BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-87/10

ON THE REDESIGN OF THE AMERICAN HOUSING SURVEY WEIGHTING PROCEDURE

by

Edward Gbur
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Larry Ernst

Report completed: August 1985

Report issued: April 1987

Table of Contents

1. Introduction

The American Housing Survey (AHS) is conducted biennially by the Census Bureau for the Department of Housing and Urban Development to obtain information about the size and composition of the residential housing inventory, changes in the inventory resulting from new construction and losses, indicators of housing and neighborhood quality, as well as information about selected characteristics of the occupants. The first survey in the series was taken in 1973 under the name Annual Housing Survey and was repeated annually until 1981 when it became a biennial survey.

Although the primary objective of AHS is to provide housing inventory information between decennial censuses, many AHS estimates are not directly comparable to Census figures. Discussions of the comparability issue can be found in Current Housing Reports (Series H-150) and in a memorandum entitled "Evaluation of the accuracy of housing data" (Charles Jones, SMD and Arthur Young, HOUS to William Butz, DIR, dated August 6, 1984). More detailed discussions of specific aspects of the problem can be found in memoranda for the record entitled "Estimation of the housing inventory" (Paul Harple, HOUS, dated September 9, 1982) and "Differences between 1980 Census counts and AHS estimates of housing units by race" (Barbara Williams, HOUS, dated October 17, 1984).

AHS is a multistage cluster sample survey conducted by personal interview over a three month period in the fall of each survey year. There are 461 PSUs, of which 156 are self-representing (SR). The remaining PSUs are selected from 220 strata and are nonself-representing (NSR). The details of the sample design can be found in Current Housing Reports (Series H-150). Once a housing unit is selected, it remains in subsequent surveys unless it is demolished, lost through disaster, converted to nonresidential use or group quarters, or moved from the site. Each year additional units representing new construction are added to the sample.

Currently AHS is undergoing extensive redesign in preparation for the eleventh survey to be conducted in the fall of 1985. Using the 1980 Census, the strata have been redefined and an entirely new sample has been drawn. The sampling design has been changed from two independent samples known as the A and C Samples (cf., Technical Paper 40) to a one PSU per stratum design. A new variance estimation procedure will be required. Personnel from the Longitudinal Surveys Branch in SMD are conducting research on the noninterview and first stage (sampling of NSR PSUs) adjustments in the weighting procedure. This work will probably lead to changes in these stages. This report describes the results of research dealing with various aspects of the weighting procedure beyond the noninterview and first stage adjustments.

In Section 2 of this report we briefly describe the current AHS weighting scheme. The data file, variables, and survey years used in our work are given in Section 3. Section 4 contains a discussion of the stability of the estimates of selected characteristics over time. Potential improvements in the current control factors are described in Section 5 and Section 6 contains the results of a study of alternative sets of cell configurations for the second and third stages of the current weighting scheme. The results of our research on the current raking procedure are presented in Section 7. Section 8 deals with new control factors which were investigated. Our conclusions and recommendations are set forth in Section 9 along with a brief discussion of future research topics.

2. An Overview of the Current AHS Weighting Procedure

A complete description of the 1983 AHS weighting procedure can be found in memoranda entitled "AHS-National: Specifications for weighting regular year X (1983) data" (Charles Jones, SMD to Thomas Walsh, DSD, dated November 19, 1983) and "1983 AHS-National: Independent control counts to be used in year X (1983) weighting" (Charles Jones, SMD to Thomas Walsh, DSD, dated March 13, 1984). Similar memoranda exist for previous AHS data. A brief outline of the scheme is included here for completeness.

A basic weight is assigned to each unit in the sample based on its probability of selection. This basic weight is multiplied by the following set of factors:

1. a duplication control factor,
2. a noninterview adjustment factor,
3. a first stage factor using adjustment cells defined by region of the country, residence, vacancy, and tenure status,
4. a second stage factor using adjustment cells defined by region of the country and "construction status",
5. a third stage factor using adjustment cells defined by residence, tenure, race and sex of the householder for occupied units and by vacancy status for vacant units,
6. a fourth stage factor which repeats the second stage cells,

7. a fifth stage factor which repeats the third stage cells. Although the same factors have been used for all ten previous AHS, there have been some changes in adjustment cells and sources of control counts.

The second and fourth stage sample cell totals include Type B and C noninterviews and certain ineligible vacant losses; the remaining stages do not include these records. The second and fourth stages use Survey of Construction (SOC) based control counts and the third and fifth stages use Current Population Survey (CPS) and Housing Vacancy Survey (HVS) based control counts. The last four adjustment stages can be viewed as a raking procedure having two cycles with two steps per cycle. More detailed discussions of the cell definitions, control counts, and raking procedure are given in subsequent sections in which related research is described.

3. Data Used in the Research

3.1 Design, Variables, and Data

The data used in this research was obtained from the AHS National Longitudinal File - Abbreviated. Each record of this file contains all information on one unit: the unit's identifiers and, for each of the ten AHS survey years, the unit's final weight and the values of approximately twenty five selected variables. For those surveys in which there is no record for the unit, the appropriate section of the record is blank. The file contains 103,643 records.

