RESEARCH REPORT SERIES
(Statistics #2003-03)

Transparent File Construction for the State of New Jersey in Census 2000

Michael Ikeda and Julie Tsay

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: July 30, 2003

Transparent File Construction for the State of New Jersey in Census 2000

Michael Ikeda and Julie Tsay, Statistical Research Division, U.S. Census Bureau

Abstract

The Decennial Census is traditionally an enumeration of persons and housing units together with characteristics of interest. Previously used methods for adjusting for net coverage only reflect coverage error for persons. For consistent handling of traditional tabulations we need coordinated adjustment of persons and housing units. This report describes the construction of a data file for Decennial short form data based on such a coordinated adjustment given the presence of external data constraints. The example presented is for the state of New Jersey and uses estimates of persons from the Accuracy and Coverage Evaluation (March 2001 estimates) as the external constraints.

Key Words: quadratic programming, coverage adjustment, household estimates

I. Introduction

Traditionally, the Decennial Census is an enumeration of persons and housing units together with characteristics of interest. The data collected are, after further processing, placed on a data file which is the source of subsequent tabulations.

Previously used methods for adjusting for net coverage only reflect coverage error for persons. Added or deleted persons are placed on the data file under a special classification, but are not associated with housing units. For consistent handling of traditional tabulations we need coordinated adjustment of persons and housing units. In this report, we describe the construction of a data file for Decennial short form data given the presence of external data constraints. The specific example presented is for the state of New Jersey and uses estimates of persons in housing units from the Accuracy and Coverage Evaluation (A.C.E.) survey (March 2001 estimates).

[*Note: The estimates from the March 2001 A.C.E. were determined to be unacceptable because A.C.E. failed to measure significant numbers of erroneous census enumerations and did not adequately measure residency status. As such, it is estimated that the March 2001 A.C.E. overstated the net undercount by at least three million persons and Census 2000 data products were not adjusted. Thus, the example presented here should be viewed in this context and would not be indicative of Census 2000 coverage errors. For further research efforts, it is recommended that the estimates from A.C.E. Revision II be used.*]

Section II gives some brief background and overview about the transparent file. Section III gives an overview of the methodology for construction of the New Jersey transparent file. Section IV gives an overview of the results. Section V provides a general summary and outlines possibilities for future work. Section VI contains a brief list of references.

The appendices contain additional methodological details and results. Appendix 1 contains a summary of results for each of the 210 poststrata in New Jersey. Appendix 2 contains documentation of software and data files and the schema for the A.C.E. poststratification. Appendix 3 contains details on the housing unit categories. Appendix 4 contains the file layout for the transparent (research) file. Appendix 5 contains details of the donor selection procedure.

Documentation of previous work on transparent file construction, along with additional background information and related references, can be found in Isaki, Ikeda, Tsay, and Fuller (2000) and Isaki, Tsay, and Fuller (2001). The basic methodology for the current work is very similar to the methodology for the previous work.

II. Background

The 2000 Decennial Census produces a Short Form file which is the source of further tabulations. Persons in households can be associated with housing unit characteristics, and their relationship to the householder is recorded.

There are also persons who are not in households. These people have similar information on person characteristics but are associated with group quarters characteristics, rather than housing unit characteristics. Relationship information is not collected for these persons and they are excluded from A.C.E.

Diagram 1. 2000 Census Short Form File

2000 Census Short Form File

For each geographic block:

Households / Persons	Non-household Category
Persons listed by household with relationships, characteristics and housing unit data.	Persons listed with characteristics. Group Quarters data.

Previously used methods for adjusting for undercount modify the Short Form file by adding two special categories of "Adjustment persons". One category of "Adjustment persons" is reserved for undercounted persons, the other category is reserved for overcounted persons. These persons have short form person characteristics but are not associated with housing units and lack relationship information. They thus resemble the "Non-Household" persons, in terms of the data available for them.

However, the "Adjustment Persons" are really persons in households, since the estimates used to calculate the "Adjustment Persons" are estimates of persons in households. It would therefore be desirable to adjust both persons and households, in order that we can analyze household results allowing for undercoverage and overcoverage.

One solution to this problem is the creation of a "Transparent File" that incorporates the estimates of coverage.

Definition: Given a census operation that includes sampling and estimation, we define a transparent file to be a census data file that is devoid of any evidence of sampling and estimation.

A transparent Decennial Census data file would -

i) have the appearance of an enumeration with effective weights of one to avoid non-integer estimates

ii) be constructed by duplicating or eliminating households at the block level

iii) contain a listing of housing units and persons with their short form data and block identification

A transparent Decennial Census data file maintains the A.C.E. (or other) estimates, has the structure of a traditional Census data file, and assigns all data from households to the household category.

While early planning for the 2000 Census included the possibility of a transparent file as part of production, it was decided that operational considerations ruled this out. One of the main problems was a requirement that the transparent file maintain (as much as possible) A.C.E. synthetic estimate totals at low levels of geography. This requirement considerably complicates the process of constructing a transparent file.

Instead, a transparent file would be constructed as a research file, starting with a single state. We wanted a state that would provide a good test of the methodology. We thought that the state should be moderately large, although not one of the very largest. Some degree of ethnic diversity was also desirable. Consultation with Louisa Miller of Population Division provided three suggestions: New Jersey, Arizona, and Virginia. We decided on New Jersey, because it had slightly above average proportions of both Blacks and Hispanics, as well as an above average proportion of Asians.

III. Methodology

A. Introduction

The basic idea in constructing a Transparent File is to classify households enumerated in the

Census by relevant characteristics and estimate the true number of households. Then delete or duplicate households on the Census file to attain the number of households required. The final product is the Transparent File. Not all Census variables were included on the research Transparent File, although the missing variables can be added.

Many steps are involved in the construction of a Transparent File. In the following, we describe the general nature of each step and in later sections, refer to appendices for details. The methodology outlined here is very similar to that outlined in Isaki, Ikeda, Tsay, and Fuller (2000) and Isaki, Tsay, and Fuller (2001). The methodology uses Quadratic Programming combined with raking to obtain estimates of the true number of households. These household estimates must maintain the person control totals. The estimates are then integerized and households on the Census file are deleted or duplicated. The addition or deletion of households takes place in three stages: 4+ person households, 2-3 person households, 1 person households. Control totals are adjusted after each stage. The final Transparent File does not necessarily maintain the person control totals, although the donor selection process does attempt to minimize differences from the person control totals. One difference from previous work is that the state-level person poststratum totals that are used as person controls are synthetic estimates instead of direct A.C.E. estimates. The housing unit categories and person poststrata are also somewhat more detailed in the current work.

B. Regression Estimation

In the estimation step, we seek housing unit weights, w_{ik}, for the $i^{\underline{th}}$ housing unit (HU) category in the $k^{\underline{th}}$ geographic block. The weights are obtained subject to state-level A.C.E. synthetic poststratum totals and A.C.E. synthetic block person totals. The HU categories i are defined by size of household (1, 2, 3, 4, 5, 6+); race/ethnic origin of householder (Non-hispanic White/Other, Non-hispanic Black, Hispanic, Non-hispanic Asian and Remainder); age of householder (0-29, 30-49, 50-64, 65+), sex of householder ("female, spouse present" is placed in the same household category as male for Black and Hispanic households); tenure; mail return category (low/high); MSA status (large MSA/other) and several categories of an "undercount variable".

The "undercount variable" breaks down households into several categories if the householder is a female with no spouse present and a renter and either Black or Hispanic. The categories are based on the composition of the household and depend on the age of the householder, presence of children, and presence of adult non-elderly Black or Hispanic males.

