collect the data traditionally collected by the decennial census long form. Development started in 1996 in four sites: Rockland County, NY; Brevard County, FL; Multnomah County, OR; and Fulton County, PA. In 1997, four more sites were added: Douglas County, NE; Otero County, NM; Franklin County, OH, and Houston, TX. Though it is an annual survey, the questionnaires are mailed out and the data collected monthly.

Eventually, in 2003, the survey will be in full-scale production, mailing 250,000 questionnaires per month to every county in the nation. Annual profiles will be produced for all states, cities, counties, metropolitan areas, or population groups for 65,000 or more people. For smaller areas, two to five years of data will be accumulated to produce estimates similar to those of the census long form.

1.1 Sampling Procedure

For each site, a systematic sample was drawn from the Master Address File (MAF) for the site. The MAF is a file of all addresses in a county developed from Census Bureau and US Post Office address listings. The 12-month sampling rates for both small governmental units (SGUs), those with less than 2,500 population, and non-SGUs are shown below. One twelfth of the selected sample was mailed each month.

Table 1A: ACS Base Sampling Rates: 1996 & 1997

	1996			1997		
Site	Non-SGU	SGUs	Sample Size(mo)	Non-SGU	SGUs	Sample Size(mo)
Brevard, FL	0.15	0.30	2,595	0.03	0.09	532
Rockland, NY	0.15	0.30	1,238	0.03	0.09	251
Multnomah, OR	0.15	0.30	3,431	0.03	0.09	703
Fulton, PA	na	0.30	166	na	0.09	50
Douglas, NE				0.15	0.30	2,355
Otero, NM *				0.03	0.09	76
Franklin, OH				0.03	0.09	1,132
Houston, TX**				0.03	0.09	3,440

* The sampling rate was increased to 0.15 in November and December 1997.
** The Houston, TX site consisted of two counties: Fort Bend and Harris.

1.2 Data Collection

Three data collection modes were used to conduct the 1996 and 1997 American Community Surveys: Mail, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI). These three modes are described below.

Mail Phase: The Mail phase began with a pre-notice letter mailed to each sample address on the second to last Wednesday of the month preceding the sample month. The ACS Questionnaire was mailed one week later. One week after that, a reminder card was mailed to all sample housing units. A replacement questionnaire was mailed two weeks later if the original questionnaire had not yet been checked in.

CATI Phase: Approximately five weeks after the mailing of the initial ACS questionnaire, the CATI staff began contacting nonresponding sample addresses by telephone. This phase lasted for approximately one month.

CAPI Phase: The CAPI universe consisted of all outstanding non-response cases remaining after the completion of the CATI phase. A 1 in 3 subsample was selected from these outstanding cases and forwarded to the Field Representatives. Field Representatives visited each assigned household and conducted an interview. The CAPI phase also lasted for approximately one month. For Otero County, NM some addresses were un-mailable. Two-thirds of these cases were selected and sent directly to CAPI.

As an illustration, table 1B shows the pattern of ACS data collection for Rockland County NY in 1997. For example, in February 1997, 251 questionnaires were mailed for panel 9702. Of these, 114 returned as mail responses in February 1997, 56 were CATI or late mail returns received in March, and 81 were late mail returns in April or were subjected to CAPI subsampling.

Table 1B: Mail Panel Month vs. Tabulation Month: Rockland County NY, 1997

Mail Panel	Tabulation Year & Month														Total
	1996													1998	
	11&12	01	02	03	04	05	06	07	08	09	10	11	12	01&02	
9611	814	425													1239
9612	542	259	436												1237
9701		118	45	88											251
9702			114	56	81										251
9703				117	41	94									252
9704					131	38	81								250
9705						140	41	70							251
9706							130	32	89						251
9707								122	38	92					252
9708									125	30	96				251
9709										131	33	87			251
9710											115	35	101		251
9711												111	35	105	251
9712													120	131	251
Total	1356	802	595	261	253	272	252	224	252	253	244	233	256	236	5489

2.0 Description of the Source Files

The data used in weighting the ACS sample was assembled from several sources. These sources were:

2.1 Sample Address/Control File:

The address/control file contained the status and outcome codes for every housing unit address included in the ACS sample. This file also contained the geographic codes (tract, block, address, etc.) and sampling stratum for every sample address.

2.2 Master Address File (MAF)

The MAF was the sampling frame. It contained geographic and other information for every address in each site. This file was constructed from the 1990 Census Address Control File (ACF), modified by periodic deliveries of the USPS Delivery Sequence File (DSF). The 1996 and 1997 samples were originally drawn in September of 1995 and 1996 respectively. After data collection for a sample year was completed, a later version of the MAF was used for to:

- To update the sample's original geographic codes, and
- To produce tract level housing unit counts to be used in the final stages of the weighting.