Discussions with SMD personnel led to the selection of the following variables for use in the evaluation of any proposed changes in the estimation procedure and in the study of the stability of estimates over time. We shall refer to these variables as evaluation variables. The variables are:

1. the number of cooperatives and condominiums,
2. the number of occupied one unit structures,
3. the number of occupied mobile homes,
4. the number of units lacking complete plumbing,
5. the number of vacant units for rent,
6. the number of households with a black head,
7. the number of households with a female head,
8. the number of low income households,
9. the number of owner occupied units with a low value,
10. the number of renter occupied units with low rent,
11. the number of renter occupied units in which the rent is a large proportion of the income,
12. the number of urban year round units,
13. the number of units with persons 65 or older who are living alone,
14. the number of units built before 1949.

The definitions of terms such as one unit structure, cooperative, condominium, and lacking complete plumbing can be found in Current Housing Reports (Series H-150). Each of the above fourteen variables either appears on the longitudinal file or can be directly calculated from variables on the file. The cutoff values for variables 8, 9, 10, and 11 are defined in later sections of this report.

For the research on the weighting adjustment factors and the raking procedure described in Sections 6 and 7, data from the 1980 survey was used for the work involving estimated levels. There was a sample reduction in rural areas in the 1981 survey so that it was not selected. Since the AHS is now a biennial survey, estimates of change were calculated over a two year time span. As a result, the 1983 survey was not chosen to estimate levels since its selection would require the use of the 1981 survey for estimates of change. The 1978 and 1980 surveys were used for the work involving estimates of change. For the research on the stability of selected estimates over time, all ten AHS were utilized.

As indicated in the Introduction, the new AHS design will be a one PSU per stratum design. To make our work compatible with the new sample design, the 85 C Sample PSUs which were not also in the A Sample, were eliminated from the data file. In addition, half of the records were deleted in each of the 25 A Sample PSUs which had been selected twice. This reduced the sample used in our work to 156 SR PSUs and 110 pairs of NSR PSUs. To maintain, at least approximately, the relative contributions of the SR and NSR PSUs to the sample estimates, the weight associated with each unit from an NSR PSU was multiplied by 1.5. Although estimates obtained from this reduced data file are not directly comparable to published AHS estimates, the benefits which accrue from calculating estimates and estimated variances based on the new sample design outweigh this disadvantage.

3.2 Variance Estimation

Three methods for calculating variance estimates were considered; the balanced half sample method with and without reweighting each half sample replicate and the Taylor series method. The balanced half sample method with reweighting is too costly in terms of computer resources and was discarded.

The balanced half sample method without replacement was initially considered for the research involving the comparison of alternative second and third stage cell configurations in the weighting procedure. However, since this research essentially involves comparing the effectiveness of different stratifications in reducing the variances of the estimators, the balanced half sample without reweighting can not be used. This is illustrated by the following example constructed by Larry Ernst (SRD). Consider the number of units in a given post-stratum (cell) as an evaluation variable. Suppose we have simple random sampling and only one post-stratification factor. Since all units in the given stratum have the characteristic and all units in the remaining strata do not, the variance of the estimator should be zero. If the variance is estimated using the balanced half sample without reweighting, a positive, sample dependent value is obtained for the variance estimate. Thus, the variance estimate contains extra variation which clouds any comparison of different post-stratification cell configurations. Although we are primarily interested in the relative merits of various cell configurations, it still would be necessary to assume that the additional variation did not alter the order of the relative variances of the evaluation variables. Since there is no way to check this assumption, the method was not used.

The Taylor series method was used to calculate variance estimates for the comparison of alternative second and third stage cell configurations. Details of the method are sketched in Appendix 1. For the research on the stability of selected estimates over time, the shortcomings of the balanced half sample without reweighting are not as critical and programming convenience led to its use. Since there are 266 strata, 272 replicates were constructed so that full orthogonal balance could be attained.

4. Stability of Selected Estimates Over Time

In any longitudinal survey it is desirable to have estimators which produce "stable" estimates over time. Estimates which fluctuate wildly from one time point to the next may be accurately reflecting the behavior of the characteristic of interest and in fact, may be stable. Therefore, stability must be defined in terms of the deviation of the estimate from the parameter value and not in terms of the values of the estimator itself. Deviations can be measured by the first or second order moments of the estimators. Use of the first order moment would lead to a definition based on the relative error or its expected value, the relative bias. Use of the second order moment would lead to a definition given in terms of the variance or mean square error of the estimator.

We have chosen the latter approach and will define stability in terms of the coefficient of variation (CV)

$$CV(\hat{\theta}) = 100 \cdot se(\hat{\theta}) \;/\; E(\hat{\theta}), \qquad (1)$$

which can be estimated from the data. If the estimated CVs are approximately constant or are decreasing over time, then we shall consider the series of estimates to be stable. A related but distinct problem is the magnitude of the CVs. If the CVs are large but relatively constant, then the estimates are stable but consideration should be given to reducing their variance.

Estimates and estimated variances were calculated for each of the fourteen evaluation variables listed in Section 3. The definitions of variables 8, 9, and 10 were forced to vary with time. These changes were made subjectively in an attempt to compensate for inflation. More sophisticated changes were not considered even though they may have provided a better evaluation.

Low income households were defined as those whose reported income was less than a cutoff value which is approximately the poverty level in the survey year for a family of 2.5 members. The family size was chosen after an examination of the median number of persons per unit (Table A-1, Current Housing Reports, Series H-150). Approximate poverty level figures were obtained from CPS Reports, Series P-60 (Table A-2). Differences due to housing unit based figures in AHS and family based figures in CPS were ignored. The upper limit for a low rent unit was defined to be approximately 50% of the median monthly rent reported in CHR, Series H-150. Units of low value were defined to be those whose value, as reported by the occupant, were less than a cutoff value of approximately 50% of the median value of a unit reported in CHR, Series H-150. Since the value of a unit is a categorical variable in the longitudinal data file, there was some unavoidable variation in these values. The cutoff values for these three variables are listed in Table 1. For variable 11, units using more that 35% of their available income for rent were considered as having a high rent to income ratio. Median values reported in CHR, Series H-150 ranged from 23% to 29% during this period.

