SAMPLE DESIGN, ESTIMATION, AND PRESENTATION OF SAMPLING ERRORS FOR THE 1980 CENSUS PROVISIONAL ESTIMATES OF SOCIAL, ECONOMIC AND HOUSING CHARACTERISTICS

Milton C. Fan, John H. Thompson, Jay Kim and Henry F. Woltman, U.S. Bureau of the Census

I. INTRODUCTION

As part of the 1980 census procedure, an Early National Sample (ENS) was designed to provide data users early access to sample data on selected social, economic and housing characteristics for the United States, each state and the District of Columbia, and for the 38 Standard Metropolitan Statistics Areas (SMSA's) with one million or more persons. This paper describes the sample design, sample selection procedure, sample estimates, variance estimates and the method of presenting sampling errors used for the ENS.

During the 1980 census, seven data items were asked of the more than 226.5 million persons who were enumerated in the census. In addition, a few data items were asked at each of over 88 million housing units, the number depending on whether the unit was occupied or vacant. This so called short-form information, constitutes the 100-percent or complete count data from the census.

Additional data items called census sample or long-form information were collected from a sample of persons and housing units. The census sample was a systematic sample of persons and housing units selected at one of two sampling rates, one-in-six and one-in-two. The sampling unit for the 1980 census was the housing unit, including all of its occupants. For persons living in group quarters, the sample unit was the person. In incorporated places of less than 2500 persons, one-half of all housing units and persons in group quarters were to be included in the sample. In other places one-sixth of the housing units and persons in group quarters were sampled. The purpose of this sampling plan was to provide more reliable estimates for small places. When both sampling rates were taken into account across the United States, approximately 19 percent of all persons and housing units were included in the census sample.

Clerical editing was required on the sample or long forms; this included coding of occupation, place of work, income and other write-in answers. Consequently sample data were expected to become available later than the complete count. Budgetary and other problems introduced additional delay in the coding operation on the full census sample. Therefore an Early National Sample of the long form questionnaires was drawn and the coding and editing of these questionnaires were expedited. A supplementary report "Provisional Estimates of Social, Economic and Housing Characteristics" was published in the spring of 1982, at least six months earlier than corresponding data could have been published from the full census sample.

II. SAMPLE DESIGN AND SAMPLE SELECTION PROCEDURE FOR EARLY NATIONAL SAMPLE

The Early National Sample (ENS) was developed to provide data users with early sample estimates for all states (and the District of Columbia) and the 38 SMSA's with one million or more persons. The sample was designed to provide certain minimum levels of reliability on estimates of per capita income for each of the above geographic areas.

The ENS can essentially be viewed as a stratified two-stage sample of enumeration districts (ED's) 1/ and persons and housing units within the ED's. The ED's in the nation were stratified into 98 strata; 48 strata consisting of the whole or state portions of the 38 SMSA's with one million or more persons; 24 strata consisting of the balances of the 24 states containing a portion of one or more of the 38 SMSA's; and 26 strata consisting of the 26 states which did not contain any portion of the 38 SMSA's. For the first stage, the 1980 Census ED's were arranged into clusters and these ED clusters were sampled systematically with probability proportionate to size (PPS). The second stage of selection was simply the census sample of persons and housing units within the selected ED's. The use of ED's as the first stage unit was required in order that the selected ED's could be given priority status during the necessary coding operations. It was also not possible to sample within ED's due to the administrative requirements of the Census Processing Operations.

1. Sample Size Criteria and Determination

The sample sizes (of ED's) within each stratum were initially chosen to achieve at most a 5-percent coefficient of variation (CV) on the stratum estimate of per capita income. The 5-percent CV criterion for sample size determination was the result of requests from the Census Bureau subject matter specialists. However, it should be noted that in states containing a portion of one or more of the 38 SMSA's with one million or more persons, the state level CV's of per capita income will be considerably less than 5-percent. Thus, to make the state estimates comparable, it was decided to achieve at most a 3.5-percent CV on state estimates of per capita income for the 26 states containing no portion of the 38 SMSA's with one million or more persons. For the purposes of sample size determination, the ED clusters were assumed to be composed of only one ED.

a. Notation

The notation is given in the context of an arbitrary stratum or interest.

N denotes the total number of persons for the stratum of interest.

M denotes the total number of ED's for the stratum.

m denotes the number of ED's to be sampled from the stratum. This is the number to be determined.