2.3 Edited Data Files

The edited data files contained the edited response data. The edits handled item nonresponse by imputing from other reported information or by hot-deck imputation. For each state the edited data files consisted of two subsets:

- Edited responses to Household questions (one record per housing unit)
- Edited responses to Population questions (one record per person)

2.4 Population Control Counts

This file contained the most recent population estimates for the counties in the ACS survey[1]. These independent estimates were produced using demographic analysis by the Census Bureau's Population Division and consisted of housing unit (non-Group Quarters) population estimates broken down by:

- Age (one year intervals to age 85)
- Sex (Male, Female)
- Race (White, Black, American Indian, Asian/Pacific Islander (API))[2]
- Ethnicity (Hispanic/Non-Hispanic)

3.0 Preliminary Operations

3.1 MAF Operations

As mentioned above, two operations were performed with the most current version of the MAF available after data collection was completed (around February of 1997 and 1998 respectively for the 1996 and 1997 surveys). These operations were to:

- Update the geography (county, tract, block, and place) of the original sample.
- Produce counts of valid housing units by tract for each site. These counts were used later in the final stages of the weighting.

3.2 Edits and Record Selection

The edits were applied to the raw housing unit and population response data. While not strictly a part of the weighting process, certain edits had to be completed in order to weight each record. The edits imputed for item nonresponse using other data reported by the household, or if necessary, by substituting a value from a nearby neighbor (hot deck imputation).

In the event of multiple responses from one sampled address, the Record Selection Algorithm selected the response to be retained. It also reclassified some records as nonresponses and, rarely, created additional person records for a household. As part of the record selection process: a status (occupied, vacant, delete, etc.), a mode (Mail, CATI, or CAPI), and a tabulation month assigned to every housing unit address, both responses and nonresponses.

3.3 Creation of the Initial Weighting Files

Two initial weighting files were created for each site: a housing unit file with one record for each sampled housing unit address, and a population file with one record for each person in responding housing units. These contained the necessary variables needed to compute the weighting factors.

3.4 Disclosure Avoidance Data Swapping

Some housing units and their members, with characteristics unique within their block group, were swapped with similar housing units in other block groups within the same site. This was done to reduce the possibility that any information about an individual housing unit or its members could be deduced from the tables and the public use files produced. Approximately 1% of the occupied housing units in each site were swapped.

4.0 Computation of the Initial Housing Unit Weighting Factors

The 1996 and 1997 surveys were processed separately but basically the same set of weighting factors was used. The weighting factors used in the 1996 and 1997 ACS fell into four general categories:

- Base weight and mode adjustments (BW, SSF, VMS)
- Sample size adjustments (SRF, FAF)
- Nonresponse adjustments (NIF1, NIF2, MBF)
- Post-stratification adjustments to controls (HPF1, HPF2, PPSF, PPF)

4.1 Base Weight (BW)

This weight was assigned to every housing unit address and was basically the inverse of an address's sampling rate. An adjustment was made to account for the probability that a housing unit on the MAF had already been sampled by other Census Bureau surveys, including the 1996 ACS.

The table below shows the ranges of base weights assigned for Small Governmental Unit (SGU) and non-SGU areas of each site. The wide range of values in 1997 for the four 1996 sites reflects the adjustment made to unduplicate the 1997 sample from the 1996 sample.

Table 4A: Range of Base Weights: 1996 & 1997

Site	1996		1997	
	Non-SGU	SGUs	Non-SGU	SGUs
Brevard, FL	6.63 - 6.67	3.33 - 3.35	27.4 - 38.5	7.44 - 11.5
Rockland, NY	6.64 - 6.67	3.32 - 3.34	27.2 - 42.4	7.39 - 11.5
Multnomah, OR	6.59 - 6.67	3.33 - 3.35	27.5 - 40.3	7.16 - 11.2
Fulton, PA	Na	3.33	na	7.27 - 11.2
Douglas, NE			6.63 - 6.67	3.33 - 3.35
Otero, NM			33.1 - 33.3	11.1
Franklin, OH			33.2 - 33.3	11.1 - 11.2
Houston, TX			33.2 - 33.3	11.1 - 11.2

4.2 Sample Reduction Factor (SRF)

This factor was used for the 1997 sample only and affected only the addresses in the four 1996 sites that were mailed in 1996 and tabulated in 1997. In the four 1996 sites, the sample sizes were reduced from approximately 15% and 30% to 3% and 9% in 1997. Since the survey is based on the month a response was received (not when the questionnaire was mailed), January and February 1997 would have had unusually large numbers of responses compared to the other ten months of 1997. To reduce the possible bias to the editing and weighting, these 'carried-over' records were sub-sampled down to the standard 1997 rates. The sample reduction factor reflects the increased weight assigned to the records that were retained.