Table 1. Cutoff Values Used in the Definitions of Variables 8, 9, and 10

Year	Low Income	Variable Low Value		Low Rent
1973	$3100	$12500	(5*)	$65
1974	3600	12500	(5)	70
1975	3900	14500	(5)	80
1976	4100	17500	(6)	85
1977	4400	17500	(6)	90
1978	4700	20000	(7)	100
1979	5200	24500	(9)	110
1980	6000	24500	(9)	120
1981	6600	27500	(10)	135
1983	7200	30000	(11)	160

*Category label from the longitudinal file is given in parantheses.

Estimated CVs are given in Table 2 and plots of the CVs against survey year are shown in Figure 1 for each variable. The dotted line connecting the 1981 and 1983 estimates emphasizes the change in the time period between consecutive surveys. The scale on the vertical axis is the same for each variable to facilitate comparisons.

As might be expected, as the number of units possessing a characteristic increases, the estimated CV decreases. The number of cooperatives and condominiums and the number of units having a high rent to income ratio are prime examples. The deletion of the rural supplement and a 6% across the board sample reduction in 1981 may be the cause of the increase in the CV for the number

Table 2. Estimated Coefficients of Variation for the Fourteen Evaluation Variables

Variable

Survey Year	Coops & condos	1-unit struct	Mobile homes	Lacking plumbing	Vacant rentals	Black head	Female head
1973	13.4**	0.7	4.8	5.3	3.1	2.3	1.1
1974	9.0	0.6	5.2	5.2	3.3	2.1	1.4
1975	5.9	0.6	5.4	4.8	3.7	2.2	1.1
1976	4.7	0.7	5.1	4.7	3.1	2.1	1.1
1977	4.8	0.7	5.5	5.2	2.8	2.1	1.1
1978	5.1	0.8	5.6	5.2	3.8	2.1	1.3
1979	5.7	0.8	4.7	5.3	4.1	2.1	1.3
1980	5.5	0.8	4.6	5.5	3.4	2.1	1.1
1981	4.9	0.9	5.7	6.4	4.4	2.6	1.3
1983	4.3	0.8	5.0	5.4	4.1	2.3	1.2

Variable

Survey	Low income*	Low value*	Low rent*	High rent -income ratio	Urban year round	Persons 65+ alone	Old units
1973	1.8	3.0	2.4	14.7	0.8	2.1	0.9
1974	1.6	3.0	2.9	14.3	0.8	2.1	0.9
1975	1.5	2.9	2.5	11.8	0.8	1.7	0.9
1976	1.5	2.6	2.6	11.6	0.8	2.0	0.9
1977	1.7	3.0	2.8	10.5	0.8	2.0	0.9
1978	1.6	3.0	2.6	11.2	0.9	2.1	0.9
1979	1.7	2.7	2.6	10.7	0.8	2.0	1.0
1980	1.4	2.8	2.6	9.0	0.8	2.0	0.9
1981	1.5	3.2	3.1	9.6	0.9	2.2	1.1
1983	1.8	3.0	2.8	7.4	0.9	2.1	1.0

* Variables whose definitions were changed (see text).
** Entries are percentages.