The main change to the housing unit categories from previous transparent file work is the addition of the mail return and MSA status categories.

The households were grouped into categories to reduce computation time and to facilitate imputation. By grouping HUs, we are able to use standard software packages for our estimation problem. Full details on the housing unit categories are found in Appendix 3. Selecting the housing unit categories for New Jersey involved a substantial amount of trial and error. We plan to explore methods for automating the process of defining housing unit categories.

The basic estimation procedure used is regression, with a raking component. Additional details are given just below in Section III-C.

C. Quadratic Programming (QP) Iteration

The basic estimation method used is a combination of quadratic programming and raking. The quadratic programming method differs from the usual regression estimator in that weights are restricted to be nonnegative. The raking is used because of the number of block restrictions. The procedure consists of several steps. Do the QP step and get a solution. Then use the solution in the raking operation to maintain block controls. Then use the block control results in a new QP step, and so forth. Note that the last step in the iteration is a QP step.

Let

$$a_{rk} = \hat{X}_{Tk} \left(\sum_{i=1}^{n} \sum_{j=1}^{m} c_{r-1,i} \, x_{ijk} \right)^{-1}$$

where $c_{0i} = 1$,
r is an iteration index,
x_{ijk} is the number of Census persons in housing unit category i, poststratum j, and block k,
and $\hat{X}_{Tk}$ is the A.C.E. synthetic total for block k.

Given the vector $\mathbf{c}_{r-1}$, a new vector of weights ($c_{r,1}, c_{r,2}, \ldots, c_{r,n}$) is chosen to minimize

$$f(\mathbf{c}) = \sum_{i=1}^{n} (c_{ri} - 1)^2 \, h_i$$

with respect to c_{ri} , subject to

$$\sum_{i=1}^{n} c_{ri} \left(\sum_{k=1}^{B} a_{rk} x_{ijk} \right) = \hat{X}_j \quad ; \quad j = 1,2,..,m$$

where $c_{ri} > 0$
h_i is the number of Census housing units in housing unit category i,
B is the number of blocks in the state,
and $\hat{X}_j$ is the A.C.E. state-level synthetic person total for poststratum j.

After four iterations, the weight for housing unit type i and block k is $w_{ik} = a_{4k} \; c_{4i}$. The weight construction combines elements of raking and of least squares regression estimation.

If a software program could accommodate a QP program with all of the observations and several thousand constraints in a reasonable amount of computer time, then there would be no need to form HU categories and use a raking procedure. A single application of QP would suffice and the estimator provided by the solution would be 'close' to a regression estimator. Additional details on the Quadratic Programming procedure can be found in Isaki, Ikeda, Tsay, and Fuller (2000) and Isaki, Tsay, and Fuller (2001).

D. Integer Estimates

If h_{ik} represents the number of HUs in category i in block k in the Census, then $w_{ik}\ h_{ik}$ is an estimator of the number of housing units for the Transparent File, where w_{ik} is the weight from the QP. Since $w_{ik}\ h_{ik}$ is not an integer, we implement an integer rounding procedure that converts $w_{ik}\ h_{ik}$ to its "nearest" integer in such a manner that the sums

$$\sum_i w_{ik} h_{ik} \text{ and } \sum_k w_{ik} h_{ik}$$

are maintained.

Let t_{ik} denote the integerized $w_{ik}\ h_{ik}$, obtained by using a controlled rounding procedure from Cox and Ernst (1982). Define,

$$U_{ik} = t_{ik} - h_{ik}.$$

Then U_{ik} is the number of households required to supplement (add to) the Census file if U_{ik} is positive, and the number to delete (subtract from) the Census file if U_{ik} is negative. The determination of HUs to add to (or delete from) the Census file is accomplished by a donor selection routine which is described below.

E. Donor Selection

In selecting HUs in the Census file for addition or deletion, we first work with housing units of four or more persons. Then after completing that group, we work with housing units of two and three persons and, finally, work with one-person housing units. The development of the Transparent File in such stages affords the possibility of adjusting state person controls and block total person controls at successive stages. To summarize, the Transparent File is constructed by adding HUs to and deleting HUs from the Census file. The donor selection process is the most time-consuming part of the entire procedure. A single donor selection phase (e.g. 4+ person units) could take more than a week for the state of New Jersey. We are exploring methods for speeding up the donor selection phase. Further details on the donor selection process can be found in Appendix 5.

The U_{ik} for four or more person HUs is strictly determined by the QP method and integer rounding. The ratio of the synthetic block estimates and the unrounded Transparent File block estimates are quite close to one and the synthetic state person controls are maintained. Hence, using integer rounding derived U_{ik} values and random donor selection of HUs should result in a Transparent File of four or more person HUs that is close to the A.C.E. block total person estimates and the state person controls. The differences would be due to the integer rounding to obtain the U_{ik} and the variation of types of persons in the i^{th} HU category. Given these expectations, the four or more person HUs are selected (for addition or deletion) strictly based on the U_{ik}.

A condition check using total population of each block is then performed on housing units with four or more persons in the transparent file. First, the synthetic block estimate is compared with the transparent file total. All blocks with transparent file total exceeding the synthetic total are considered completed if the block contains no 1-, 2- or 3- person households. Otherwise they are subject to a condition check. The condition is that the total person count in households with four or more persons on the Transparent File in the block plus the Census person count in households with 1-, 2-, and 3- persons be less than or equal to the total person synthetic estimate. To meet the condition, 4+ person households are randomly eliminated from the transparent file.

Once the four or more person HUs in the Transparent File are identified, the QP method is applied to the one, two and three person HUs to obtain weights, w_{ik}, that are constructed to maintain A.C.E. synthetic block person totals and synthetic state person controls minus those persons in four or more person households on the Transparent File. The w_{ik} are then used to obtain the U_{ik}. The same condition check that is used for the 4+ person housing units is used for elimination of 2- and 3- person housing units. A complication is that the quadratic programming step may run fine for all households, but fail for the 1-3 person HUs.

Once the Transparent File is complete with two-person and three-person HUs, the single person HUs on the Transparent File are determined by raking the estimated HUs by category and block that were obtained from the preceding QP application on 1-, 2- and 3- person HUs. The raking is done by tract and controlled rounding is performed to integerize the raked estimates. For each major race, the difference between the A.C.E. synthetic tract estimate and the persons in the Transparent File (persons in 2^+ HUs) is allocated to their respective single person HU categories.

Two complications can arise with the single unit procedure. One complication is that the tract total race difference may be negative. The second is that there may not exist any single-person HUs of a certain race in the tract. In the first situation, housing units of the problem race are eliminated from the transparent file in a prescribed manner (see Appendix 5). In the second situation, the recommended procedure is to combine the 'problem' race with another race which has a positive difference larger than the absolute value of the negative difference and containing some single-person unit HUs. If the concern is lack of single units, the remedy is to collapse the 'problem race' with another race according to the Long Form race collapsing criteria until single-person HUs appear in the combined race group. Then, raking can be implemented and the usual U_{ik} determined. The procedure maintains race or combined race at the tract level and also at the state level. Donor selection is on the basis of the U_{ik}.

IV. Results

There are 112,746 blocks in the state of New Jersey; 111,951 blocks contained 1 to 1573 housing units; 34 percent of the blocks contained fewer than 11 HUs.