Table 4B: Sample Reduction Factors: 1997

Site	1997	

	Non-SGU	SGUs
1996 Sites	5.0	3.33
1997 Sites	na	na

4.3 CAPI Subsampling Factor (SSF)

This factor was 1.0 for all Mail and CATI cases, 3.0 for those records selected in CAPI subsampling, and zero for those not selected. The actual assignment of a value for SSF was complicated by late mail returns received during the month of CAPI operations. The table below shows the effect of this interaction. For Otero County NM, some addresses were unmailable. A two-thirds sample of these were sent directly to CAPI and for these cases SSF = 1.5.

Table 4C: Assignment of SSF to CAPI Universe Cases

Mail status	In CAPI subsample	Not selected for CAPI
Late mail return received	SSF = 1*	SSF = 1
No mail return	SSF = 3	SSF = 0
*Assuming that the mail return was selected over the CAPI return (if any) by the record selection algorithm.		

4.4 Variation in Monthly Response by Mode (VMS)

This factor made the total weight of the Mail, CATI, and CAPI subsampled records tabulated in a month equal to the total weight of all cases originally mailed for that month. Twelve factors were computed for each site, one for each month of the year.

To compute the twelve values of VMS for each site for each sample year, define:

$$M_{sp} = \sum_{i \in s,p} BW_i$$

$$MAIL_{sm} = \sum_{\substack{j \in s,m \\ Mode=Mail}} (BW_j \times SRF_j \times SSF_j)$$

$$CATI_{sm} = \sum_{\substack{j \in s,m \\ Mode=CATI}} (BW_j \times SRF_j \times SSF_j)$$

$$CAPI_{sm} = \sum_{\substack{j \in s,m \\ Mode=CAPI}} (BW_j \times SRF_j \times SSF_j)$$

where:

M_{sp} = for site *s*, the weight of all questionnaires mailed in month *p*
$MAIL_{sm}$ = for site *s*, the weight of the mail cases tabulated in month *m*
$CATI_{sm}$ = for site *s*, the weight of the CATI cases tabulated in month *m*
$CAPI_{sm}$ = for site *s*, the weight of the CAPI cases tabulated in month *m*
s = site code (1, 2, ... , 8)
p = panel month (the month for which the questionnaire was mailed)
m = tabulation month (the month an address was tabulated)

Then VMS for site *s* in month *i* was defined as:

$$VMS_{si} = \frac{M_{si} - MAIL_{si}}{CATI_{si} + CAPI_{si}}$$

Where i = 1... 12

This value of VMS was applied to all CATI and CAPI housing units in site *s*, in tabulation month *i.* For all Mail responses, VMS = 1.0. For addresses whose questionnaires were mailed in November or December but not received until the following year, VMS = 0. The table below is an extract showing the weighted totals of the components of VMS for Rockland County, NY (Site 2) for several months in 1997.

I	M_{si}	$MAIL_{si}$	$CATI_{si}$	$CAPI_{si}$	VMS_{si}
05	8101	5093	730	2967	0.8138
06	8100	4568	910	2616	1.0019
07	8215	4387	644	2246	1.3248
08	8114	4639	752	2704	1.0051
09	8134	4685	573	2757	1.0357
10	8113	4126	696	3170	1.0310

So for September 1997: $VMS_{2,09}$ = (8134 - 4685) / (573 + 2757) = 1.0357

The table below shows the range of values of VMS for the eight sites.

Table 4D: Range of Values of VMS: 1996 & 1997

Site	1996	1997
Brevard, FL	0.89 - 1.13	0.82 - 1.29
Rockland, NY	0.91 - 1.13	0.72 - 1.43
Multnomah, OR	0.78 - 1.16	0.86 - 1.26
Fulton, PA	0.79 - 1.28	0.77 - 1.87
Douglas, NE		0.88 - 1.19
Otero, NM		0.74 - 1.36
Franklin, OH		0.86 - 1.29
Houston, TX		0.91 - 1.14

4.5 Noninterview Factor #1 (NIF1)

This was the first of two factors that adjusted the weight of all respondents to account for both respondents and nonrespondents. NIF1 was a tract-based nonresponse adjustment; it increased the weight on responding housing units to account for both respondents and similar nonrespondents. 'Similar' for this factor meant within the same tract and of the same building type (single or multi-unit). To compute NIF1 define:

$$RES1_{cbt} = \sum_{\substack{j \in c,b,t \\ OccResp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

$$NON1_{cbt} = \sum_{\substack{j \in c,b,t \\ NonResp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

where:

