CENSUS 2000 ICM: STRATIFICATION AND POSTSTRATIFICATION

Eric Schindler and Richard Griffin
Eric Schindler, Bureau of the Census, Washington, DC

KEY WORDS: Raking, Heterogeneity

In 1990, the synthetic estimation technique developed for census adjustment assumed that within poststrata undercount rates are constant across all subpopulations. A poststratum, the finest level for which direct coverage estimates are produced, is usually defined as a function of demographic and/or geographic characteristics. Poststrata are defined so as to minimize the impact of failure of the synthetic assumption; that is, to minimize heterogeneity within poststrata. This paper will use 1990 census and Post Enumeration Survey (PES) data to assess the use of raking to create additional poststratification cells for the Census 2000 Integrated Coverage Measurement (ICM) in terms of variance and heterogeneity. Pearson correlations are used to assess heterogeneity at the poststratum level.

I. INTRODUCTION

In the 1990 Post Enumeration Survey (PES), data from 150,000 housing units were used to estimate the coverage error in the census counts for 357 poststrata defined by 4 Census regions, 3 metropolitan area size classes, 5 race/Hispanic origin groups, 2 tenure categories, and 7 age/sex combinations. Major changes are being planned for the Census 2000 Integrated Coverage Measurement (ICM) program. In order to improve the acceptability of the ICM estimates in response to recent Supreme Court rulings, the total population estimate for each state will be based only on data from within the state. To accommodate this requirement for direct state estimates, the total sample size is being increased to 750,000 housing units. The sample size for most states with fewer than 10,000,000 residents will be about 10,000 to 15,000 housing units. These sample sizes should be sufficient to produce coefficients of variation of about 0.50% for each state. Larger sample sizes will be allocated to the seven larger states so as to produce standard errors of about 60,000 in each state. Experience from 1990, when the number of poststrata had to be reduced from 1392 to 357 to obtain acceptable standard errors, showed that 10,000 to 15,000 housing units will be large enough to provide reliable estimates for only about 30 poststrata.

Within each state, it will be necessary to produce reliable estimates for race/origin groups by age/sex by tenure for substate areas such as large cities of metropolitan statistical areas. If the 5 PES race/origin groups and 7 PES age/sex groups and 2 tenure categories are used and 6 areas are defined, a total of (5x7x2x6=)420 poststrata would be required, far more than the sample will support.

This paper discusses an empirical study in which iterative proportional fitting, or raking, is employed to develop coverage error estimates and corrections for large numbers of poststrata. Since developed by Deming in 1940, raking has been used to weight data from the census long form. Purcell and Kish (1980) suggest a similar approach for more general postcensal local area estimates. While previous raking approaches adjusted the results of a survey to match the census, this approach corrects the initial enumeration estimates to match the final census estimates produced by the ICM coverage quality check survey.

Section II discusses the available data and the construction of simulations. Section III discusses results. Section IV lays the groundwork for possible use in 2000.

II. DESIGN AND ESTIMATION

POSTSTRATUM DEFINITION

The 1990 PES data for California are used[1]. There are about 11,000 housing units in the PES sample, about as many as the typical small state sample will have in 2000. Thus, the discussion of this paper will be more applicable to Colorado or North Carolina in 2000 than to California. California's sample size will be 80,000 to 100,000 housing units and both the sample design and estimation will be more complex. Five variables are used in poststratification:

- The five PES race/Hispanic origin groups: (1) non-Hispanic white or other, (2) Black, (3) Hispanic white or other, (4) American Indians (not living on reservations), and (5) Asians and Pacific Islanders.
- The seven PES age/sex categories: (1) under 18, (2) male 18-29, (3) female 18-29, (4) male 30-49, (5) female 30-49, (6) male 50+, and (7) female 50+.
- The two PES tenure categories: (1) owner and (2) renter
- Based on the location of the available PES sample, six substate areas in California: (1) Los Angeles City, (2)

The authors are mathematical statisticians in the Decennial Statistical Studies Division. This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

[1] American Indians living on reservations were dropped from this study. This population will be sampled separately in 2000 and the results will be added to those for the non-reservation or "urban" American Indians.

the remainder of Los Angeles County, (3) the central cities of other MSAs with over 250,000 residents, (4) the central cities of MSAs with fewer than 250,000 residents, (5) suburban areas of the MSAs in (3) and (4), and (6) non-MSA areas. It is assumed that the state was stratified by these areas for sample selection.