V^2 denotes the relative variance of the per capita income estimate based on a sample of m ED's.

B^2 denotes the between ED component of V^2.

W^2 denotes the within ED component of V^2.

$\overline{N} = N/M$ is the average ED size in the stratum.

N_i denotes the 100-percent census count of persons for the i^{th} ED.

Y_{ij} denotes the person income for the j^{th} person in the i^{th} ED.

y_{ij} denotes the person income for the j^{th} sample person in the i^{th} ED.

n_i denotes the number of sample persons in the i^{th} ED.

$\overline{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$ is the sample per capita income for the i^{th} ED.

$\overline{Y}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$ is the per capita income for the i^{th} ED.

$\overline{\overline{Y}} = \sum_{i=1}^{M} \sum_{j=1}^{N_i} Y_{ij}/N$ is the per capita income for the entire stratum.

$$B^2 = \frac{1}{\overline{N}^2} \frac{\sum_{i=1}^{M} N_i^2 (\overline{Y}_i - \overline{\overline{Y}})^2}{(M-1)\ \overline{\overline{Y}}^2},$$

$$W^2 = \frac{\sum_{i=1}^{M} N_i S_i^2}{N\overline{\overline{Y}}^2}, \text{ where}$$

$$S_i^2 = \frac{\sum_{j=1}^{N_i} (Y_{ij} - \overline{Y}_i)^2}{N_i - 1}$$

b. Assumptions

Several assumptions were made in determining the ED sample size. They were:

1) The first stage selection of ED's within strata would be made with simple random sampling without replacement;

2) The second stage selection of persons within ED's would be made independently by ED and by simple random sampling without replacement. Furthermore, it was assumed that the second stage sampling fraction would be 1-in-6; 2/

3) It was also assumed that the population relative variance of per capita income would be relatively constant from stratum to stratum; and

4) The increase, if any, in the population relative variance of per capita income from 1970 to 1980 could be measured by comparing 1969 and 1979 Current Population Survey (CPS) data.

Assumption 1) is thought to be conservative as a well designed systematic PPS sampling scheme will tend to yield CV's that are lower than those that would arise under the simple random sample scheme. Assumption 2) ignores the effect of within household clustering of income on the variance since the sampling unit for the census sample is the housing unit. Since the within ED variance component is shown later to be quite small in comparison to the between ED variance component, ignoring this clustering effect is of little consequence. Assumptions 3) and 4) were made upon the review of a sample set of empirical data.

c. Derivation of Sample Size

The estimate of per capita income for a specific stratum is given by

$$\overline{\overline{y}} = \frac{\sum_{i=1}^{m} N_i \overline{y}_i}{\sum_{i=1}^{m} N_i}$$

Given assumptions 1) and 2) above, the relative variance of $\overline{\overline{y}}$, the estimate of per capita income, V^2 (Hansen at el [1] page 253) may be expressed as:

$$V^2 = \left(\frac{1}{m} - \frac{1}{M}\right) B^2 + \frac{5}{\overline{N}m} W^2,$$

Given the above formulation of V^2, clearly a value of $m < M$ can be found to make V^2 as small as desired 3/. For any given level of desired CV, say C, the corresponding value of m which yields a value of V less than or equal to C is

$$m \geq \frac{B^2 + \frac{5W^2}{\overline{N}}}{C^2 + \frac{B^2}{M}}, \text{ and } m \leq M \qquad (1)$$

For each stratum, the value of $\overline{N}$ and M are available from 1980 Census preliminary complete count data. If values of B^2 and W^2 could be approximated, the sample size required to achieve any level of C could be found. Values of B^2 and W^2 were generated from the 1970 census sample records that were used to form the study population for 1980 census weighting and variance estimation studies (Woltman et at [2]). The computed values of B^2 and W^2 for some places in California are shown below in Table 1. As may be seen from Table 1 that B^2 is generally 0.25 or less and that 3.1 may be an acceptable value to assume for W^2.