RES1 = weighted sum of the respondents of occupied housing units
NON1 = weighted sum of the non-respondents
c = county code
b = building type (single or multi-unit)
t = tract code

Then NIF1 was computed as:

$$NIF1_{cbt} = \frac{RES1_{cbt} + NON1_{cbt}}{RES1_{cbt}}$$

This value of NIF1 was applied to all occupied, respondent housing units in county *c*, tract *t*, of building type *b*. For vacancies and non-housing units (businesses, vacant lots, etc.), NIF1 = 1.0. For nonrespondents, NIF1 = 0. The table below is an extract showing the weighted totals ($RES1_{cbt}$ and $NON1_{bt}$) for several tracts in Rockland County, NY (county 36 087) for single units (b = 1).

b	tract(t)	$RES1_{cbt}$	$NON1_{cbt}$	$NIF1_{cbt}$
1	010900	2103	0	1.0000
1	011000	686	32	1.0464
1	011100	3350	132	1.0395
1	011200	1971	0	1.0000
1	011300	3430	246	1.0717
1	011401	2890	200	1.0693
1	011402	2277	72	1.0314
1	011501	2283	0	1.0000

So for this county, for single-units in tract 011300: $NIF1_{36087,\ 1,\ 011300}$ = (3430 + 246) / 3430 = 1.0717

The range of values of NIF1 for the eight sites is shown below.

Table 4E: Range of Values of NIF1: 1996 & 1997

Site	1996	1997
Brevard, FL	1.00 - 1.16	1.00 - 1.32
Rockland, NY	1.00 - 1.11	1.00 - 1.36
Multnomah, OR	1.00 - 1.22	1.00 - 1.24
Fulton, PA	1.00 - 1.05	1.00 - 1.02

Douglas, NE		1.00 - 1.29
Otero, NM		1.00 - 1.06
Franklin, OH		1.00 - 1.24
Houston, TX		1.00 - 1.69

4.6 Non-Interview Factor #2 (NIF2)

The second nonresponse adjustment factor was NIF2. This factor was computed in the same manner as NIF1 except that similar housing units were now defined as being tabulated in the same month instead of in the same tract, and NIF2 was computed given that NIF1 has already been applied. Define:

$$RES\,1_{sbm} = \sum_{\substack{j \in s,b,m \\ Occ\ Resp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

$$RES\,2_{sbm} = \sum_{\substack{j \in s,b,m \\ Occ\ Resp}} (BW_j \times SRF_j \times SSF_j \times VMS_j \times NIF\,1_j)$$

$$NON\,1_{sbm} = \sum_{\substack{j \in s,b,m \\ Nonresp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

where:

RES1 = weighted sum of the respondents of occupied housing units
RES2 = weighted (including NIF1) sum of the respondents of occupied housing units
NON1 = weighted sum of the non-respondents

s = site code
b = building type (single or multi-unit)
m = tabulation month

Then NIF2 was computed as:

$$NIF\,2_{sbm} = \frac{RES\,1_{sbm} + NON\,1_{sbm}}{RES\,2_{sbm}}$$

This value of NIF2 was applied to all responding occupied housing units in site *s*, tabulation month *m*, of building type *b*. For vacancies and non-housing units, NIF2 = 1.0. For nonrespondents, NIF2 = 0. The table below is an extract showing the weighted totals ($RES1_{sbm}$, $RES2_{sbm}$ and $NON1_{sbm}$) for several tabulation months (m) in 1997 for Rockland County, NY (site 2) for single units (b = 1).

b	m	$RES2_{sbm}$	$RES1_{sbm}$	$NON1_{sbm}$	$NIF2_{sbm}$
1	05	6061	5922	0	0.9771
1	06	5438	5321	234	1.0216
1	07	5127	5025	265	1.0316
1	08	5541	5427	112	0.9996
1	09	6087	5955	0	0.9783
1	10	6010	5885	102	0.9963

Then NIF2 for August 1997, for single-units is: $NIF2_{2,1,08}$= (5427 + 112) / 5541 = 0.9996

The range of values of NIF2 for the eight sites is shown below.

Table 4F: Range of Values of NIF2: 1996 & 1997

Site	1996	1997
Brevard, FL	0.98 - 1.02	0.99 - 1.04
Rockland, NY	0.98 - 1.03	0.92 - 1.09
Multnomah, OR	0.98 - 1.04	0.97 - 1.03
Fulton, PA	0.98 - 1.04	0.99 - 1.10
Douglas, NE		0.98 - 1.04
Otero, NM		0.97 - 1.17
Franklin, OH		0.98 - 1.05
Houston, TX		0.97 - 1.05

4.7 Mode Bias Factor (MBF)

This factor was an attempt to compensate for the bias resulting from not taking the mode of response into account when calculating NIF1 and NIF2. The concern was that there were systematic differences between the households that returned ACS mail forms and those that did not.