Some Census summary statistics (excludes Group Quarters) for New Jersey are as follows:

Characteristics	
1. Persons	8,219,529
2. Housing units	3,310,275
3. Occupied Housing units	3,064,645
4. Vacant Housing units	245,630
5. Blocks (non-empty)	111,951
6. Tracts	1,938
7. Counties	21

The race/ethnicity groups are White (Non-hispanic White and other), Black (Non-hispanic Black), Hispanic, and Asian (Non-hispanic Asian, Native Hawaiian or Pacific Islander, American Indian).

The race/ethnicity groups are based on the poststratum groups for the Census 2000 A.C.E. Some descriptive statistics for New Jersey are as follows:

Characteristics	Census Population	Census Housing Units	A.C.E. Estimate
1. White Owner	4,454,965	1,670,215	4,472,603
2. White Renter	1,092,542	548,633	1,108,658
3. Black Owner	496,953	156,864	499,907
4. Black Renter	590,063	229,524	612,097
5. Asian Owner	301,158	81,194	302,795
6. Asian Renter	186,512	67,065	189,815
7. Hispanic Owner	405,177	103,200	409,159
8. Hispanic Renter	692,159	207,950	720,521
9. Total	8,219,529	3,064,645	8,315,555
10. Owner	5,658,253	2,011,473	5,684,464
11. Renter	2,561,276	1,053,172	2,631,091

New Jersey estimates from the Census, A.C.E., and the Transparent File for the 210 A.C.E. poststrata in New Jersey can be found in Appendix 1. More details on the A.C.E. postratum groups can be found at the end of Appendix 2 or in DSSD memorandum Q-37 (see references).

Quantiles of relative differences between A.C.E. and Transparent File poststratum estimates are given below. The differences are mostly fairly small in magnitude. Positive relative differences indicate an A.C.E. estimate larger than the Transparent File estimate. Note that relative differences of large absolute magnitude are generally associated with very small poststrata.

Quantile	Relative Difference (%)
100% Maximum	11.5385
95%	2.3256
90%	1.4728
75%	0.3425
50% Median	0.0000
25%	-0.5882
5%	-5.7143
0% Minimum	-16.6667

Some statistics on relative differences (in percent) between the A.C.E. and the Transparent File poststratum estimates grouped by various characteristics are given below:

Characteristic	Number of poststrata	Minimum	Maximum	Range	Median
White	56	-2.0997	1.3980	3.4977	-0.00161
Black	56	-8.3916	11.5385	19.9301	-0.09690
Hispanic	56	-16.6667	3.6082	20.2749	-0.10441
Asian	42	-6.6148	6.5217	13.1365	-0.37849
Age under 18	30	-0.0176	2.9371	2.9574	0.30096
Male 18 to 29	30	-5.7143	3.4765	9.1908	0.06278
Female 18 to 29	30	-2.1739	1.8868	4.0607	0.06107
Male 30 to 49	30	-9.0909	2.3810	11.4719	-0.17118
Female 30 to 49	30	-4.5574	3.6082	8.1657	0.06942
Male 50+	30	-16.6667	8.2192	24.8428	-0.19516
Female 50+	30	-16.3043	11.5385	27.8428	-0.56581
Owner	105	-16.3043	8.2192	24.5235	0.14826
Non-Owner	105	-16.6667	11.5385	28.2051	-0.12065
High Return Rate	84	-8.3916	3.4765	11.8681	0.02882
Low Return Rate	84	-16.6667	11.5385	28.2051	0.05368
Large MSA	84	-1.1490	1.3980	2.5469	0.04951
All Other TEAs	84	-16.6667	11.5385	28.2051	0.00000

Additional results are shown in Table 1-3. Note that no statistical testing was done on any of the differences. Any statements about relative magnitudes are descriptive only and do not imply statistical significance.

Table 1 provides selected (mostly race by tenure) summary totals of the 210 poststrata. Most of the Transparent File totals in the table are fairly close to the A.C.E. totals although the differences in the totals for Asians are relatively larger (the difference for Asian owners is the largest in the table at slightly more than one A.C.E. standard error).

Table 2 provides a comparison between the census and the Transparent File for housing units by race and tenure, as well as a comparison between person differences and housing unit differences. Based on Table 2, it appears that adjusted persons have a tendency to be in smaller sized households for Asian renters and for all groups of owners. Note that the percent increase for Asian renter units is the largest in the table, but the percent increase for Asian renter persons is similar to the overall percent increase for renters.

Table 3 provides summary statistics for blocks and tracts for differences by race or tenure. Let TF_k, SYN_k and CEN_k denote the transparent file estimate, synthetic estimate, and Census count for a characteristic in the k-th area, respectively, where an area can be a block or a tract, and let N denote the number of areas with a non-zero population for the given characteristic. Define

$$\text{Mean Squared Difference} = \text{MSD}_k = N^{-1} \sum_{k=1}^{N} (TF_k - SYN_k)^2$$

$$\text{Mean Absolute Difference} = \text{MAD}_k = N^{-1} \sum_{k=1}^{N} |TF_k - SYN_k|$$

$$\text{Relative Difference} = \text{RD}_k = (TF_K - SYN_K)\,[CEN_K]^{-1}$$

The differences at the block level in total persons are due to rounding. The differences in other categories are larger than those for total persons, because no direct restrictions were imposed on those categories at the block level. Smaller groups tend to have larger relative differences at both the block and tract levels. The absolute and mean squared tract-level differences by tenure are larger than the corresponding tract-level differences by race because there is an attempt to control by race at the tract level in the donor selection process. The mean absolute differences appear to be reasonably small at the block and tract levels, although the relative differences can be fairly variable for smaller groups.

Table 1. Person Summary Estimates for the State of New Jersey

Category	Census	A.C.E.	s.e of A.C.E.	Transparent File	Difference A.C.E. -T.File
Black Persons	1,087,016	1,112,003	5104.57	1,112,111	-107
Black Owner	496,953	499,907	2768.01	499,606	301
Black Owner age 0-17	142,215	144,520	823.61	144,063	457
Black Renter	590,063	612,096	4093.05	612,505	-408
Hispanic Persons	1,097,336	1,129,680	5172.42	1,130,049	-369
Hispanic Owner	405,177	409,159	2197.53	408,831	328
Hispanic Renter	692,159	720,521	4420.95	721,218	-697
Asian Persons	487,670	492,610	3101.07	491,094	1,516
Asian Owner	301,158	302,795	2584.20	300,056	2,739
Asian Renter	186,512	189,815	1843.09	191,038	-1,223
White Persons	5,547,507	5,581,261	18083.97	5,582,301	-1,039
White Owner	4,454,965	4,472,603	17084.55	4,472,999	-396
White Renter	1,092,542	1,108,659	5860.28	1,109,302	-643
Total Persons	8,219,529	8,315,554	20369.56	8,315,555	-1

Table 2. Housing Unit Estimates for the State of New Jersey 2000 Census

Category	Census	Transparent File	Difference T. File-Census (Persons)**	Difference T. File-Census (HU)	% Difference (Persons)**	% Difference (HU)
Black	386,388	396,228	25,095	9,840	2.31	2.55
Black Owner	156,864	159,462	2,653	2,598	0.53	1.66
Black Renter	229,524	236,766	22,442	7,242	3.80	3.16
Hispanic	311,150	323,157	32,713	12,007	2.98	3.86
Hispanic Owner	103,200	105,346	3,654	2,146	0.90	2.08
Hispanic Renter	207,950	217,811	29,059	9,861	4.19	4.74
Asian	148,259	152,095	3,424	3,836	0.70	2.59
Asian Owner	81,194	81,135	-1,102	-59	-0.36	-0.07
Asian Renter	67,065	70,960	4,526	3,895	2.43	5.81
White	2,218,848	2,268,814	34,794	49,966	0.63	2.25
White Owner	1,670,215	1,714,601	18,034	44,386	0.40	2.66
White Renter	548,633	554,213	16,760	5,580	1.53	1.02
Owned HUs	2,011,473	2,060,544	23,239	49,071	0.41	2.44
Rented HUs	1,053,172	1,079,750	72,787	26,578	2.84	2.52
Occupied HUs	3,064,645	3,140,294	96,026	75,649	1.17	2.47

* * Refer to Table 1. Person estimates are based on person characteristics.