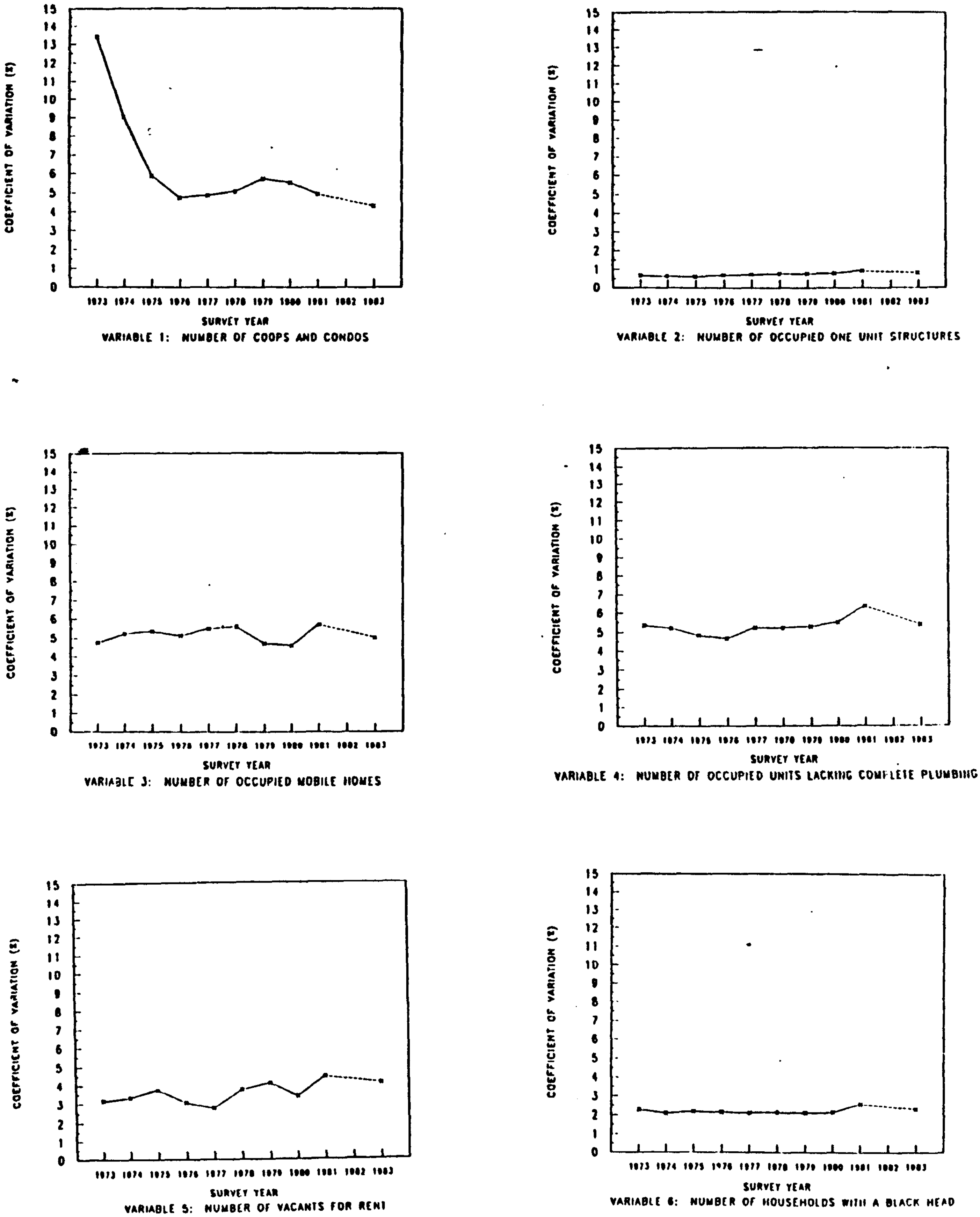
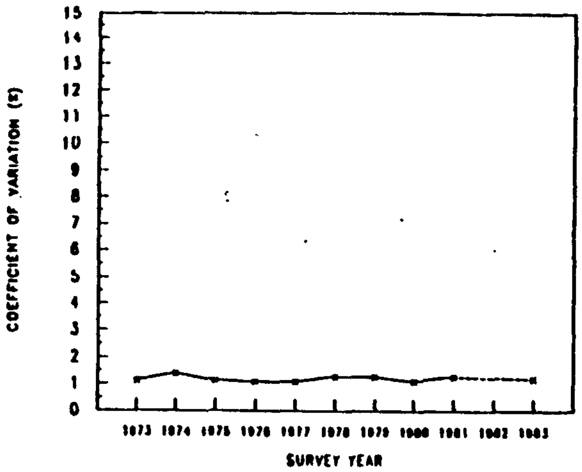
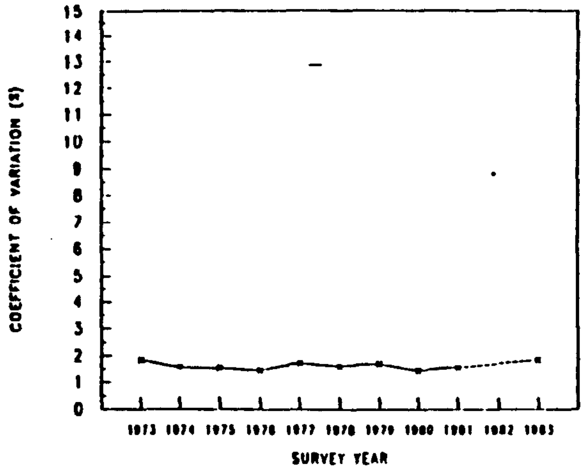


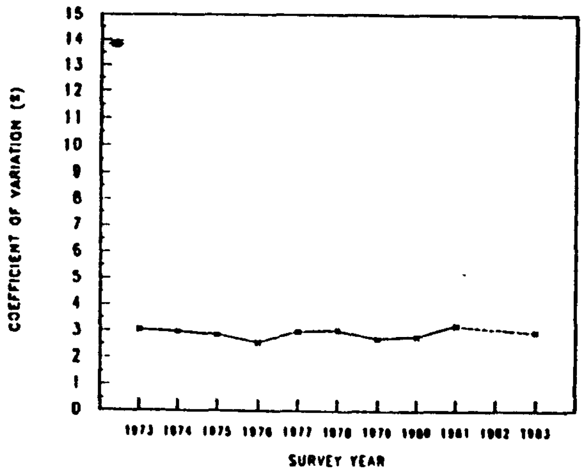
Figure 1. Plots of Coefficients of Variation Versus Survey Year for Each Evaluation Variable