- The last variable, "household stability," identifies households which may have better coverage. A household is "stable" if either (1) there is only one resident and that resident is over age 50, or (2) there are 2 to 7 residents, the first two residents are over age 30 and are of the opposite sex[2], and any additional residents are under age 18. Households with more than seven residents or with any resident between 18 and 29 are never "stable." This variable imitates a similar variable found to be well correlated with mail return rates (Word, 1997). The 40% of residents in households which are "stable" are undercounted by about 1.6%. The 60% in "non-stable" households are undercounted by about 5.5%. The fact that all 18-29 year olds are in "non-stable" households accounts for about half of this difference. The difference holds across tenure and across race and age groups for some but not all of the geographic areas.

RAKING

The raking process is carried out in two dimensions with two sets of poststrata. The poststratum sets were designed not to exceed about thirty poststrata and to include all variables from the 1990 PES plus household stability. The inclusion of an additional variable quantifying the difficulty of collecting a neighborhood is being considered.

- Poststratum Set I: 5 race/origin * 7 age/sex = 35 poststrata. Because of the limited number of American Indians in the sample, all American Indians were included in the same poststratum. Therefore, there were only 29 final poststrata. In 2000 additional collapsing of poststrata might be necessary in some states to eliminate cells with little or no ICM sample.
- Poststratum Set J: 6 areas * 2 tenure * 2 household stability = 24 poststrata[3].

Potentially there are 600 non-zero cells in the two dimensional matrix.

Four matrices of size 29*24 are defined. The cells in these matrices are called "interior" cells. Summations to the 29 or 24 subtotals defined by the two poststratum sets are call "marginals" or marginal subtotals. The entry in the I,J cell of the first matrix is the estimated weighted number of initial enumeration persons, $E_{I,J}$, simultaneously in poststrata I and J. The entry in the I,J cell of the second matrix is the estimated weighted number of initial enumeration persons who are correctly enumerated, $CE_{I,J}$. The entry in the I,J cell of the third matrix is the estimated weighted number of persons from the independent PES enumeration, $P_{I,J}$. The entry in the I,J cell of the fourth matrix is the weighted estimated number of persons in $P_{I,J}$ who can be matched to the initial enumeration, $M_{I,J}$.

Marginal totals for each of the dimensions are obtained by adding over the other dimension: $E_I = \sum_J E_{I,J}$; $CE_I = \sum_J CE_{I,J}$; $P_I = \sum_J P_{I,J}$; $M_I = \sum_J M_{I,J}$,and similarly for the J dimension. The Dual System Estimate (DSE), the DSE undercount rate UC, and the DSE coverage factor (CF) for each marginal cell are: $DSE_I = CE_I \times P_I / M_I$, $UC_I = \frac{DSE_I - INIT_I}{DSE_I}$, and $CF_I = \frac{CE_I}{INIT_I} \times \frac{P_I}{M_I}$ [4], where $INIT_I$ would be the initial enumeration estimate for the I marginal cell. $INIT_I$ is not available so E_I is used as an approximation.

The objective of the raking process is to transform the 29x24 matrix $Q_{1,I,J} = E_{I,J}$ into another 29x24 matrix such that the appropriate sums are exactly equal to one set of marginal DSEs and almost equal to the other set of marginal DSEs. A new matrix, $Q_{2,I,J}$, is obtained by multiplying the sum of the I. entries in $Q_{1,I,J}$ by DSE_I over the sum of the I. entries. This is equivalent to the traditional DSE with 29 poststrata with all I. interior cells of the matrix having the same coverage factor. Matrix, $Q_{3,I,J}$, is obtained by multiplying the sum of the .J entries in $Q_{2,I,J}$ by DSE_J over the sum of the .J entries. Matrix, $Q_{4,I,J}$, is obtained by multiplying the sum of the I. entries in $Q_{3,I,J}$ by DSE_I over the sum of the I. entries. $Q_{5,I,J}$ and $Q_{6,I,J}$ are obtained by repeating the operations for $Q_{3,I,J}$ and $Q_{4,I,J,K}$. A 29x24 matrix of coverage factors is estimated by: $CF_{I,J} = Q_{6,I,J}/Q_{1,I,J}$. Since the last operation controls to the marginal totals for Poststratum Set I, the 29 I. sums of the $Q_{6,I,J,K}$ entries over the 600 (actually 595 since 5 possible non-zero cells have no sample) non-zero interior cells are exactly equal to the

[2] If "relationship" were available, the first two residents would have to be married and other residents would have to be their children for the household to be considered "stable."