The first step in computing MBF was to calculate an alternative noninterview adjustment, NIFM, using only the CAPI respondents in the denominator. The underlying assumption was that the characteristics of nonrespondents were most like those of the hardest-to-get respondents.

4.7a Noninterview Factor - Mode (NIFM)

This factor was similar to NIF2 in that housing units were grouped by tabulation month (as well as by building type). However NIFM adjusts the weight of just the CAPI respondents to account for both CAPI respondents and all non-respondents. MAIL and CATI cases receive a value of NIFM = 1.0. This factor was not applied directly but rather as part of computing MBF. To compute NIFM, define:

$$RESM_{sbm} = \sum_{\substack{j \in s,b,m \\ CAPI\ Resp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

$$NONM_{sbm} = \sum_{\substack{j \in s,b,m \\ Nonresp}} (BW_j \times SRF_j \times SSF_j \times VMS_j)$$

where:

RESM = weighted sum of the CAPI respondents
NONM = weighted sum of all non-respondents
s = site code
b = building type (single or multi-unit)
m = tabulation month

Then NIFM was computed as:

$$NIFM_{sbm} = \frac{RESM_{sbm} + NONM_{sbm}}{RESM_{sbm}}$$

For purposes of computing MBF, this value of NIFM was applied to all CAPI responding occupied housing units in site *s*, tabulation month *m*, of building type *b*. For Mail and CATI respondents, and for vacancies and non-housing units, NIFM = 1.0. For nonrespondents, NIFM = 0. The table below is an extract showing the weighted totals ($RESM_{sbm}$ and $NONM_{sbm}$) for several tabulation

months in 1997 for Rockland County, NY (site 2) for single units (b = 1).

b	m	$RESM_{cbt}$	$NONM_{cbt}$	$NIFM_{cbt}$
1	06	921	234	1.2873
1	07	816	265	1.2873
1	08	1131	112	1.0990
1	09	1550	0	1.0000
1	10	2102	102	1.0487

Then NIFM for August 1997, for single-units is: $NIFM_{2,1,08}$= (1131 + 112) / 1131 = 1.0990.

The range of values of NIFM for the eight sites is shown below.

Table 4G: Range of Values of NIFM: 1996 & 1997

Site	1996	1997
Brevard, FL	1.00 - 1.19	1.00 - 1.28
Rockland, NY	1.00 - 1.21	1.00 - 1.29
Multnomah, OR	1.00 - 1.27	1.00 - 1.20
Fulton, PA	1.00 - 1.18	1.00 - 1.08
Douglas, NE		1.00 - 1.27
Otero, NM		1.00 - 1.07
Franklin, OH		1.00 - 1.14
Houston, TX		1.00 - 1.19

4.7b Computing MBF

This factor made the total weight, within specified weighting cells, the same as if NIFM had been used instead of NIF1 and NIF2. For any specified group of housing units, the total weight could now be computed two ways. One was to use the nonresponse adjustments NIF1 and NIF2. The other way was to use the other nonresponse adjustment NIFM. MBF was the factor that, when applied to NIF1 and NIF2, caused these two results to be equal for a specific grouping of housing units. For computing MBF, housing units within a site were grouped by tenure, tabulation month, and the marital status of the householder. For each group two weighted totals were computed, one using NIF1 and NIF2, the other using NIFM.

$$N_{somr} = \sum_{\substack{j \in s,o,m,r \\ Occ\ Resp}} (BW_j \times SRF_j \times SSF_j \times VMS_j \times NIF1_j \times NIF2_j)$$

$$M_{somr} = \sum_{\substack{j \in s,o,m,r \\ Occ\ Resp}} (BW_j \times SRF_j \times SSF_j \times VMS_j \times NIFM_j)$$

where:

N = weighted sum computed using NIF1 x NIF2

M = weighted sum for the same cell computed using NIFM
s = site code
o = tenure (owner or renter)
m = tabulation month
r = marital status (married/widowed or single/divorced/separated - based on the status of a householder)

Then MBF was computed as:

$$MBF_{somr} = \frac{M_{somr}}{N_{somr}}$$

This value of MBF was applied to all responding occupied housing units in site *s*, with tenure *o*, in tabulation month *m*, with marital status *r*. For vacancies and non-housing units, MBF = 1.0. After MBF was applied, the weight of a housing unit was:

Housing Unit Weight = BW x SRF x SSF x VMS x NIF1 x NIF2 x MBF

Note that NIFM was not used directly; it was included indirectly as part of the computation of MBF.