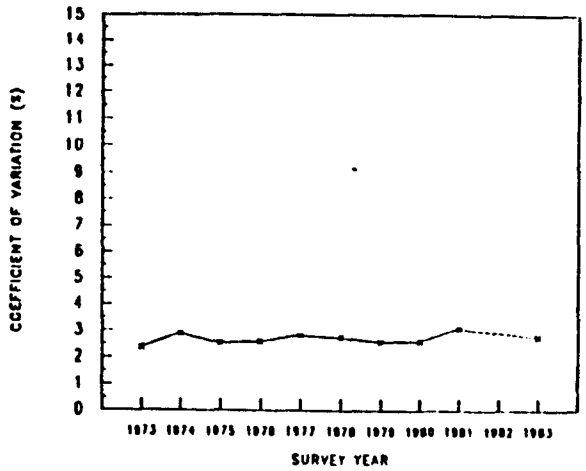
VARIABLE 7: NUMBER OF HOUSEHOLDS WITH A FEMALE HEAD



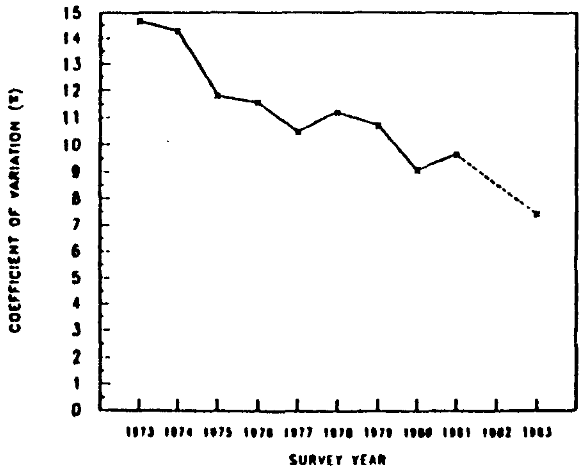
VARIABLE 8: NUMBER OF HOUSEHOLDS WITH LOW INCOME



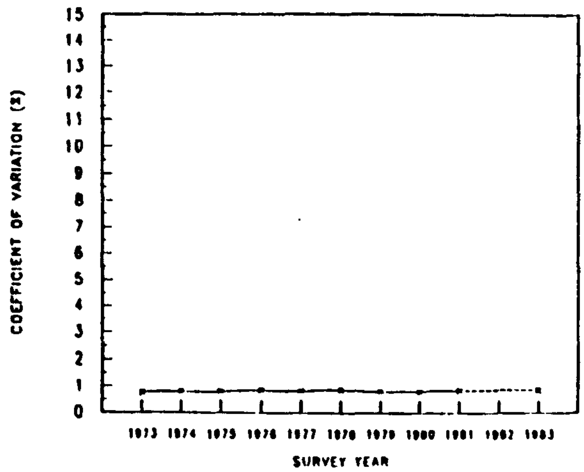
VARIABLE 9: NUMBER OF OWNER OCCUPIED UNITS WITH LOW VALUE



VARIABLE 10: NUMBER OF RENTER OCCUPIED UNITS WITH LOW RENT

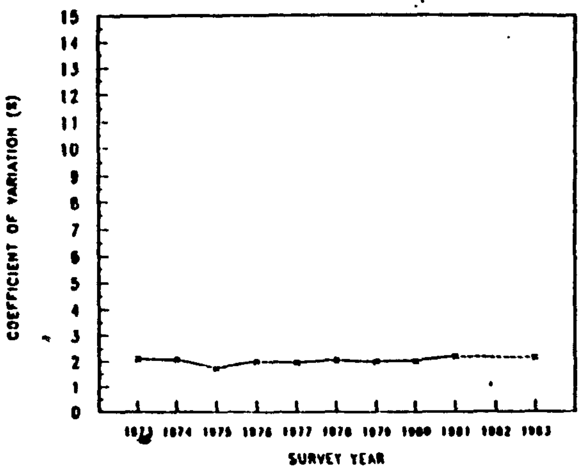


VARIABLE 11: NUMBER OF RENTERS WITH A HIGH RENT TO INCOME RATIO

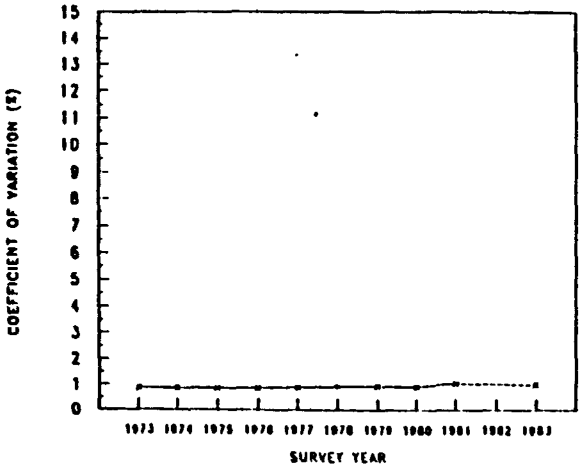


VARIABLE 12: NUMBER OF URBAN YEAR ROUND UNITS

Figure 1 (continued)



VARIABLE 13: NUMBER OF PERSONS 65 OR OLDER LIVING ALONE



VARIABLE 14: NUMBER OF UNITS BUILT BEFORE 1949

Figure 1 (continued)

of occupied mobile homes and the number of units lacking complete plumbing. A 7% across the board sample reduction in 1977 may explain the increase in CVs for some variables from 1976 to 1977.

The number of occupied mobile homes, vacants for rent, and occupied units lacking complete plumbing appear to be the least stable variables. These evaluations are subjective and the overall patterns in the plots may have a reasonable explanation.

The sum of the relative variances (square of the CV) for the fourteen evaluation variables provides a measure of the overall stability of AHS over time. These sums, along with the sum of the variances, are given in Table 3. The increasing trend in the sum of the variances is offset by the increasing level of the estimates, thereby producing a general downward trend in the sum of the relative variances. This downward trend indicates that, to the extent that the evaluation variables are representative of the entire survey, AHS has been relatively stable over time. The

Table 3. Sums of the Variances and Relative Variances for the Evaluation Variables over Time

Survey Year	Sum of Variances	Sum of Relative Variances
1973	571.5*	0.049**
1974	590.5	0.038
1975	582.7	0.027
1976	644.6	0.024
1977	686.7	0.023
1978	791.2	0.026
1979	774.9	0.024
1980	772.6	0.020
1981	1077.8	0.025
1983	1003.4	0.018

* Entries are in billions.
**Entries are unitless numbers.

large entry for 1981 may reflect the deletion of the rural supplement and the sample reduction.