[3] Three or more dimensional raking including partially overlapping poststrata (such as area by tenure by minority/non-minority in the second dimension and race by tenure by household stability in the third) may give better direct estimates for the "non-official" marginals, but collapsing to eliminate cells with little or no ICM sample would be much more difficult.

[4] A small adjustment for whole person imputations in the initial enumeration (these persons are not included in the E-sample file) has been omitted.

29 DSEs defined for Poststratum Set I. The interior cells which add up to each cell on the marginal have heterogeneous coverage factors which reflect coverage differences for the variables not in Poststratum Set I, namely area, tenure, and household stability.

With the raking approach, synthetic estimation at any level proceeds by multiplying the appropriate coverage factors for the 600 interior cells times the corresponding initial enumeration estimates $E_{I,J}$. Compared with the 29 coverage factors available from traditional dual system estimation, raking provides the opportunity for much finer distinctions of subpopulations. Instead of a single factor for Asian males 30-49, there are 24 factors based on area, tenure, and household stability.

STANDARD ERROR ESTIMATES

Standard errors are estimated by the jackknife procedure. The PES was a cluster sample with 358 block clusters selected in California. One block cluster at a time is removed and the full raking process is repeated. The standard errors of the coverage factors are estimated by:

$$SE(CF_{I,J}) = \sqrt{\sum_{m=1}^{358} (\tilde{CF}_{m,I,J} - \bar{CF}_{m,I,J})^2} \quad \text{or}$$

$$SE(CF_{I}) = \sqrt{\sum_{m=1}^{358} (\tilde{CF}_{m,I} - \bar{CF}_{m,I})^2} \quad \text{for I=1 to 29, where}$$

$\tilde{CF}_{m,I}$ or $\tilde{CF}_{m,I,J}$ is the coverage factor without block m, and the averages are taken over all blocks, and similarly for $SE(CF_J)$. Note that $SE(DSE_\cdot) = SE(CF_\cdot) \times INIT_\cdot$.

III. RESULTS

FOR MARGINAL SUBTOTALS

By design, the final estimates and standard error estimates based on the raking are exactly equal to the direct DSEs for Poststratum Set I defined by 5 race/origin groups and 7 age/sex groups. The corresponding totals for the state are equal within rounding error at 33,099,703. Since these estimates are closest to the direct estimates required by law for reapportionment and redistricting, they will be considered the "primary" estimates for this paper. Without raking, synthetic estimates would be made using the 29 coverage factors from this poststratification. On the other hand, if Poststratum Set J were used as the only stratification, the total estimate would be 28,271 higher at 33,127,974. These direct estimates are ratio adjusted to the "official" estimate 33,099,703.

The purpose of the raking approach is to permit good estimates for variables that cannot be included in the primary poststratification. For example, for the six areas defined in California, Table 1 compares three DSEs: (1) for the direct stratification using Poststratum Set J (the target estimates, since this is probably close to the best estimate that can be made for these substate areas with the available sample size), (2) for the raking process, and (3) for synthetic estimates from the primary poststratification using Poststratification Set I. The undercount rates for the raking process are within 0.03% of those for the directly estimated DSEs for each of the six regions, less that 1% of the total coverage correction. On the other hand, the differences compared to the target direct estimates using only the primary poststratification and synthetic estimation are as much as 1.46% or 27% of the total coverage correction. That is, in Los Angeles County and the non-MSA areas, synthetic estimation found only about three-fourths of the undercount found by either direct estimation or raking. Synthetic estimation also found about 12% less undercount in Los Angeles City, but 20% more undercount in the suburban areas excluding Los Angeles county and 10% more undercount in the large central cities.