Table 1--1970 Census Between E.D. Relvariance B^2 and Within E.D. Relvariance W^2 by Selected Places in California

Place Name	Total Population	Number of E.D.'s	Average Persons Per E.D.	Between E.D. Relvariance (B^2)	Within E.D. Relvariance (W^2)
Madera	16,000	18	895	0.23	3.0
Novato	30,200	49	616	0.19	2.9
San Rafael	38,400	48	800	0.06	3.0
Costa Mesa	70,600	67	1,054	0.10	2.5
Garden Grove	120,300	124	970	0.05	2.5
Laltabra	40,700	45	904	0.07	2.7
Seal Beach	24,200	32	757	0.02	1.9
Yorba Linda	11,900	13	912	0.02	3.2
Auburn	6,600	12	550	0.07	2.5
San Clemente	16,400	21	779	0.09	2.7
San Diego	670,000	523	1,281	0.23	3.1
San Francisco	682,500	755	904	0.23	3.0
San Jose	436,200	577	756	0.13	2.6

Population CV's for per capita income for 37 of the 38 SMSA's with one million or more persons were computed from 1970 census data. We found that these per capita income CV's were remarkably consistent from SMSA to SMSA-ranging from 1.2 to 1.4 percent-giving some justification for assumption 3), above. Population CV's for mean persons income 4/ and per capita income were also calculated from 1969 and 1979 data from the Census Bureau's Current Population Survey. A comparison of these CV's indicated that while there did not appear to be an increase in the per capita income CV's, there was some increase for mean persons income. This increase was seen to be as large as 1.088. To be conservative, it was assumed that the values of B^2 and W^2 had increased by $(1.088)^2$ between 1970 and 1980. Substituting these values of B^2 and W^2 into equation (1), and setting the value of C to 0.05 or 0.035 yielded the desired sample sizes.

The achieved sample sizes were relatively constant from stratum to stratum for those strata which corresponded to SMSA's-ranging from about 118 to 128 ED's. The sample sizes achieved for those strata corresponding to states without a portion of one of the 38 large SMSA's were slightly more variable 5/ ranging in general from 200 to about 260 ED's.

Finally, it should be noted that deviations in the value of W^2 will have little impact on the resulting sampling sizes. This follows since $W^2/\bar{N}$ is quite small. However, deviations in the value of B^2 may have a significant effect on the size of the sample required to achieve the desired levels of reliability.

2. Sample Selection Procedures

The selection procedures were performed by systematically sampling ED clusters with PPS, where the measure of size was the preliminary 100-percent census population count for the ED cluster. This was accomplished as follows:

1. First the ED's were sorted within each stratum into county, place, MCD (minor civil division), and census tract order;

2. The ED's were next arranged into clusters of one or more ED's so that each cluster contained a number of persons that was at least one-fourth of the average ED size for the stratum. The clustering of ED's was to eliminate the possibility of selecting a small ED in the sample and assigning this ED a huge weight. The rationale of imposing the minimum ED cluster size was to ensure that the maximum of initial weights was consistent with other census weighting objectives. The clustering was performed by repeatedly passing the sorted file of ED's. On the first pass, an ED that was too small was combined with the smaller of the ED directly preceeding or succeeding it on the file. On each subsequent pass this procedure continued with ED clusters that were too small;

3. ED clusters were next systematically sampled with probability proportionate to the size of each ED cluster so that the desired number of clusters fell in sample.

4. Additionally, a systematic sample of ED's containing only vacant housing unit was selected.

The final sample included a total of 17,143 ED's selected in the manner outlined above. This is approximately 5 percent of all ED's in the United States. However, only slightly more than one and a half percent of the housing units and persons in the United States were included in the sample.

III. Estimation

The estimation procedure for the ENS was based on the assignment of weights to sample records of persons and housing units. For each of the 98 strata, an estimate of a specific stratum sample total was produced by summing the weights of sample persons or housing units in the stratum that possess the characteristic for which the total is being estimated. The weights used in producing the estimates were obtained from a raking ratio (iterative ratio) procedure. (Raking ratio procedures have also been examined by a wide group of authors, for example Brackstone and Rao [3].) The raking ratio procedure for 1980 census was discussed and compared empirically with other alternative procedures (Kim et. al. [4]).

The weighting arrays used for the raking ratio procedure are shown in Appendix 3 in [4]. For population characteristics the weighting array was three dimensional using column categories defined by race, origin, sex and age, row categories defined by family type and size of household, and categories for the third dimension defined by householder/nonhouseholder status.

For occupied housing unit characteristics, the weighting array was a matrix using column categories defined by tenure, race/ origin of the householder, and value/rent, and row categories defined by family type and size of household.