Table 3. Summary Statistics of Transparent File and Synthetic Person Estimates at the Block and Tract Levels for the State of New Jersey

BLOCKS	Number of Non-Zero Blocks	Mean Squared Difference	Mean Absolute Difference	Quantiles of Relative Difference			
Characteristic	N	MSD	MAD	5% of RD_k	95% of RD_k	Mean of TF_k	Mean of CEN_k
Total Persons	111951	0.20	0.31	-0.018	0.014	74.28	73.42
Owners	108759	4.41	0.91	-0.059	0.054	52.24	52.03
Renters	71719	6.54	1.14	-0.074	0.169	36.70	35.68
Blacks	44986	6.03	1.29	-0.350	0.486	24.72	24.16
Hispanics	59576	5.39	1.21	-0.453	0.601	18.97	18.42
Whites	108344	5.82	1.23	-0.079	0.085	51.53	51.21
Asians	42257	10.65	1.68	-0.992	0.885	11.62	11.54

TRACTS	Number of Non-Zero Tracts	Mean Squared Difference	Mean Absolute Dfference	Quantiles of Relative Difference			
Characteristic	N	MSD	MAD	5% of RD_k	95% of RD_k	Mean of TF_k	Mean of CEN_k
Total Persons	1938	1.46	0.85	-0.0005	0.0005	4290.79	4241.24
Owners	1935	317.18	11.99	-0.020	0.021	2936.17	2924.16
Renters	1935	315.98	11.94	-0.025	0.034	1361.27	1323.66
Blacks	1928	35.93	3.54	-0.111	0.121	576.82	563.80
Hispanics	1927	22.01	2.78	-0.061	0.073	586.43	569.45
Whites	1938	25.20	2.43	-0.004	0.006	2880.44	2862.49
Asians	1913	55.23	4.09	-0.178	0.168	256.71	254.92

V. Summary and Future Work

The implementation of the transparent file methodology on the state of New Jersey generally seems to have been successful. The resulting person estimates are similar to the A.C.E. estimates and the housing unit estimates seem reasonable based on our initial examination.

There are some operational issues that need to be addressed. We need to substantially speed up the donor selection phase. Currently it can take more than a week to complete one donor selection phase for the state of New Jersey. The entire process has three donor selection phases (4+ person hh, 2-3 person hh, 1 person hh), although the single-person phase takes less time.

In addition, we need to automate the process. The current procedure required a substantial amount of trial and error. One problem is that it is possible for the quadratic programming step to work fine for all households but fail when we do the quadratic programming on just the 1-3 person households.

The obvious extension is to continue the process on additional states. We would prefer to have some resolution of the operational issues before we try this. We might also want to consider incorporation of results from the A.C.E. Revision II. Other sets of person controls could also be used, such as the demographic analysis results. One problem is that the demographic analysis results are only available at the national level. We would need some way of extending the demographic analysis person estimates down to lower levels of geography. Another possible extension is to incorporate housing unit controls into the procedure, such as those from the A.C.E. housing unit dual system estimates.

VI. References

1. Cox, L.H. and Ernst, L.R. (1982). Controlled Rounding. INFOR 20, 423-432.

2. Isaki, C.T., Ikeda, M.M., Tsay, J.H. and Fuller, W.A. (2000). "An Estimation File that Incorporates Auxiliary Information", Journal of Official Statistics, vol. 16, no. 2, pg. 155-172.

3. Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2001) "Report on Transparent File Construction for the Decennial Census". Attachment to Census Bureau internal memorandum from Cary T. Isaki to Tommy Wright, "Report on Transparent File Work through Sacramento Dress Rehearsal", June 11, 2001.

4. Memorandum from D. Kostanich to D. Stoudt, January 5, 2001, "Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation (U.S.) - Reissue of Q-29," DSSD Census 2000 Procedures and Operations Memorandum Series #Q-37.

APPENDIX 1

New Jersey 210 Poststrata Summary

Details on the poststratification scheme can be found at the end of Appendix 2.

	Poststratum	Census	A.C.E.	Transparent File	Difference (A.C.E-T.F)	(Difference/A.C.E.)*100
1	011	910554	918480	917837	643	0.0700
2	012	176192	179475	179369	106	0.0591
3	013	171966	169617	169291	326	0.1922
4	014	581114	591433	591719	-286	-0.0484
5	015	615689	622415	621906	509	0.0818
6	016	595972	599128	599450	-322	-0.0537
7	017	707583	708755	710595	-1840	-0.2596
8	051	93009	88187	87885	302	0.3425
9	052	28316	27903	27791	112	0.4014
10	053	26495	25028	24912	116	0.4635
11	054	72737	68630	68411	219	0.3191
12	055	70811	66955	66019	936	1.3980
13	056	83101	82705	82736	-31	-0.0375
14	057	102979	101591	102737	-1146	-1.1281
15	251	51276	53300	53241	59	0.1107
16	252	7571	7664	7676	-12	-0.1566
17	253	7332	7302	7318	-16	-0.2191
18	254	30537	31111	31112	-1	-0.0032
19	255	32409	33053	32993	60	0.1815
20	256	23769	24173	24213	-40	-0.1655
21	257	24545	24482	24542	-60	-0.2451
22	291	9731	10045	10020	25	0.2489
23	292	2069	2124	2109	15	0.7062
24	293	2002	2055	2047	8	0.3893
25	294	6933	6745	6735	10	0.1483
26	295	7100	7028	7059	-31	-0.4411
27	296	6340	6545	6564	-19	-0.2903
28	297	6833	6674	6712	-38	-0.5694
29	331	149923	154395	154270	125	0.0810
30	332	74202	77428	77343	85	0.1098
31	333	78878	80396	80362	34	0.0423
32	334	134008	137339	137409	-70	-0.0510
33	335	129835	132959	132885	74	0.0557
34	336	80915	80944	81126	-182	-0.2249
35	337	128243	125527	126042	-515	-0.4103
36	341	52211	52707	52679	28	0.0531