Table 1: DSEs and Undercount Rates for Six Regions in California with Three Poststratification Schemes

	Wgtd E	Dir DSE	Raked DSE	Synth DSE
CA	31804737	33,099,703		
Los Angeles City	3,726,837	3,935,309	3,934,253	3,908,532
	UC Rate	5.30%	5.27%	4.65%
	% Diff in UC Rate		-0.48%	-12.25%
Los Angeles County	3,852,927	4,085,734	4,085,202	4,023,577
	UC Rate	5.70%	5.69%	4.24%
	% Diff in UC Rate		-0.22%	-25.57%
Large Central Cities	4,209,197	4,366,938	4,367,096	4,383,909
	UC Rate	3.61%	3.62%	3.99%
	% Diff in UC Rate		0.10%	10.33%
Small Central Cities	2,777,721	2,922,231	2,921,530	2,911,776
	UC Rate	4.95%	4.92%	4.60%
	% Diff in UC Rate		-0.46%	-6.90%
MSA not Central City	15839863	16331425	16332983	16430545
	UC Rate	3.01%	3.02%	3.60%
	% Diff in UC Rate		0.31%	19.44%
Not MSA	1,398,192	1,458,065	1,458,639	1,441,362
	UC Rate	4.11%	4.14%	3.00%
	% Diff in UC Rate		0.92%	-27.06%

Examining how well the raked estimates match the DSEs along all 24 marginals for Poststratum Set J, Graph 1 depicts the undercount rates for the three sets of estimates. The raked undercount rates (triangles) always overlap the target direct undercount rates (squares), but the synthetic

undercount rates (diamonds) vary substantially. Table 2 shows the undercount rates and standard errors for direct estimates, raked estimates, and synthetic estimates. The direct DSEs based on Poststratum Set J give nearly unbiased estimates of the true populations. The raked estimates of the undercount rates are all within 0.14% (average difference 0.07%) of the direct estimates. The synthetic estimates based on Poststratum Set I differ from the direct estimates by as much as 5.12% (average difference 2.26%). Since the average undercount rate in the marginal cells is only 4.32%, missing the target value by 2.26% cannot be considered very good.

The jackknife standard errors for the raked estimates average minimally smaller (less than 1%) than the standard errors for the direct estimates. However, the synthetic estimates are based on larger samples, so their standard errors average 65% smaller.

For mean square error, assume that the direct estimates are unbiased and parallel the development of Ghosh and Rao (1994):

$$MSE(DSE_{Rake}) = VAR(DSE_{Rake}) + Bias^2$$

$$= VAR(DSE_{Rake}) + E[DSE_{Rake} - DSE_{Direct}]^2$$

$$= VAR(DSE_{Rake}) + E[(DSE_{Rake} - DSE_{Direct})^2] - VAR(DSE_{Rake} - DSE_{Direct})$$

we can approximate MSE by:

$$\hat{MSE}(D\hat{SE}_{Rake}) \approx \hat{VAR}(D\hat{SE}_{Rake}) + (D\hat{SE}_{Rake} - D\hat{SE}_{Direct})^2 - \hat{VAR}(D\hat{SE}_{Rake} - D\hat{SE}_{Direct})$$

DSE_{Rake} and DSE_{Direct} are highly correlated and the second and third terms are approximately equal and approximately 0. For all 24 marginal cells, $\hat{RMSE}(D\hat{SE}_{Rake})$ is within 1% of $\hat{SE}(D\hat{SE}_{Rake})$.

Similarly for the synthetic estimates:

$$\hat{MSE}(D\hat{SE}_{Synth}) \approx \hat{VAR}(D\hat{SE}_{Synth}) + (D\hat{SE}_{Synth} - D\hat{SE}_{Direct})^2 - \hat{VAR}(D\hat{SE}_{Synth} - D\hat{SE}_{Direct})$$

In this case the second and third terms are neither approximately equal nor approximately 0. With the available sample size, $\hat{VAR}(D\hat{SE}_{Synth} - D\hat{SE}_{Direct})$ is larger than $(D\hat{SE}_{Synth} - D\hat{SE}_{Direct})^2$ for 9 of the 24 marginal cells, so much so that $\hat{MSE}(D\hat{SE}_{Synth})$ is negative for all of them. These estimates are shown in Table 2.