Briefly, the estimation was performed in the following steps:

a) inflate the ED sample counts by an initial weighting factor that is a product of the inverse of the sample selection probability of ED's and the inverse of the observed Census sampling rate within the ED. The observed sampling rate refers to person, occupied housing unit or vacant housing unit depending on which sample counts are being inflated using the rate.

b) cumulate the initial inflated sample count into the weighting array and obtain the 100-percent counts for each marginal category of the weighting array;

c) collapse the weighting array, if necessary;

d) adjust the initial inflated sample interior cell counts of the weighting array based on the 100-percent marginal counts;

e) and finally, allocate the adjusted interior cell counts to each sample record, as weights.

The mathematics and procedures used for the estimation procedures are very similar to those described in [2]. However, due to the large size of the sample that was to be adjusted in each stratum, it was necessary to use the following collapsing criteria.

a) The uninflated marginal sample count for the marginal category must be at least 30 for the person weighting array and 15 for the housing unit weighting matrices;

b) The 100-percent marginal count must be at least 300 for the person weighting array and 150 for the housing unit weighting matrices;

c) The ratio of the 100-percent marginal count to the initial inflated marginal count must be greater than 0.5 annd less than 2.5.

These criteria were developed to insure collapsing probabilities that would be consistent with the full census sample weighting procedures [2].

IV. ESTIMATION AND PRESENTATION OF SAMPLING ERRORS

The sampling errors were to be presented in the form of design effects (e.g., Kish [5]). The design effect is the ratio of the variance of an estimate for a specific sampling design to the variance of the corresponding estimate based on a simple random sample of the same sample size. These presentations of sampling errors were employed in previous census publications (Waksberg et. al. [6]), and in publications from other such surveys, e.g., Kalton et. al. [7].

To prepare the ENS sampling error estimates, it was first necessary to estimate the sampling error for each of the 1120 data items to be published. It was then highly desirable to combine the 1120 data items into a small number of groups for publication. This was accomplished by computing design effects for each of the 1120 data items, and grouping the data items based on homogeneity of design effect. Furthermore, it was desirable to combine the ninety areas for which data was to be published (United States, states and the District of Columbia, and the 38 SMSA's with one million or more persons) into groups with similar design effects. This resulted in the formation of 36 data item categories for which design effects were published for 11 groups of publication areas. These appear in Table D in [8].

The square root of the design effect is less affected by extreme values, and may, therefore, be preferred when an average design effect is required (Kish [5] p. 579). Hence, the design effect published, and as discussed later in this paper, is actually the square root of the design effect as commonly defined. The design effects were averaged over the data items in the group for each publication area. The group design effects were then averaged over the publication areas in the publication area group and the average was used in determining the standard error for all data items in the data item group and for all publication areas in the publication area group. The actual methodology used in the ENS publications to present the estimated sampling errors, perform the variance estimation and to calculate and group the design effects is described as follows:

1. Presentation of Sampling Errors in the ENS Publications

The ENS report [8] contains four tables for estimating standard errors. Two of the tables show the unadjusted standard errors associated with a simple random sampling design.

One of the table applies to estimates of total, the other one to percents. They show the values of

$$\hat{SE}(\hat{X}) = \sqrt{f \hat{X} (1 - \hat{X}/N)} \text{ and } \hat{SE}(\hat{p}) = \sqrt{\frac{f\hat{X}}{N^2} (1 - \hat{X}/N)},$$

Where N is total population, $\hat{X}$ is the estimate of characteristic total, p is the estimated percentage, and f is the averaged value of initial weighting factors f = 62.39≐62, approximately the inverse of the overall combined sampling fraction for the country.

The third table shows groupings of publication areas (U.S., states and the District of Columbia, and 38 SMSA's with one million or more persons) that must be used in conjunction with the fourth table. The fourth table reflects the design effects for publication areas; it provides factors to be applied to either of the first two tables. The data user is required to obtain the unadjusted standard error from either of the first two tables, then find the publication area of interest in the third table and obtain its publication area group number, and finally use the fourth table to obtain the factor for the type of data item of interest (e.g., labor force status, veteran status, school enrollment) and the publication area group given in the third table. Then the data user multiplies the design effect shown from the fourth table by the unadjusted standard error to obtain an estimate of the standard error of the ENS statistic of interest. Tables A through D in [8] illustrate how these tables appeared in the ENS report.