The table below is an extract showing the weighted totals (N_{somr} and M_{somr}) for Rockland County, NY (site 2) in 1997.

o	m	r	N_{somr}	M_{somr}	MBF_{somr}
1	06	1	4486	4487	1.0003
1	06	2	915	906	0.9904
1	07	1	4870	4889	1.0040
1	07	2	786	790	1.0052
1	08	1	4408	4384	0.9946
1	08	2	815	834	1.0233
1	09	1	4418	4403	0.9967
1	09	2	1313	1307	0.9951

For Rockland County owner-occupied (o=1) housing units, in August (m=08), where the householder was married or widowed (r=1), MBF is: $MBF_{2,1,08,1}$ = 4384 / 4408 = 0.9946.

The range of values of MBF for the eight sites is shown below. In general, MBF was higher for temporarily occupied housing units than for permanently occupied ones. It also tended to be higher for renters than for owners.

Table 4H: Range of Values of MBF: 1996 & 1997

Site	1996	1997
Brevard, FL	0.99 - 1.05	0.97 - 1.07
Rockland, NY	0.99 - 1.03	0.94 - 1.07
Multnomah, OR	0.98 - 1.03	0.98 - 1.03
Fulton, PA	0.98 - 1.01	0.92 - 1.04
Douglas, NE		0.99 - 1.05
Otero, NM		0.98 - 1.01

Franklin, OH		0.99 - 1.01
Houston, TX		0.99 - 1.04

4.8 Furlough Adjustment Factor (FAF)

This factor only applied to the 1996 survey. It adjusted the weights of the February 1996 CAPI records to account for the 'missing' January 1996 CAPI cases caused by the furlough of late 1995/early 1996. This value was approximately 2.0 for February CAPI cases and exactly 1.0 for all others.

4.9 First Housing Unit Post Stratification Factor (HPF1)

By the time data collection for a sample year was completed, the original sample was almost a year-and-a-half old. In order to account for any new construction that had occurred since the sample was selected and to reflect any geographic coding changes, the weighted number of housing units was compared to housing unit counts from a newer Master Address File (MAF). The factor HPF1 made the weighted number of housing units in a tract equal to the MAF housing unit count for the tract. Define

$$HU_{ct} = \sum_{j \in ct} (BW_j \times SRF_j \times SSF_j \times VMS_j \times NIF1_j \times NIF2_j \times MBF_j \times FAF_j)$$

where:

HU = weighted estimate of housing units in county *c*, tract *t*
c = county code
t = tract code

$$HPF1_{ct} = \frac{MAFHU_{ct}}{HU_{ct}}$$

Then HPF1 was computed:

Where: MAFHU = a count of the valid Master Address File (MAF) addresses in county *c*, tract *t*

This value of HPF1 was applied to all housing unit addresses (including vacancies and non-housing units) in county *c*, tract *t*. The table below is an extract showing the totals ($MAFHU_{ct}$ and HU_{ct}) for Rockland County, NY (county 36 087).

t	$MAFHU_{ct}$	HU_{ct}	$HPF1_{ct}$
011000	740	901	0.8209
011100	4615	4321	1.0681
011200	2214	2525	0.8767
011300	5562	4933	1.1275
011401	3448	3216	1.0720

Then HPF1 for tract 011300 is: $HPF1_{36087,\ 011300}$ = 5562 / 4933 = 1.1275

The range of values of HPF1 for the eight sites is shown below.

Table 4I: Range of Values of HPF1: 1996 & 1997

Site	1996	1997
Brevard, FL	0.92 - 2.45	0.71 - 1.47
Rockland, NY	0.96 - 1.36	0.82 - 1.35

Multnomah, OR	0.60 - 1.94	0.83 - 1.80
Fulton, PA	0.98 - 1.02	0.96 - 1.04
Douglas, NE		0.95 - 1.33
Otero, NM		0.94 - 1.17
Franklin, OH		0.66 - 1.74
Houston, TX		0.62 - 2.29

5.0 Computation of the Person Post Stratification Factor (PPSF)

After computing HPF1 for housing units, the weights of the persons in the housing units were computed. Initially, each person in a housing unit was assigned the weighting factors (BW, ..., HPF1) of their associated housing unit. Then an iterative process was run to compute PPSF.

This factor was then applied to individual person records based on their age, race, sex and Hispanic origin. It adjusted person weights so that the weighted sample of persons approximately matched county population control counts by age, race, sex, and Hispanic origin. These control counts were provided by the Census Bureau's Population Division and reflect an estimate of the population of the county at July 1st.