The Pearson correlation coefficients between the raked undercount rates and the direct undercount rates for Poststratum Set J are greater than 0.999 for weighted (by size of the marginal subtotal) or unweighted data. The coefficients between the synthetic undercount rates estimated from Poststratum Set I and either the direct or raked undercount rates are approximately 0.74 for unweighted data and 0.82 for weighted data. Although all are within sampling error, the very close approximation to the direct estimates for the important cells on the margins of the raking matrix is a clear advantage of raking over synthetic estimation.

The synthetic assumption states that all persons within a poststratum have similar coverage properties. If the 29 poststrata of Poststratification Set I after collapsing satisfied the synthetic assumption, the synthetic estimates for substate areas would be equal to the direct estimates. This failure of the synthetic assumption is to be expected. Most states are too complex for 29 poststrata to be sufficient to satisfy wide ranging data needs reflecting differences in race, age, sex, tenure, and geography. The raking approach allows matching or near matching on two (or more) sets of about 30 poststrata instead of just one.

INTERIOR CELLS

With synthetic estimation, all sample persons in each of the 29 marginal cells, I, are assigned the same undercount rate and coverage factor. The same coverage factor is then applied to all persons in interior cell I,J, J=1,..,24. For non-Hispanic white and other owners less than 18 years old the synthetic undercount rate is about 3.00%. With raking the rates in the interior cells of the raking matrix corresponding to this group vary from a 0.80% overcount to a 8.99% undercount depending on the additional variables: site, tenure, and household stability. The range of the raked interior cell undercount rates is about 10% in each of the race/origin by age/sex marginal poststrata with the higher undercount rates for renters in "non-stable" households. Raking maintains the heterogeneity in the two sets of marginals defined by the two poststratification sets by maintaining and sometimes creating heterogeneity in the finest level of cells.

Examining the interior cells also shows the effect of including the household stability variable. Recall that all housing units with persons 18-29 are not "stable", so this high undercount population is always "non-unstable." Overall there is about a 4% difference between "stable" and "non-stable" residents in the undercount rates for the twelve site by tenure comparisons. About half of this difference is caused by the 18-29 year olds and about half is caused by 4% to 7% differences for the remaining population in just four of the site by tenure comparisons (LA county: owners and renters; small central cities: renters; and large central cities: renters).

Even though the cells in the interior of the raking matrix have little sample, it is possible to compare the raked estimates with the interior cell direct estimates. There are 595 cells in the two dimensional matrix with sample persons. Poststratification by these cells was simulated and

direct Dual System Estimates and CVs were calculated. 108 of these 595 cells have more than 100 (unweighted) persons in the PES sample. For these 108 cells the synthetic CVs average 2.0%, the raked CVs average 3.0%, and the direct CVs, based on 595 poststrata, average 5.3%. The direct CVs are much larger than the synthetic or raked CVs for smaller cells with the averages over all 595 cells equal to 3.0%, 4.0%, and 17.7%, respectively. With the small sample sizes and high standard errors, poststratification by all variables simultaneously to obtain estimates for the interior cells is not a viable option.

Unlike the situation for subtotals along the marginals of the raking matrix, raking does not improve the observed differences for the 595 nonnegative cells compared to synthetic estimation. For the 108 large cells, the observed differences between the synthetic estimates and the direct estimates average 5.7%; the observed differences between the raked estimates and the direct estimates average 4.4%. Over all 595 cells, these observed differences both average 9.4%. However, since the direct estimates of the interior cells have large standard errors, not matching the direct estimates for these small cells (for example, male API owners 30-49 in "stable" households in small central cities) cannot be considered a serious problem. The advantage of the raking approach is that the coverage factors generated for these smallest cells permit the estimates for the larger cells on the marginals of the raking matrix to closely approximate the corresponding direct estimates.

IV. SUMMARY

The Census Bureau developed and has sixty years of experience with iterative proportional fitting or raking. Raking appears to be an acceptable means of making the Dual System Estimates of the Integrated Coverage Measurement program in Census 2000 approximate a range of subtotals more extensive than could be achieved with simple poststratification and synthetic estimation. The raking process maintains or creates heterogeneity in the interior cells of the matrix to reflect coverage differences between the populations in several alternative sets of marginal poststrata, with lower variances than could be obtained by direct estimation, but with some bias.