5. Improvements in the Current Control Totals

The adjustment factors and independent control totals for the AHS weighting scheme were briefly described in Section 2. In this section, we elaborate on the current methods of obtaining independent control totals and suggest some improvements.

5.1 Third and Fifth Stage Control Counts

The third and fifth stage control totals are based on estimates obtained from CPS for occupied units and from HVS for vacant units. A description of the procedure is given in a paper entitled "Analysis of Census Bureau national housing inventory estimates" by David Bateman (1977 ASA Proceedings of the Social Statistics Section, 53-59) and is summarized below.

An estimate of the number of occupied units is obtained as follows:

1. CPS estimates of the number of occupied units are obtained for the 35 months preceding the execution of the AHS weighting procedure.
2. Twelve month moving averages are constructed from the estimates in Step 1.
3. A straight line is fit to the 24 moving averages.
4. The estimated total number of occupied units for the survey date is the appropriate predicted value from the fitted regression line.

A relative frequency distribution of occupied units for the third stage control cell configuration is obtained from CPS by averaging the sample distribution from the four quarters centered around the AHS survey date. The independent control counts are obtained by multiplying the estimated total from Step 4 by this average distribution.

The independent control counts for vacant units are obtained by averaging the vacancy distribution from HVS for the two quarters centered around the survey date and multiplying the result by the AHS sample total number of vacants.

The regression procedure described above is apparently motivated by a desire to smooth the random fluctuations in the original data. Unfortunately, this procedure has some undesirable properties. First, the averaging process used to obtain the dependent variable has created a strong correlation structure in addition to that already existing in the original data series. The large correlations between consecutive values must be accounted for in the modeling process. Second, the regression model is used to predict the twelve month average of the number of occupied units rather than the number of occupied units on the survey date. It is unclear why the average is the quantity to be predicted, especially since the controls for the vacant units are not in terms of twelve month averages.

The difficulties with the regression procedure are illustrated with data obtained from Jim Hartman (SMD). The data consist of monthly CPS estimates of the number of occupied units for the three year period from January, 1978 to December, 1980. A plot of the data is shown in Figure 2 and the twelve month moving averages are plotted in Figure 3. A simple linear regression model of the form

$$\bar{y}_t = \beta_0 + \beta_1 x_t + u_t, \qquad \text{for } t=1, \ldots, 24 \qquad (2)$$

where $\bar{y}_t$ = the average number of occupied units for the t-th set of twelve months,

x_t = the midpoint of the t-th set of twelve months,

was fit to the data using ordinary least squares. The fitted equation is given by

$$\hat{y}_t = 76.39 + 0.13\, x_t .$$

The regression was highly significant ($F = 19{,}302.0$) and $R^2 = 0.999$. However, the residual plot shown in Figure 4 contains a strong pattern indicative of a correlated error structure. The Durbin-Watson test for autocorrelation was highly significant. Inclusion of higher order polynomial terms in the model (2) will not correct this problem. The correlation structure must be incorporated into the model.

A simple regression model which attempts to account for the correlation structure is given by

$$\bar{y}_t = \beta_0 + \beta_1 x_t + u_t, \qquad (3)$$

$$u_t = \rho u_{t-1} + e_t , \qquad \text{for } t=1, \ldots, 24$$

where e_t are $iid(0,\sigma^2)$. The model (3) assumes that the errors u_t follow a first order autoregressive model. A fit of the model (3) to the data indicated that a more complex model for the error structure is required. Hence, the regression approach for the twelve month averages was abandoned.

More complex error structures than that given in (3) can be easily modeled using time series techniques. This approach was explored using the original data series rather than the twelve month averages for two reasons. First, the number of occupied