The first iteration adjusted for race, sex and age group. Define:

$$POP\,1_{rsa} = \sum_{i \in r,s,a} (BW_i \times SRF_i \times \ldots \times HPF\,1_i)$$

$$CC\,1_{rsa} = \sum_{j \in a} (CCPOP_{rsj})$$

where:

POP1 = ACS weighted population estimate for race *r*, sex *s,* age group *a*
CC1 = Population control counts for race *r*, sex *s,* age group *a*
CCPOP = Population control counts for race *r*, sex *s,* and single-year age *j*
r = race code (White, Black, Other)
s = sex code (Male, Female)
a = age group (a grouping of five or more ages, such as: 0-4, 5-9, 10-19, etc.)

$$PPSF(R)_{rsa} = \frac{CC\,1_{rsa}}{POP\,1_{rsa}}$$

Then PPSF(R) for race *r*, sex *s,* and age group *a* was computed as:

After PPSF(R) was applied to all person weights, a second iteration was run that adjusted the weighted population to match the control counts by Hispanic origin, sex and age group. (The age groups used for the Hispanic origin adjustment may have been different than the ones used for the race adjustment). Define:

where:

POP2 = ACS weighted population estimate for Hispanic origin *h*, sex *s,* age group *a* after PPSF(R) has been applied
CC2 = Population control counts for Hispanic origin *h*, sex *s,* age group *a*
CCPOP = Population control counts for Hispanic origin *h*, sex *s,* and single-year age *j*
h = Hispanic origin code (Hispanic, Non-Hispanic)
s = sex code (Male, Female)
a = age group

Then PPSF(H) for Hispanic origin *h*, sex *s,* and age group *a* was computed as:

PPSF(R) was then recomputed after all previously computed values of PPSF(R) and PPSF(H) from the prior iterations had been applied to each person. This process of alternating race and Hispanic adjustments was repeated up to five more times ending with a race adjustment. The final value of PPSF for a person was the product of all of the PPSF(R)"s and PPSF(H)"s computed. The range of values of PPSF for each site is shown in the table below.

Table 5A: Range of Values of PPSF: 1996 & 1997

Site	1996	1997
Brevard, FL	0.8 - 1.9	0.8 - 2.2
Rockland, NY	0.6 - 1.8	0.6 - 1.9
Multnomah, OR	0.7 - 1.6	0.6 - 1.7
Fulton, PA	0.7 - 1.5	0.9 - 1.3
Douglas, NE		0.7 - 1.5
Otero, NM		0.5 - 2.4
Franklin, OH		0.6 - 1.4
Houston, TX		0.6 - 1.9

This completed the weighting of persons. The final person weight is shown below. The last step was to convert the final person to an integer value in a controlled rounding process that ensured that the rounded estimates of population by race, sex, or Hispanic origin were close to the unrounded estimate for that same block, tract, or county.

$$Final\ Person\ Weight = BW \times SRF \times SSF \times \ldots \times HPF1 \times PPSF$$

6.0 Computation of the Final Housing Unit Weighting Factors

6.1 Principal Person Factor (PPF)

This factor transferred some of the over or under-coverage detected in the person weighting onto the housing units. The PPSF factor of one of the householders of a housing unit was assigned to that housing unit. When assigned to a housing unit, this factor was renamed PPF. PPF for un-occupied housing units was 1.0.

6.2 Second Housing Unit Post Stratification Factor (HPF2)

Like HPF1, this factor made the number of housing units in a tract equal to the current MAF control count totals. It was computed in the same fashion as HPF1 except that the term in the denominator, HU_{ct} was computed as:

$$HU_{ct} = \sum_{j \in ct} (BW_j \times SRF_j \times SSF_j \times \ldots \times MBF_j \times FAF_j \times HPF1_j \times PPF_j)$$

HPF2 was then computed in the same manner as HPF1 using the same MAF tract housing unit counts ($MAFHU_{ct}$) used for computing HPF1. In a final step, the final housing unit weight (shown below), was converted to an integer value in a controlled rounding process that ensured that the rounded estimate of housing units within any individual block, tract, or county was within 1.0 of the unrounded estimate for that same block, tract, or county.

$$Final\ HU\ Weight = BW \times SRF \times SSF \times \ldots \times HPF1 \times PPF \times HPF2$$

7.0 Results After Each Weighting Stage

7.1 Housing Unit Factors

Table 7A below shows the change in total number of housing units for each 1997 site after each weighting factor was applied. For each site, the weighted totals were the same for factors VMS through MBF. Basically this was the weighted total of all the questionnaires mailed for each site in 1997. For all sites, except Fulton, the weighted total increased after HPF1 was applied. This reflects the new construction captured by adjusting to a more current MAF than the one used for initial sampling. The effect of PPF was mixed, some counties increase, some decrease, and some have very little change. Finally, the application of HPF2 brings the totals back to the MAF housing unit count.