In contrast, traditional synthetic estimation misses the direct estimation targets defined by alternative poststratification sets and has homogeneous coverage factors within poststrata. It does, however, have lower standard errors in interior cells than the raked estimates.

The 1990 PES sample size in California was approximately what is being planned for most states in 2000. If raking is used in 2000, stratification should take into account all poststratification variables whether primary, i.e., intended for the "official census estimates" for congressional apportionment and used in the last stage of the raking procedure, or secondary, i.e., used for substate estimates. In 1990 a cluster sample of blocks, classified by race, tenure, and geography within the nine Census Divisions, was selected. Only synthetic estimates were possible for states and substate areas. In 2000, blocks can be classified by race and tenure within well defined substate areas desired for publication. Adequate sample sizes will be required to control variances for each marginal cell for which the raked and direct estimates will automatically be close[5].

California, New York, Texas, and Florida will have much larger than average ICM sample sizes. It may be possible to explicitly define several areas as a first stage of stratification and poststratification and rake within them instead of across them, or to use more extensive poststratification in all dimensions of the raking matrix.

The inclusion of tract-level mail response rate as an additional stratification variable is being considered to reflect neighborhood coverage difficulties similar to the way the tenure and household stability variables reflect housing unit coverage difficulties. Some decisions such as the appropriate areas for publication will have to be made on a state by state basis depending on the characteristics of the specific populations in each state. Also, more collapsing than that done in this study will be required. For example each marginal poststratum should probably contain at least about 1% of its state's sample, or about 300 persons. In 2000, most states will not have sufficient numbers of Blacks or Hispanics or Asians or American Indians to maintain all sex and age categories. It will have to be determined whether collapsing age/sex within race or collapsing race within age/sex is more appropriate.

V: REFERENCES

GHOSH, M. and RAO, J.N.K. (1994). Small Area Estimation: An Appraisal. *Statistical Science,* Vol 9, No 1, pp.55-93.

PURCELL, N.J. and KISH, L. (1980). Postcensal Estimates for Local Areas (or Domains). *International Statistical Review*, 48, pp.3-18.

WORD, David L., 1997, "Who Responds/Who Doesn't? Analyzing Variation in Mail Response Rates during the 1990 Census", U.S. Census Bureau Population Division Working Paper Series #19, March 10, 1997.

[5] Considering the limited sample size, this could be an easier task than designing an optimum sample for a more extensive set of poststrata.