37	342	35424	36873	36853	20	0.0542
38	343	34109	34452	34456	-4	-0.0116
39	344	51454	53383	53383	0	0.0000
40	345	43592	43809	43784	25	0.0571
41	346	31174	30610	30693	-83	-0.2712
42	347	41596	40296	40445	-149	-0.3698
43	391	5668	5679	5680	-1	-0.0176
44	392	1727	1815	1795	20	1.1019
45	393	1885	1974	1972	2	0.1013
46	394	3819	3944	3951	-7	-0.1775
47	395	3963	4119	4112	7	0.1699
48	396	2113	2088	2088	0	0.0000
49	397	2602	2498	2499	-1	-0.0400
50	401	1206	1246	1242	4	0.3210
51	402	608	681	688	-7	-1.0279
52	403	641	713	718	-5	-0.7013
53	404	976	994	1007	-13	-1.3079
54	405	950	957	976	-19	-1.9854
55	406	365	381	389	-8	-2.0997
56	407	455	451	455	-4	-0.8869
57	411	93786	95646	95269	377	0.3942
58	412	18841	19064	19009	55	0.2885
59	413	19090	19567	19524	43	0.2198
60	414	46497	47684	47654	30	0.0629
61	415	55349	56386	56540	-154	-0.2731
62	416	36297	36580	36595	-15	-0.0410
63	417	46126	46280	46684	-404	-0.8729
64	421	45782	46212	46138	74	0.1601
65	422	12523	12164	12144	20	0.1644
66	423	12563	12222	12202	20	0.1636
67	424	22644	23006	22987	19	0.0826
68	425	26172	25986	25907	79	0.3040
69	426	21833	21136	21090	46	0.2176
70	427	30479	28999	28934	65	0.2241
71	431	2500	2515	2512	3	0.1193
72	432	481	489	472	17	3.4765
73	433	497	514	506	8	1.5564
74	434	1268	1272	1265	7	0.5503
75	435	1459	1444	1437	7	0.4848
76	436	1034	1024	1024	0	0.0000
77	437	1261	1245	1252	-7	-0.5622
78	441	147	147	144	3	2.0408
79	442	32	35	37	-2	-5.7143
80	443	27	29	29	0	0.0000

81	444	48	49	49	0	0.0000
82	445	62	61	60	1	1.6393
83	446	75	73	67	6	8.2192
84	447	80	78	75	3	3.8462
85	451	89923	94202	94091	111	0.1178
86	452	21546	23458	23507	-49	-0.2089
87	453	27717	29381	29410	-29	-0.0987
88	454	36114	37611	37813	-202	-0.5371
89	455	48592	50117	50303	-186	-0.3711
90	456	16818	16755	16910	-155	-0.9251
91	457	27121	27385	27665	-280	-1.0225
92	461	109494	114817	114780	37	0.0322
93	462	25681	27068	27050	18	0.0665
94	463	35985	38664	38621	43	0.1112
95	464	38359	39582	39599	-17	-0.0429
96	465	53526	54071	54046	25	0.0462
97	466	21931	21317	21239	78	0.3659
98	467	34648	35004	34800	204	0.5828
99	471	728	756	754	2	0.2646
100	472	202	212	205	7	3.3019
101	473	198	203	204	-1	-0.4926
102	474	323	340	342	-2	-0.5882
103	475	434	436	430	6	1.3761
104	476	142	143	155	-12	-8.3916
105	477	166	159	169	-10	-6.2893
106	481	202	205	205	0	0.0000
107	482	35	35	35	0	0.0000
108	483	47	46	47	-1	-2.1739
109	484	41	42	41	1	2.3810
110	485	52	51	50	1	1.9608
111	486	11	11	11	0	0.0000
112	487	27	26	23	3	11.5385
113	491	84815	85614	85020	594	0.6938
114	492	20169	20838	20827	11	0.0528
115	493	20556	20974	20952	22	0.1049
116	494	43949	44877	44974	-97	-0.2161
117	495	49212	49865	49890	-25	-0.0501
118	496	23313	23329	23354	-25	-0.1072
119	497	28644	29118	29400	-282	-0.9685
120	501	37331	37308	37239	69	0.1849
121	502	12522	12831	12774	57	0.4442
122	503	11472	11356	11331	25	0.2201
123	504	19808	20299	20320	-21	-0.1035
124	505	19434	19318	19272	46	0.2381

125	506	13072	12577	12584	-7	-0.0557
126	507	14495	14358	14385	-27	-0.1880
127	511	1988	2041	2012	29	1.4209
128	512	301	305	301	4	1.3115
129	513	328	337	339	-2	-0.5935
130	514	809	820	861	-41	-5.0000
131	515	1075	1116	1106	10	0.8961
132	516	407	397	403	-6	-1.5113
133	517	422	414	413	1	0.2415
134	521	345	346	342	4	1.1561
135	522	79	86	84	2	2.3256
136	523	96	106	104	2	1.8868
137	524	158	157	164	-7	-4.4586
138	525	193	194	187	7	3.6082
139	526	89	86	86	0	0.0000
140	527	95	92	107	-15	-16.3043
141	531	126137	128949	128655	294	0.2280
142	532	56095	61022	61005	17	0.0279
143	533	47930	50300	50353	-53	-0.1054
144	534	67226	70973	71302	-329	-0.4636
145	535	66351	66584	66743	-159	-0.2388
146	536	25323	26178	26347	-169	-0.6456
147	537	33335	33247	33629	-382	-1.1490
148	541	86061	87770	87472	298	0.3395
149	542	37580	43423	43223	200	0.4606
150	543	31906	33817	33790	27	0.0798
151	544	41087	43671	43743	-72	-0.1649
152	545	38249	38955	39002	-47	-0.1207
153	546	14864	15411	15492	-81	-0.5256
154	547	17600	17658	17847	-189	-1.0703
155	551	698	722	718	4	0.5540
156	552	244	280	280	0	0.0000
157	553	213	219	220	-1	-0.4566
158	554	336	354	380	-26	-7.3446
159	555	270	283	282	1	0.3534
160	556	100	101	108	-7	-6.9307
161	557	92	89	95	-6	-6.7416
162	561	192	201	199	2	0.9950
163	562	41	49	50	-1	-2.0408
164	563	54	62	63	-1	-1.6129
165	564	56	66	72	-6	-9.0909
166	565	67	76	77	-1	-1.3158
167	566	30	36	42	-6	-16.6667
168	567	22	25	29	-4	-16.0000

169	571	466	488	474	14	2.8688
170	572	141	154	153	1	0.6493
171	573	159	150	149	1	0.6666
172	574	333	365	373	-8	-2.1917
173	575	347	328	328	0	0.0000
174	576	207	227	232	-5	-2.2026
175	577	231	218	213	5	2.2935
176	581	327	356	355	1	0.2809
177	582	199	204	214	-10	-4.9019
178	583	199	218	222	-4	-1.8348
179	584	239	244	248	-4	-1.6393
180	585	210	230	228	2	0.8695
181	586	84	86	84	2	2.3255
182	587	84	92	86	6	6.5217
183	591	82111	81511	79832	1679	2.0598
184	592	18285	17970	17696	274	1.5247
185	593	18917	19405	19093	312	1.6078
186	594	48062	48982	48394	588	1.2004
187	595	56798	57038	57357	-319	-0.5592
188	596	32263	32530	32275	255	0.7839
189	597	36153	36652	36582	70	0.1909
190	601	41608	41489	41038	451	1.0870
191	602	22972	24184	24654	-470	-1.9434
192	603	23614	23766	23979	-213	-0.8962
193	604	37698	38906	39425	-519	-1.3339
194	605	32410	32876	32941	-65	-0.1977
195	606	10960	11092	11159	-67	-0.6040
196	607	11514	11465	11756	-291	-2.5381
197	631	1795	1876	1853	23	1.2260
198	632	426	401	419	-18	-4.4887
199	633	389	405	401	4	0.9876
200	634	1079	1071	1096	-25	-2.3342
201	635	1095	1141	1193	-52	-4.5574
202	636	864	867	875	-8	-0.9227
203	637	1037	1016	1068	-52	-5.1181
204	641	1357	1430	1388	42	2.9370
205	642	417	470	476	-6	-1.2766
206	643	356	403	406	-3	-0.7444
207	644	686	707	737	-30	-4.2433
208	645	699	732	725	7	0.9563
209	646	361	351	369	-18	-5.1282
210	647	518	514	548	-34	-6.6147

APPENDIX 2

DOCUMENTATION OF SOFTWARE, DATA FILES, AND POST-STRATIFICATION USED IN THE STATE OF NEW JERSEY TRANSPARENT FILE CONSTRUCTION

This document gives an overview of the programs along with input and output files for constructing the transparent file for the State of New Jersey. The basic steps include classifying occupied housing units (HU) by relevant characteristics, determining the HU weight, estimating the number of households and converting the estimates to integers, and deleting or duplicating households on the census file to obtain the desired transparent file. The schema for A.C.E. poststratification in Census 2000 can be found at the end of this Appendix.