2. Variance Estimation

For a given data item and stratum the variance estimator used for the ENS is

$$\hat{Var}(\hat{X}) = \frac{n}{n-1} \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{m_i} \hat{X}_{ij} \right) - \frac{\hat{X}}{n} \right]^2, \text{ and}$$

where n denotes the number of sampled ED clusters in the stratum;

m_i denotes the number of sampled ED's in the i^{th} ED cluster;

$\hat{X}_{ij}$ denotes the weighted data item total resulted from raking ratio procedure for the j^{th} ED in the i^{th} ED cluster in the stratum; and

$$\hat{X} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \hat{X}_{ij}.$$

Isaki and Pinciaro [9] investigated the performance of this and other variance estimators for sampling designs similar to that used in the ENS. In general, this estimator will have a positive (negative) bias if the within systematic sample intraclass correlation between the first stage unit totals is negative (positive).

3. Calculation and Grouping of Design Effects

a. Calculation of Individual Design Effects

For a given data item and publication area, the design effect was calculated as follows:

$$F_1 = \left[\frac{\sum_{k=1}^{M} \hat{Var}(\hat{X}_k)}{62 \left(\sum_{k=1}^{M} \hat{X}_k \right) \left(1 - \frac{\sum_{k=1}^{M} \hat{X}_k}{\sum_{k=1}^{M} N_k} \right)} \right]^{1/2}$$

where M denotes the number of strata in the publication area.

$\hat{Var}(\hat{X}_k)$ denotes the variance estimates described above for the k^{th} stratum in a particular publication area.

$\hat{X}_k$ denotes the weighted data item total for the k^{th} stratum in a publication area.

N_k denotes the complete census count (of persons or housing units) for the k^{th} stratum in a publication area.

b. Combining the Design Effects for Data Item Groups Within Publication Areas

The individual data items were grouped within each publication area based on subjective judgement about the similarity of design effects.

For each group of data items within each publication area, a group design effect was calculated as:

$$F_{1G} = \sum_{\ell=1}^{N_G} P_{G\ell} F_{1G\ell}$$

where N_G denotes the number of data items in the G^{th} group of data items.

$F_{1G\ell}$ denotes the individual design effect factor for ℓ^{th} data item in the G^{th} group of data items

$$P_{G\ell} = \sum_{k=1}^{M} \hat{X}_{k\ell} \Big/ \sum_{\ell=1}^{N_G} \sum_{k=1}^{M} \hat{X}_{k\ell}$$

where M denotes the number of strata in the publication area and

$\hat{X}_{k\ell}$ denotes the item total for ℓ^{th} data item and the k^{th} stratum.

c. Grouping Publication Areas

89 publication areas including all states (and the District of Columbia) and the 38 SMSA's with one million or more persons were grouped into 10 groups. United States was put into a separate group. The groupings were done using a clustering procedure. The clustering procedure as initially developed by Friedman and Rubin [10], and subsequently modified by Jewett [11], is essentially a "hill-climbing pass" algorithm. The variables used for clustering were group design effects of eight population characteristics; school enrollment, years of school completed, labor force status, family income, person poverty status, unemployment, occupation, and unrelated individuals income. Group identification for each publication area is given in Table C in [8].

For a publication area group, the group design effect for a given group of data items was obtained as a weighted average over all publication areas in the same group where the 100-percent census count of persons was used as the weight. The averaged group design effects for each publication area group are given in Table D in [8].

V. Coefficient of Variation of Per Capita Income Estimates

To evaluate the assumptions used in the sample design the CV's (coefficient of variation) on estimates of per capita income from the ENS were estimated for all states (including the District of Columbia) and the 38 SMSA's with one million or more persons. The documentation for calculating these CV's is given [12]. Table 2 gives the desired CV on estimates of per capita income as specified in the sample design and the estimated CV on estimates of per capita income as achieved in the ENS.

It may be seen from Table 2, that the CV's on estimates of per capita income from the ENS were less than the desired CV for all states and SMSA's, except for the District of Columbia, Dallas-Fort Worth, Houston, and Miami SMSA's. Except for the Miami SMSA, the differences were not of great concern and were tolerable considering the rather broad assumptions used to design the sample and the fact that the variance estimator may be biased upward. For Miami SMSA the CV from the ENS was about two times the desired level. Further review indicated that this resulted from an unusual distribution of income in the selected ED clusters for the Miami SMSA.