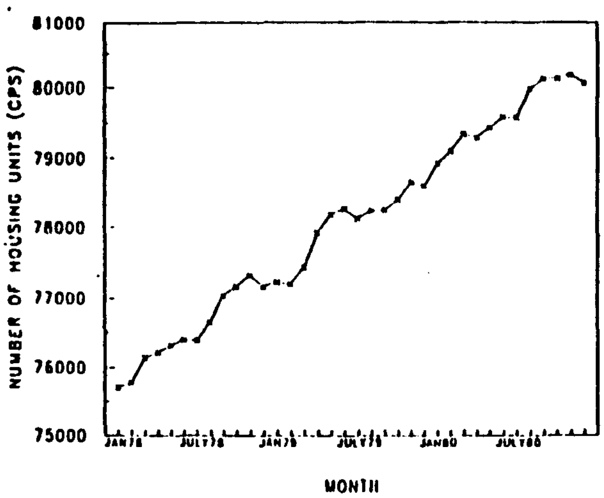


Figure 2. A Plot of the Estimated Number of Housing Units Versus Time

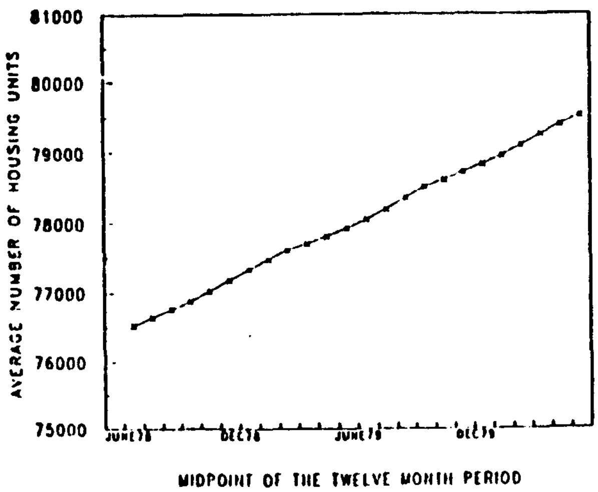


Figure 3. A Plot of the Twelve Month Moving Averages Versus Time

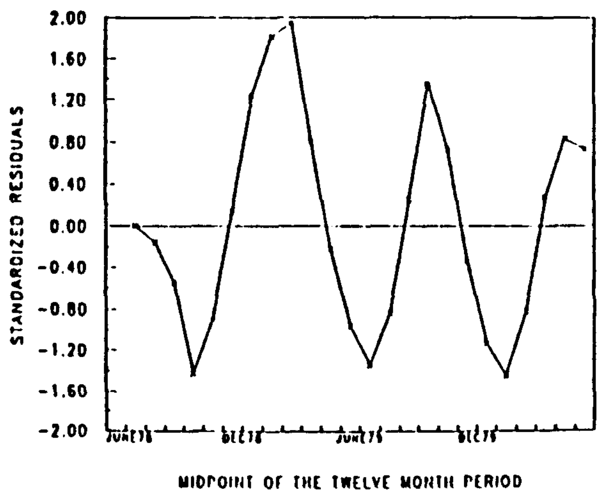


Figure 4. A Plot of the Standardized Residuals from the Fit of Model (3)

units on the survey date is a more reasonable variable to predict and second, we do not have to deal with the additional, artificial correlation structure imposed on the dependent variable by the averaging process.

The CPS data series was extended to cover the period from June, 1971 to October, 1980. With the assistance of William Bell (SRD), time series models were fit to the appropriate portions of the series to obtain the required estimates for the 1978 and 1980 AHS. A plot of the entire series is shown in Figure 5. The upward trend in the series indicated nonstationarity. Thus, the first differences (the series formed by calculating the differences between estimates for successive months) were modeled. A plot of the first differences is shown in Figure 6. Initial analyses indicated that the first differences could be modeled as a fourth order moving average with a 12 month seasonal component and a constant term; i.e.,

$$d_t = \gamma + (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)(1 - \psi B^{12})e_t , \qquad (4)$$

where y_t = CPS estimate of the number of occupied units in month t,

$d_t = y_t - y_{t-1}$,

B is the backshift operator $(Be_t = e_{t-1})$,

e_t are iid$(0,\sigma^2)$.

The estimated coefficients and their estimated standard errors, the mean square error, and the forecast and actual values for the 1978 and 1980 survey dates, along with the 95% confidence limits for the forecast, are given in Table 4. The autocorrelation