I Classification of the Occupied Housing Unit Category

The **cor-hucat.sas** program, a SAS program that reads the Hundred Percent Edited Detail File (HEDF) , classifies households by relevant characteristics and creates the output file **hucat_nj.dat.** with housing unit category attached to the housing unit record . The housing unit category record layout is in Appendix 3.

II Housing Unit Weight Determination (w_{ik})

1. Initialization

crtqpdat.sas is a program that creates the initial phase counts of persons in housing unit category i, ACE poststratum j and block k (x_{ijk}); The initial phase housing unit counts in housing unit category i in block k (h_{ik}) and the A.C.E. total persons control (X_j). The program includes two input files and three output files for the running of the quadratic programming.

<u>Input Files</u>

1) File with total persons of each poststratum from A.C.E.
2) **hucat_nj.dat**

<u>Output Files</u>

1) File of Person Controls (**syntot-nj.dat)**

This file contains 4 variables:

a) Poststratum code
b) Total persons in each poststratum from Census
c) Total persons in each poststratum from A.C.E.

d) The adjustment factor for each poststratum

2) Housing Unit File (**cnhucat-nj.dat**)

This file contains 2 variables.

a) Housing unit category code for each HU.
b) Number of initial phase housing units in each housing unit category.

3) Person File (**cnpercat-nj.dat)**

This file contains 2 variables as well.

a) Poststratum code within each HU category (PCAT)
b) Number of initial phase persons in each PCAT.

2. Running Quadratic Programming

Using the quadratic programming (QP) method to obtain housing weights for occupied housing units, the weight construction procedure consists of iteration steps of QP and a raking operation to maintain block controls to A.C.E. synthetic estimates of total persons for the block.

There are 3 input files associated with the version of **hufullr2.f** for calculating weights controlling to ACE totals.

<u>Input Files</u>

1) File of Person Controls from ACE
2) Housing Unit File from Census
3) Person File

<u>Output Files</u>

1) File of HU Weights

There is one record for each HU on the Census housing unit file that has five variables.

a) HU identifier
b) HU category 1
c) HU category 2
d) Initial Weight
e) Final Weight

2) File of HU Level Diagnostics

3) File of Person Level Diagnostics

The two diagnostic files are used by diagnostic SAS programs to check the lower constraint element and upper constraint element .

3. Iterating Weight Construction Procedure

Program **calab**.sas** is a SAS program that applies the weight from QP and raking to maintain the block controls. Then the weight construction procedure is repeated in an iterative manner by using the block control results in a new QP input. The ** is an iteration index . This program has 4 input files and 2 output files.

<u>Input Files</u>

1) File of housing unit weight from QP.
2) File of total persons control in each poststratum from A.C.E.
3) File with number of persons in each HU category from Census.
4) File with number of persons in each PCAT in each block.

<u>Output Files</u>

1) Adjusted Person Block Control File

   This file contains 2 variables.

   a) Block identifier
   b) Adjusted persons in each block..

2) The Person File From Adjustment .

   This file contains 2 variables.

   a) PCAT code.
   b) Number of adjusted persons in each PCAT.

III. Integer Estimates (t_{ik})

The **prernd.sas** SAS program applies the final weight of the fourth QP solution to estimate the required households. Then it reads in the Census files and creates the arrays needed for the controlled rounding program. There are 2 files associated with the version of the FORTRAN program **rndnj.f** for the controlled rounding of the estimates of households (W_{ik}*h_{ik}) to integers.

Input File: **hubkwt.dat**

Output File: **rndhu.output**

The program for calculating the undercount or overcount housing units for each category i in every block k is in **crt-donor.**sas, that reads in the output from controlled rounding and writes out the output file with the name **undhu-c**-t#.dat**. Where " ** " is the index of the county and "#" is the index of the tract. This file contains 2 variables.

a) Block identifier
b) Number of required housing units for each HU category in each block. (U_{ik})

IV. Donor selection.

For convenience, when the required housing units are selected at random from the housing units in every block, we run 4 to 5 tracts at each time to generate a group file. During this process, the **nock-c**-#.sas** is used in selecting the HUs from the Census file . This program includes 2 input files and 1 output file.

Input Files

1) Census file.
2) File with number of undercount or overcount HUs in each block .

Output File

Transparent File of Group of Tracts (**nockc**-#.dat**)

Where " ** " is the index of the county and " # " is the index of the group file of tracts , the generated output files **nockc**-#.dat** are then combined into one file called **tpf-nock-4+.dat**. At that time a condition check is performed on each block to identify if the blocks with the total persons count of households with 4 or more persons on the transparent file plus Census count of households of 1, 2 and 3 persons exceeds the total person in A.C.E (dfp1). This procedure is carried out by a program called **chkper4+.sas,** which produces an output with a listing of the dfp1 for each HU category i in block k.

To find the candidate HUs for elimination , a program called **findelm-nock.sas** is used to select the duplicated HUs with nonmixed households in dfp1. Then generates the elimination of HUs in the output file. There are 4 files (3 input and 1 output) included in the **findelm-nock.sas** program.

Input Files

1) Transparent File
2) Census File
3) Person File from A.C.E.

Output File

Eliminate HU File

Program **elm-tpf.sas** eliminates the HU from the transparent file and creates the **tpf-elm4+.dat ,** which is the Transparent File for four or more households. The four or more person HUs portion of the Transparent File is constructed first. The software for constructing the transparent file for 4 or more households resides in directory **/home/jtsay/cen2000.**

V. Transparent File Construction

Using the same QP procedure, control rounding and donor selection as was done for the 4+ HUs we constructed the transparent file for 2 and 3 person HUs, but all the software for constructing the transparent file for 2 and 3 person HUs----**tpf-elm-nock.dat** resides in subdirectory **/home/jtsay/cen2000/tpf3p.**

Software **crt-bk-tmi-nock.sas** is a program for creating control totals for 1 person households. There are 4 input files and 2 output files.

Input Files

1) Transparent File for 2+ Persons
2) HEDF File
3) A.C.E. Poststratum Person File
4) File of Housing Units Weight

Output Files

1) File of the Block Total Population Control (b_k)

   This file contains 2 variables.

   a) Block identifier.
   b) Total persons of each block.

2) File of Controls for HU Category i Types that are within Race m (t_{mi}) in Each Tract.

This file contains 4 variables.

a) Tract code
b) Race code
c) HU category
d) Number of control persons in each T_{mi} in each Tract

Programs **rak-#-**.sas** are for raking which is done by tract where # indicate the county's code and ** indicate the tract's code then controlled rounding of the estimator to integer and the computation of the undercount or overcount housing units in category i in block k (U_{ik}). Donor selection for the single HU is the same as before, and the transparent file for single housing units is **tpf-1p-nock.dat**, all software are resides in subdirectory **/home/jtsay/cen2000/tpf1p.**

Finally, we combined the transparent file for 4+ person HUs along with transparent file for 2 and 3 person HUs and the transparent file for the single HU into one file called **tpf-nock-all.dat.** The record layout for the transparent file is in Appendix 4.