1/ An Enumeration District (ED) may be thought of as approximating one census enumerator's work load.

2/ It is conservative that the second stage fraction was assumed to be 1-in-6. The sample rate of census sample of persons and housing units within ED was either 1-in-6 or 1-in-2.

3/ This is possible since $W^2/N \doteq 0$

4/ Mean person income is defined as the aggregate person income divided by total persons with income.

5/ Detailed data on 1970 and 1980 income CV's and the actual achieved sample sizes may be found in [13].

References

[1] Hansen M. H, Hurwitz W. N and Madow W.G. "Sample Survey Methods and Theory Volume 1 Methods and Applications" John Wiley and Sons, Inc., 1953.

[2] Woltman, H., Miskura, S., Thompson, J., and Bounpane, B., "1980 Census Weighting and Variance Estimation Studies, Design and Methodology" Proc. Sect. Survey Res. Meth., Amer. Stat. Assoc. 1981, 164-169.

[3] Brackstone G.J. and Rao, J.N.K. "An investigation of raking ratio estimators" to appear in Sankkya B, 1980.

[4] Kim, J., Thompson, J. Woltman, H., and Vajs, S. "Empirical Results from the 1980 Census Sample Estimation Study" Proc. Sect. Survey Res. Meth., Amer. Stat. Assoc., 1981, 170-175.

[5] Kish, L. "Survey Sampling", John Wiley and Sons, Inc., 1965.

[6] Waksberg, J., Hanson, R. and Bounpane, P. "Estimation and Presentation of Sampling Errors for Sample Data from the 1970 U.S. Census" Proceedings of the 39th session, International Statistical Institute, Vienna, August 1973, 67-82.

[7] Kalton G and Blunden R. M. "Sampling Errors in the British General Household Survey" Proceedings of the 39th Session International Statistical Institute, Vienna, August 1973, 83-97.

[8] Supplementary Report Provisional Estimates of Social Economic, and Housing Characteristic PHC80-S1-1 Government Printing Office, April 1982.

[9] Isaki C.T. and Pinciaro S.J. "Numerical Comparison of Some Estimators of Variance Under PPS Systematic Sampling" Proc. Sect. Survey Res. Meth., Amer. Stat. Association 1977, 308-313.

[10] Friedman H.P. and Rubin J. "On Some Invariant Criteria for Grouping Data," Journal of American Statistical Association, 1967, 1159-1178.

[11] Jewett, Robert S. "A Method of Stratification when Size Constrainst are Imposed", Interval Census Bureau Document (1980).

[12] Miskura, Susan M., "Computer Specifications for Computing the CV on Estimates of Per Capita Income from the Early Sample Publications National Sample" 1980 Decennial Census Evaluation and Research Memorandum Series SMD Number 38-9, Internal Census Bureau Memorandum for Robert T. O'Reagan, Prepared by Milton C. Fan, January 18, 1982.

[13] Fan, Milton C., "Documentation of Sample Size Determination for the Preliminary Sample Publications" Internal Census Bureau Memorandum for Henry F. Woltman, April 28, 1981.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of a number of colleagues for their help in preparing this paper. In particular, we wish to thank Paul Zeisset, David W. Chapman, Robert Fay and David D. Chapman for their comments, Don Dalzell for his dedicated computer programming support in sample selection and weighting operations, and Robert Jewett and Teresa Passalacqua for their excellent computer programming in the variance estimation operation. The authors also owe a great deal of thanks to Ms. Loretta Palmer for typing the manuscript.

Table 2--Desired CV on Estimates of Per Capita Income From Sample Design and, Estimates of Per Capita Income and its Estimated CV From ENS