Table 7A: Housing Unit Estimates After Each Weighting Factor was Applied: All Sites, 1997

	FL	NE	NM	NY	OH	OR	PA	TX
BW	210,301	185,941	27,382	97,091	447,089	277,198	6,653	1,358,311
SRF	210,283	185,941	27,382	97,152	447,089	276,804	6,602	1,358,311
SSF	209,805	184,832	27,300	96,771	444,387	273,982	6,480	1,351,341
VMS	211,760	186,978	27,361	97,561	450,665	280,067	6,714	1,369,916
NIF1	211,760	186,978	27,361	97,561	450,665	280,067	6,714	1,369,916
NIF2	211,760	186,978	27,361	97,561	450,665	280,067	6,714	1,369,916
MBF	211,760	186,978	27,361	97,561	450,665	280,067	6,714	1,369,916
HPF1	219,263	193,928	27,793	102,675	471,623	288,080	6,684	1,439,609
PPF	214,934	193,028	27,362	99,545	462,920	289,308	6,971	1,464,692
HPF2	219,263	193,928	27,793	102,675	471,623	288,080	6,684	1,439,609

Table 7B shows the number of housing units in Rockland county in 1997, by type, after each factor was applied. In particular, as SRF, SSF, and NIF1 were applied, the weight associated with Sample Reduction, Not-in-CAPI , and Non-Response housing units respectively, was redistributed to the other housing units in the site.

The table also illustrates the usual effect of MBF: to increase the number of temporarily occupied housing units at the expense of permanently occupied ones. In Rockland County, HPF1 increases the total number of housing units by 5.2%. And while the total number of housing units after HPF2 was the same as after HPF1, the composition of the housing unit stock (occupied, vacant, etc.) changed. In this site the net effect of PPF was to reduce the number of occupied housing units and increase the numbers of the other types. In other sites, PPF may have the opposite effect.

Table 7B: Housing Unit Estimates by Type After Each Factor was Applied: Rockland County NY, 1997

	Occupied	Vacant	Temp Occ.	Non-Resp.	Non-HUs	Not in CAPI	Sample Reduct.	Total
BW	68,935	903	228	755	1,068	19,534	5,668	97,091
SRF	71,635	1,010	228	942	1,217	22,120	0	97,152
SSF	87,513	2,678	372	2,691	3,517	0	0	96,771
VMS	88,178	2,673	367	2,731	3,612	0	0	97,561
NIF1	90,903	2,673	373	0	3,612	0	0	97,561
NIF2	90,902	2,673	374	0	3,612	0	0	97,561
MBF	90,877	2,673	399	0	3,612	0	0	97,561
HPF1	95,869	2,694	419	0	3,693	0	0	102,675
PPF	92,671	2,694	419	0	3,693	0	0	99,477
HPF2	95,657	2,777	435	0	3,806	0	0	102,675

Table 7C below compares the estimated site population before and after PPSF was applied. The "before" estimate was computed by applying all housing unit factors through HPF1 to each person in the housing unit. In general the effect of PPSF on the total population estimate for a county was relatively small, but table 7D shows that the impact varies across the races. Compared to the population controls, the American Community Survey under-estimated the numbers of Black males and 'Other Race' in Rockland County.

Table 7C: Population Estimates Before and After PPSF was Applied: All Sites, 1997

	FL	NE	NM	NY	OH	OR	PA	TX
Before	447,699	432,640	54,976	280,852	1,024,668	613,050	13,698	3,356,648
After	454,603	431,415	54,315	272,644	994,228	612,692	14,381	3,442,986
Change	+1.55%	-0.28%	-1.20%	-2.92%	-2.97%	-0.06%	+4.99%	+2.57%

Table 7D: Population Estimates By Race Before and After PPSF was Applied: Rockland County NY, 1997

	WHITE		BLACK		OTHER RACE	
	Male	Female	Male	Female	Male	Female
Before	113,082	123,341	12,404	16,978	7,351	7,696
After	110,240	115,183	14,261	15,888	8,277	8,795
Change	-2.5%	-6.61%	+15.0%	-6.4%	+12.6%	+14.3%

Footnotes:

[1] For Portland the counts were for Multnomah county only and do not include the sample portions of Clackamas and Washington counties. For Columbus, OH the counts were for Franklin County only and do not include the sample portions of Fairfield County.

[2] For weighting purposes, American Indian and API were combined into just one 'Other' category.