VI. 2000 Transparent File - Determining A.C.E. Controls for QP

The 2000 A.C.E. is designed to produce adjustment factors by region for Non-Hispanic White and Other Owners and for the entire U.S. for other races (Non-Hispanic Black, Native Hawaiian or Pacific Islander, Non-Hispanic Asian, Hispanic, American Indian or Alaska Native on Reservation and Off-Reservation American Indian or Alaska Native) and tenure types. Within race groups tenure by sub-geography, detail is provided for Non-Hispanic Black and Hispanic while only tenure is provided for the remaining race groups. Non-Hispanic White, Non-Hispanic Black and Hispanic detail is also crossed by mail return rate categories (two of them).

All categories above are crossed by seven age-sex categories. A schematic of the detail follows this section. The crossed categories are termed poststrata. Depending on the final A.C.E. distribution of sample persons, some poststrata may be combined with others. When combined, less detail, e.g. geographic, is available.

Each state will be done separately with the largest state done last in each region. For the largest states, adjust person controls for region race totals by accounting for persons already on the T. File. Note that person race controls are determined for each state. This means that for one state, Indians on reservation may stand alone while in another state, it is collapsed with another race group.

Since all sub poststratum group controls are based on synthetic estimation, we know all of the controls before T. File construction. The manner of collapsing of geographic entities is dictated by the poststratum groups. For example, if the 'All other TEA's' group is too small and exhaustive, we look to combine them with the tracts in the 'Small MSA and Non-MSA MO/MB' group.

Schematic for Post-stratification Variables

<table>
<tr><th></th><th colspan="4">High Return Rate*</th><th colspan="4">Low Return Rate*</th></tr>
<tr><th></th><th>Northeast</th><th>Midwest</th><th>South</th><th>West</th><th>Northeast</th><th>Midwest</th><th>South</th><th>West</th></tr>
<tr><td>Non-Hispanic White and Other</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Owner</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Large MSA MO/MB</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr>
<tr><td>Medium MSA MO/MB</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td></tr>
<tr><td>Small MSA and Non-MSA MO/MB</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td></tr>
<tr><td>All Other TEAS</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td><td>32</td></tr>
<tr><td>Non-Owner</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Large MSA MO/MB</td><td colspan="4">33</td><td colspan="4">34</td></tr>
<tr><td>Medium MSA MO/MB</td><td colspan="4">35</td><td colspan="4">36</td></tr>
<tr><td>Small MSA and Non-MSA MO/MB</td><td colspan="4">37</td><td colspan="4">38</td></tr>
<tr><td>All other TEAs</td><td colspan="4">39</td><td colspan="4">40</td></tr>
<tr><td>Non-Hispanic Black</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Owner</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Large MSA MO/MB
Medium MSA MO/MB</td><td colspan="4">41</td><td colspan="4">42</td></tr>
<tr><td>Small MSA and Non-MSA MO/MB
All other TEAs</td><td colspan="4">43</td><td colspan="4">44</td></tr>
<tr><td>Non-Owner</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Large MSA MO/MB
Medium MSA MO/MB</td><td colspan="4">45</td><td colspan="4">46</td></tr>
<tr><td>Small MSA and Non-MSA MO/MB
All other TEAs</td><td colspan="4">47</td><td colspan="4">48</td></tr>
<tr><td>Hispanic</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Owner</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>Large MSA MO/MB
Medium MSA MO/MB</td><td colspan="4">49</td><td colspan="4">50</td></tr>
<tr><td>Small MSA and Non-MSA MO/MB
All other TEAs</td><td colspan="4">51</td><td colspan="4">52</td></tr>
</table>

Non-Owner		
Large MSA MO/MB Medium MSA MO/MB	53	54
Small MSA and Non-MSA MO/MB All other TEAs	55	56
Native Hawaiian or Pacific Islander		
Owner	57	
Non-Owner	58	
Non-Hispanic Asian		
Owner	59	
Non-Owner	60	
American Indian or Alaska Native on Reservation		
Owner	61	
Non-Owner	62	
Off-Reservation American Indian or Alaska Native		
Owner	63	
Non-Owner	64	

* Separate median cut-off values will be formed from census data for each Race/Hispanic Origin by Tenure group.

Age/Sex Groups:

	Male	Female
Under 18	1	
18 to 29	2	3
30 to 49	4	5
50^{+}	6	7

Full details on the A.C.E. poststrata can be found in memorandum Q-37 (see references). The poststratum number is the poststratum group number followed by the age/sex group number.

APPENDIX 3

Record Layout for The State of New Jersey Housing Unit Categories

Variable			Value	
Character	Name	Description	code	Description
1	H1	Household size	1	one person
			2	two people
			3	three people
			4	four people
			5	five people
			6	six or more people
2	H2	Race of Householder	1	Non-Hispanic White
			2	Non-Hispanic Blacks
			3	Non-Hispanic Asians and rest
			4	Hispanic
3	H3	Tenure	1	Owned
			2	Rented
4	H4	Age of Householder	1	Age 0 - 29
			2	Age 30 - 49
			3	Age 50 - 64
			4	Age 65+
5	H5	Sex of Householder		For (Black or Hispanic) hh
			1	Male hh or Female hh with spouse
			2	Female hh with no spouse present
				For all other hh
			1	Male hh
			2	Female hh
6	H6	Undercount Variable	0	Household size=1 OR
				hh not Black or Hispanic OR
				Owner hh OR
				Male hh OR
				Female hh with spouse present

Female B|H renter with no spouse present AND

0 No B2|H2 no any child

1 No B2|H2 but with child < 10

2 No B2|H2 but with child > 9

3 Size=2 and hh age<30 and one B1|H1

4 Size=2 and hh age 30-64 and one B2|H2

5 Size>2 and hh age <30 and one B1|H1

6 Size>2 and hh age 30-64 and one B2|H2

7 Size>2 and hh age <30 and 2 or more B1|H1

8 Size>2 and hh age 30-64 and 2 or more B2|H2

9 otherwise

7. H7 0 Low Return Rate

1 otherwise

8. H8 1 Large MSA

0 otherwise

Remarks

hh=householder

B|H Black or Hispanic

B1|H1 if hh is Black then B1|H1 = Black males 18 to 29

if hh is Hispanic then B1|H1 = Hispanic males 18 to 29

B2|H2 if hh is Black then B2|H2 =Black males 18 to 64

if hh is Hispanic then B2|H2 =Hispanic males 18 to 64

APPENDIX 4

RECORD LAYOUT FOR TRANSPARENT FILE (RESEARCH FILE)

BLOCK RECORD

Variable	Description	Length
RT	Record Type 1	1
ST	Tabulation State	2
COUNTY	Tabulation County	3
TRACT	Census Tract	6
BLOCK	Census Block	4
BLOCKSFX	Block Group	1
BLANK1	Blank	1
DIVISION	Division	17
PLACE	Place	9
BLANK2	Blank	1
POP100	Population Count(100%)	5
HU100	Housing Count(100%)	5
BLANK3	Blank	1
BTYP	Block type	1

HOUSING UNIT RECORD

Variable	Description	Length
RT	Record Type 4	1
COUNTY	Tabulation County	3
TRACT	Census Tract	6
BLOCK	Tabulation Block Number	4
BLOCKSFX	Tabulation Block suffix	1
ID	MAF and DMAF ID	12
NRU	Nonresponse Follow-up Universe	1
NP	Number of Persons at this Housing Unit	3
FINST	Final Status of Unit	1
SFINST	Source of FINST and NP	1
UBSA	Units at Basic Street Address(BSA)	4
HSUB	Housing unit Substituted	1
RFT	Record TYPE and Form Type of the CUF Primary Return Record for the ID	17
STENURE	Tenure	2