States/1,000,000+ SMSA	Desired CV of PCI Estimate (%) From Sample Design	Estimated CV of PCI Estimate (%) From ENS	Per Capita Income Estimates From ENS	States/1,000,000+ SMSA	Desired CV of PCI Estimate (%) From Sample Design	Estimated CV of PCI Estimate (%) From ENS	Per Capita Income Estimates From ENS
SMSA's							
Anaheim-Santa Ana-Garden	5.0	3.9	9,501	Colorado	3.5	2.8	8,138
Grove, Ca.				Connecticut	3.5	2.6	8,458
Atlanta, Ga.	5.0	4.9	8,338	Delaware	3.5	3.0	7,392
Baltimore, Md.	5.0	3.5	7,501	District of Columbia	5.0	6.3	9,016
Boston, Mass.	5.0	3.1	8,012	Florida	3.5	3.0	7,593
Buffalo, N.Y.	5.0	2.5	7,142	Georgia	3.5	3.3	6,682
Chicago, Ill.	5.0	3.6	8,033	Hawaii	3.5	3.1	7,790
Cincinnati, Ohio-Ky.-Ind.	5.0	3.1	7,621	Idaho	3.5	1.9	6,224
Cleveland, Ohio	5.0	3.2	8,135	Illinois	3.5	2.4	7,789
Columbus, Ohio	5.0	3.8	7,561	Indiana	3.5	2.1	7,088
Dallas-Fort Worth, Texas	5.0	5.1	8,213	Iowa	3.5	1.7	7,224
Denver-Boulder, Colo.	5.0	4.1	9,193	Kansas	3.5	2.4	7,199
Detroit, Mich.	5.0	3.4	8,282	Kentucky	3.5	3.1	5.935
Fort Laud.-Hollywood, Fla	5.0	3.5	8,791	Louisiana	3.5	2.7	6,321
Houston, Texas	5.0	5.8	8,999	Maine	3.5	1.6	5,846
Indianapolis, Ind.	5.0	3.8	7,919	Maryland	3.5	2.3	8,154
Kansas City, Mo.-Kans.	5.0	2.5	8,302	Massachusetts	3.5	2.0	7,411
Los Angeles-Long Beach, Ca.	5.0	4.6	8,310	Michigan	3.5	2.4	7,708
Miami, Fla.	5.0	11.2	8,766	Minnesota	3.5	2.1	7,529
Milwaukee, Wis.	5.0	3.4	8,361	Mississippi	3.5	2.5	5,327
Minn.-St. Paul, Minn.-Wis.	5.0	3.2	8,898	Missouri	3.5	2.0	6,882
Nassau-Suffolk, N.Y.	5.0	4.2	8,947	Montana	3.5	2.0	6,698
New Orleans, La.	5.0	4.2	6,798	Nebraska	3.5	2.2	7,034
New York, N.Y.-N.J.	5.0	4.7	7,861	Nevada	3.5	2.4	8,561
Newark, N.J.	5.0	4.1	8,354	New Hampshire	3.5	1.4	6,931
Philadelphia, Pa.-N.J.	5.0	3.5	7,554	New Jersey	3.5	1.8	8,113
Phoenix, Ariz.	5.0	3.6	7,751	New Mexico	3.5	2.7	6,137
Pittsburg, Pa.	5.0	3.6	8,054	New York	3.5	2.7	7,469
Portland, Oreg.-Wash.	5.0	2.1	8,352	North Carolina	3.5	2.4	6,177
Riverside-San Bern-Ont, Ca.	5.0	2.7	7,146	North Dakota	3.5	1.8	6,470
Sacramento, Ca.	5.0	3.2	7,922	Ohio	3.5	1.9	7,262
St. Louis, Mo.-Ill.	5.0	2.7	7,517	Oklahoma	3.5	2.8	6,933
San Antonio, Texas	5.0	5.6	6,189	Oregon	3.5	1.6	7,431
San Diego, Ca.	5.0	3.4	7,878	Pennsylvania	3.5	1.9	7,172
San Francisco-Oakland, Ca.	5.0	4.7	9,815	Rhode Island	3.5	2.0	6,800
San Jose, Ca.	5.0	3.3	9,613	South Carolina	3.5	2.8	6,191
Seattle-Everett, Wash.	5.0	2.2	9,192	South Dakota	3.5	2.2	5,902
Tampa-St. Petersburg, Fla.	5.0	4.7	7,535	Tennessee	3.5	2.9	6,180
Washington, D.C.-Md.-Va.	5.0	2.3	10,223	Texas	3.5	2.6	7,266
				Utah	3.5	2.2	6,399
STATES				Vermont	3.5	1.8	6,245
Alabama	3.5	2.6	5,975	Virginia	3.5	2.4	7,549
Alaska	3.5	2.9	10,171	Washington	3.5	1.6	7,922
Arizona	3.5	2.6	7,030	West Virginia	3.5	2.2	6,243
Arkansas	3.5	2.2	5,467	Wisconsin	3.5	1.9	7,272
California	3.5	1.8	8,296	Wyoming	3.5	1.9	7,982