Graph 1: Undercount Rate for 24 Poststrata

Comparing Direct, Raked, and Synthetic Estimates

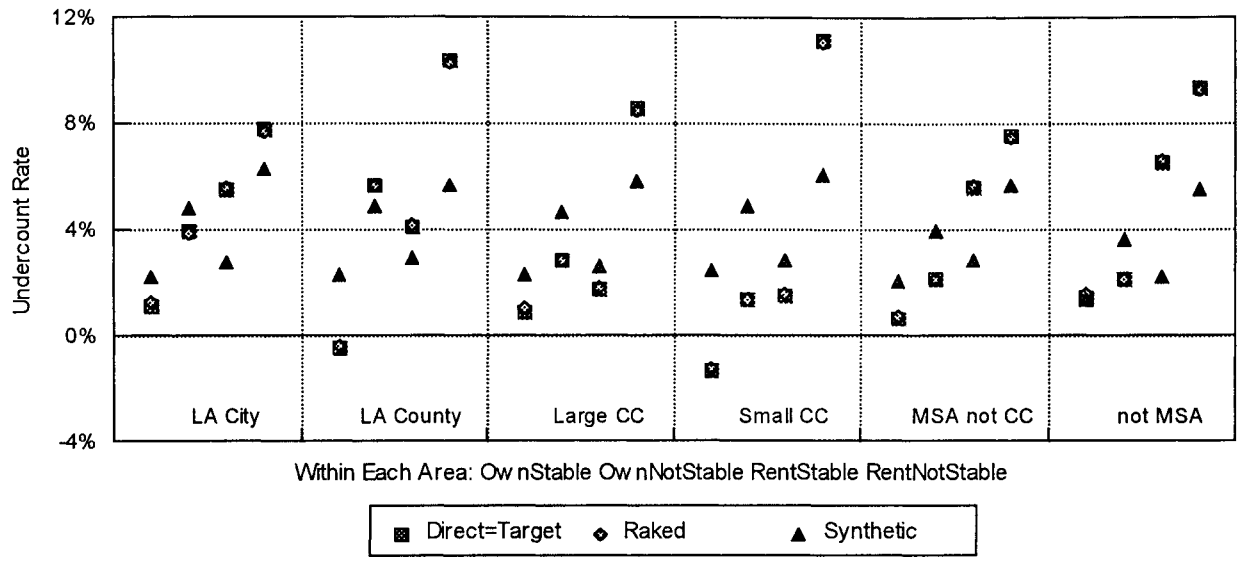

TABLE 2: Results for 24 Poststrata

Poststratum Definition			Samp Size	Weighted Sample	Direct Estimates = Target			Raked Estimates			Synthetic Estimates			
					DSE	UCrate	SE	DSE	UCrate	SE	DSE	UCrate	SE	MSE
Los Angeles City	Own	Stable	2028	706453	714561	1.13%	7797	715281	1.23%	7816	722145	2.17%	3292	/
		Non-Stable	3302	953401	992025	3.89%	13176	991611	3.85%	13143	1001466	4.80%	6347	/
	Rent	Stable	1156	514626	544810	5.54%	12139	545213	5.61%	12094	529024	2.72%	3123	11982
		Non-Stable	4529	1552357	1683913	7.81%	28250	1682148	7.72%	28071	1655897	6.25%	12871	13668
Los Angeles County	Own	Stable	581	901599	897000	-0.51%	13559	897803	-0.42%	13579	922600	2.28%	5409	21684
		Non-Stable	1048	1110849	1177695	5.68%	29859	1177143	5.63%	29819	1167427	4.85%	9184	/
	Rent	Stable	268	577634	602121	4.07%	11409	602537	4.13%	11465	594949	2.91%	5273	/
		Non-Stable	771	1262845	1408919	10.37%	37831	1407719	10.29%	37571	1338601	5.66%	11642	60631
Large Central Cities	Own	Stable	1097	1263342	1274818	0.90%	14518	1275915	0.99%	14249	1292721	2.27%	6926	8710
		Non-Stable	1664	1408073	1449158	2.84%	24274	1448583	2.80%	24037	1476669	4.65%	9504	16705
	Rent	Stable	481	511517	520479	1.72%	14243	520973	1.82%	14330	525286	2.62%	4226	/
		Non-Stable	1486	1026265	1122483	8.57%	22935	1121625	8.50%	23186	1089233	5.78%	6854	24404
Small Central Cities	Own	Stable	1267	690143	680934	-1.35%	14030	681561	-1.26%	13975	707576	2.46%	4425	22854
		Non-Stable	1496	741937	752048	1.34%	11060	751687	1.30%	11195	779798	4.86%	5612	24511
	Rent	Stable	360	235516	239046	1.48%	6310	239201	1.54%	6313	242379	2.83%	1848	/
		Non-Stable	1581	1110125	1250204	11.20%	63941	1249081	11.12%	63334	1182023	6.08%	9455	23405
MSA not Central City	Own	Stable	2424	5283540	5316871	0.63%	36636	5321869	0.72%	35944	5395245	2.07%	28679	73713
		Non-Stable	2360	6280831	6416522	2.11%	100895	6415128	2.09%	100870	6539719	3.96%	34187	69254
	Rent	Stable	567	1219961	1292148	5.59%	33079	1292925	5.64%	33089	1255917	2.86%	7887	10731
		Non-Stable	1398	3055531	3305884	7.57%	89912	3303061	7.49%	89246	3239664	5.68%	19555	/
Not MSA	Own	Stable	1933	573857	581923	1.39%	7358	582761	1.53%	7361	581377	1.29%	2532	/
		Non-Stable	1488	347793	355217	2.09%	6223	355186	2.08%	6220	360959	3.65%	2060	/
	Rent	Stable	695	143758	153789	6.52%	5110	153910	6.60%	5104	146976	2.19%	876	4623
		Non-Stable	1572	332784	367137	9.36%	15388	366782	9.27%	15219	352050	5.47%	3655	4112