HHT	Household/Family Type	2
P65	Number of people 64+ years in Household	2
P18	Number of People Under 18 years in household	2
NPF	Number of people in Family	2
NRC	Number of Related Children under 18 in household	2
TEN	Tenure	1
HUCAT	Housing Unit Category	8
DPCODE	Duplicate Code	3

PERSON RECORD

Variable	Description	Length
RT	Record Type 5	1
COUNTY	Tabulation County	3
TRACT	Census Tract	6
BLOCK	Tabulation Block Number	4
BLOCKSFX	Tabulation Block Suffix	1
PUID	Current Unit ID Number	12
REL	Relationship	3
SEX	Sex	1
QAGE	Age	3
QDB	Date of birth	8
QSPAN	Hispanic Origin Code	3
QRACE	Race Code	2
QDDP	Data Defined Person	1
QGQTYP	Person's Group Quarters Type code	3
POSTR1	Poststratum: First 2 Digits	2
POSTR2	Poststratum: Third digit	1
OC	Own Child indicator	1
RC	Related Child indicator	1
AGELONG	Age Recode (90+ recoded to 90)	2
DPCODE	Duplicate Code	3

APPENDIX 5.

DONOR SELECTION PROCESS FOR NEW JERSEY (top down)

The donor selection process for Transparent File construction for New Jersey used the 'top down' methodology (previously documented in Isaki, Tsay, and Fuller (2001)). The donor selection procedure is a 'top down' approach in which the U_{ik} households are added to the Transparent File (in stages, e.g. 4^+ HUs, then 2 to 3 person HUs, etc.) in their entirety. Then the condition check is applied to determine the number of HUs to eliminate from the file to exactly achieve the synthetic estimate (done in stages as well). So, if the synthetic estimate for a block is 67. 34, the 4^+ T. File is 75 and the Census count is 5 for persons in 1, 2 and 3 person HUs, 13 are deleted from the 4^+ T. File and 13 are added to the control totals used for 1-3 person households.

The 'top down' approach is desirable because it achieves the U_{ik} count better than the alternative 'bottom up' procedure (procedure also documented in Isaki, Tsay, and Fuller (2001)). The 'top down' approach deletes from persons in the U_{ik} rather than supplements the 1, 2, 3 person Census counts while attempting to attain the synthetic block total. Furthermore, the 'top down' adjusts to achieve the synthetic block total. It does it with the following procedure -

(i) Identify all blocks k with Transparent File 4^+ block person totals exceeding the synthetic estimate of total person for block k.

a. If there are <u>no</u> persons in the block in 1, 2, or 3 person households in the Census, then accept the 4^+ person T. File households in the block. Set the iteration $a_{5k} = \varnothing$.

b. Otherwise, suppose the number of Census persons in 1, 2, or 3 person HUs is 5 as in the above example. Here, we impose a condition check after the donor selection of 4^+ HUs. That is, we require that the final 4^+ HUs in the Transparent File for block k plus the Census 1, 2, 3 person HUs equal the Total person synthetic estimate.

So, we want

$$\text{Syn Total} = \text{T. File } 4^+ + \text{Census } 1, 2, 3.$$

So, since Syn Total = 67.34, T. File 4^+ = 75 we need to eliminate 13 persons from HUs in the T. File 4^+.

(ii) In addition, for QP application on 1, 2, and 3 person HUs we use

$$a_{5k} = [\text{Total Person Synthetic - Final T. File } 4^{+} \text{ persons}] /_{k} \sum_{i(1,2,3)} C_{4j} X_{ijk}$$

(iii) The manner of 4^{+} HU person elimination follows -

1. If less than 4 persons for removal, use the closest 4 person substituted HU

.

2. If between 4 and 8, choose among 4^{+} HU combinations that are closest to that desired.

3. If exceeding 8, randomly select HUs by first forming combinations of substituted HUs; select a combination and continue until the total persons were achieved as closely as possible. If the number required was large, we randomly selected 4^{+} HUs until we achieved or attained closeness to the target.

After the initial QP application and donor selection there were 64 blocks requiring the elimination step for a total of 489 persons eliminated in 72 households. After the QP for 1, 2, & 3 person HUs we identified 5566 blocks with negative differences (total synthetic minus T. File 2^{+}) and of these 1580 contained single units. Unfortunately, 3986 blocks had negative 1 to 6 and no single units. We accepted as completed, all 2^{+} HUs with zero single unit counts. We then performed an elimination step on 2 and 3 person HUs.

In the 2/3 HU elimination step from blocks with negative differences between total person synthetic and Transparent File 2^{+}, we eliminated 2 and 3 person substituted HUs first to convert to a positive dfp1. When there were not sufficient substitutions, we eliminated all substitutions and eliminated some non-substitutions to attain the positive dfp1 target. Note that all substituted 2/3 person HUs were subject to elimination. Blocks with negative differences and containing no census single units were considered completed and added to the T. File.

There were 1580 blocks where Synthetic total - T.F. 2^{+} total was negative and containing census single unit persons. For each of these blocks, the absolute difference was added to the census single person count and the sum represented the target elimination (dfp1). We eliminated substitutions (duplications) first. Among substitutions, if the required elimination was two or less, we randomly chose one among the two person HUs. Same for three persons. Larger figures required selecting random combinations of 2 and 3 person HUs. The largest figure was twenty-eight.

No attention was paid to race or other HU category characteristics in selection for elimination. Nearly all figures were under four and most were under three. Upon selection, the HUs were eliminated from the 2^{+} T. File. After completion, we checked the tract / race differences (T_{m}) between the synthetic total and the 2^{+} T. file for negativeness.

Elimination - Tract / Race

We obtained 1285 negative tract / race T_m in the state of New Jersey. For these tract / race, we need to eliminate race persons until T_m is positive. Three tracts contained no single units.

We used the following algorithm to identify HUs for elimination. Assume that the absolute value of the negative difference for a given tract / race, rounded up to an integer, is R. We then listed all blocks in the tract in order of increasing dfp1 where dfp1 = Synthetic - T. File 2^+ - Census 1 person total for the race in question. This places the block with the largest negative value of dfp1 first in the list.

We then considered elimination from the T. File 2^+, non-mixed HUs of the race in question in the first block in the list.

Given R we formed an inventory of sizes of HUs in the tract whose combinations exactly totaled R. Then, we scanned the HUs in the first block for the largest HU eligible for a combination. When an HU was found, it was marked for elimination and the magnitude of the block's dfp1 and the tract R were each reduced by the number of persons in the HU. After this adjustment was made, the new block with the largest negative dfp1 was searched for the largest HU eligible for the new set of combinations. This continued until the recomputed R was zero. We then proceeded to the next tract.

If there were no non-mixed substitution HUs within one of R in the tract but a non-mixed non-substitution HU within one of R existed, then the non-mixed non-substituted HU was eliminated.

If in the above, several HUs of the same size existed in a block and not all were needed, a random selection of HUs to be eliminated was undertaken in the block. In summary, the combinations of HUs, the ordering of blocks and the distribution of non-mixed substituted HUs in the block dictated the elimination process.

Upon completion of the tract / race elimination, the block differences (synthetic total - Transparent File 2^+ total) from blocks without any Census single units but with positive differences were summed and spread over the remaining block constraints in a ratio adjustment prior to raking the t_{mik}.