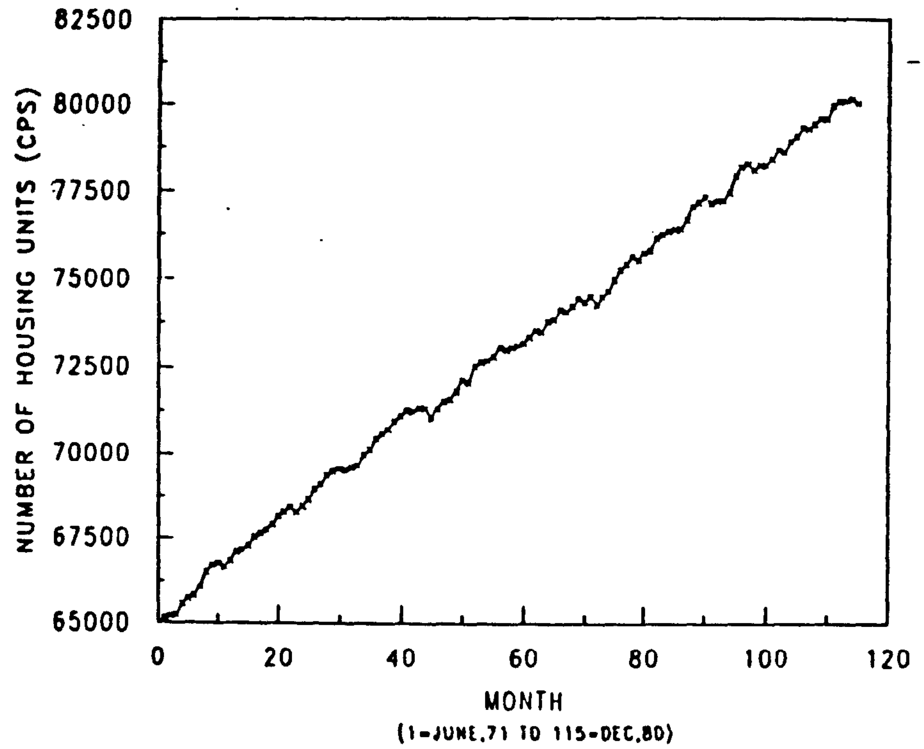


Figure 5. A Plot of the Estimated Number of Housing Units Versus Time

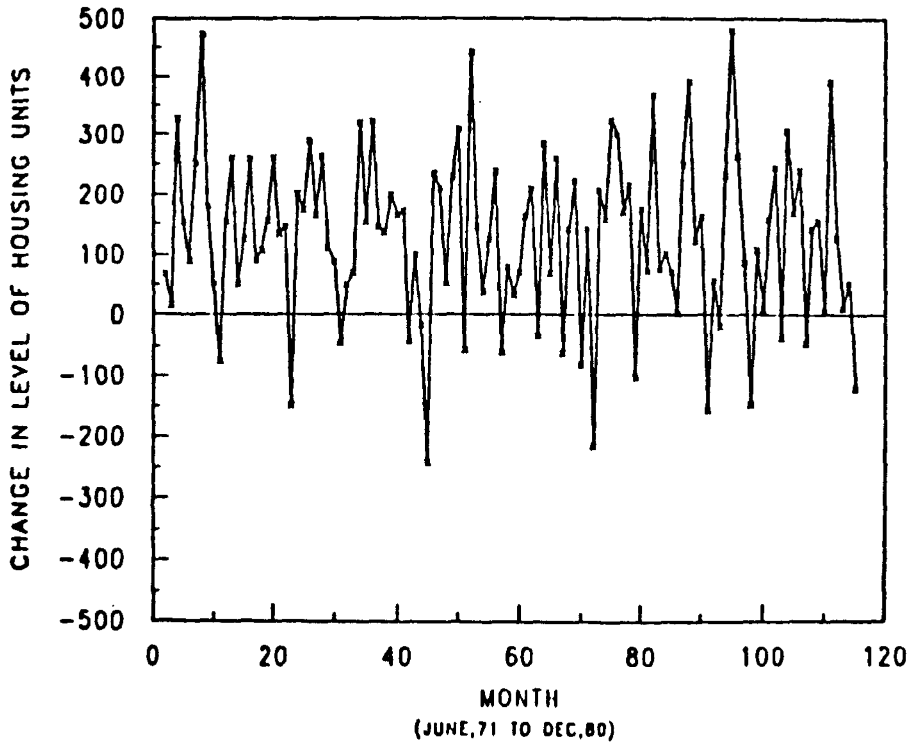


Figure 6. A Plot of the First Differences Versus Time

function of the residuals from both fits were indicative of white noise; i.e., iid$(0,\sigma^2)$ random variables.

Table 4. Results for the Time Series Model (4) Fit to the CPS Data

	1978 AHS	1980 AHS
n	88	112
γ	137.514(9.648)*	133.752(4.731)
ψ	-0.393(0.116)	-0.283(0.102)
θ_1	0.177(0.112)	0.175(0.095)
θ_2	-0.001(0.114)	0.049(0.094)
θ_3	0.174(0.113)	0.234(0.095)
θ_4	0.149(0.112)	0.243(0.096)
σ^2	16237.790	16694.705
CPS estimate	77156.0**	80120.0
prediction	77184.6	80233.2
lower 95% CL	76934.8	79979.9
upper 95% CL	77434.5	80486.5

* Standard errors are given in parantheses.
** Estimates are in thousands.

Both of the 95% confidence intervals for the predicted number of occupied units for the survey dates contain the observed CPS estimate for that month. Thus, the forecasts provide reasonable estimates of the true number of occupied units. Since the time series models take into account the correlation structure in the CPS data series, they are preferable to the simple regression based estimate.

It is worth noting that the time series model (4) is related to a straight line model. If differencing is thought of as taking derivatives (and recalling that the derivative of a straight line function is the slope of the line), then the model (4) can be reexpressed as

$$y_t = \gamma_0 + \gamma t + w_t \ ,$$
$$(1-B)w_t = (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)(1-\psi B^{12})e_t \ . \quad (5)$$

The model (5) can be viewed as a simple linear regression with a complex error structure.

For future surveys it is recommended that a time series model be fit to the CPS estimates and the predicted value for the survey date be used to obtain the independent control counts. A model of the form (5) would be a reasonable starting point but other possible models should not be excluded from consideration.

The CPS estimate of the total number of occupied housing units uses the principal person (or family) weights. There is some indication (cf., a memorandum entitled "Tabulations of monthly estimates of principal persons measuring the effect of CPS coverage improvement and second stage ratio adjustment" from David Bateman, SMD to Charles Jones, SMD, dated March 31, 1980) that the CPS estimates are biased upward. Since the difference between the CPS estimate and the predicted value from the time series model (4) is not significant for either year, the use of a time series model did not contribute further to any overestimates.

In a separate research project, alternatives to the current CPS principal person adjustment are being studied. At this time the research has not progressed far enough to speculate on the effect on AHS of any of the alternative methods. A report entitled "The Current Population Survey family weighting procedure and some alternatives" (by Edward Gbur, SRD, dated

December 1984) which describes the alternatives under consideration is available.

5.2 Second and Fourth Stage Control Counts

The second and fourth stage control factors are region of the country and "construction category". For the latter factor the housing inventory is divided into the set of units built before the most recent census (old construction) and the set of units built after the most recent census (new construction). New construction is subdivided into the time periods between consecutive AHS. Old construction units are classified as occupied or vacant while new construction units are classified by type (conventional one unit structure, conventional two or more unit structure, mobile home).

In 1981, the independent controls for conventional new construction were obtained from the Survey of Construction (SOC). In the remaining AHS, these cells had no independent controls; i.e., the second stage sample totals were used as controls for both the second and fourth stages. Old construction controls for occupied units are obtained by subtracting the new construction total from the CPS based estimate of the total number of occupied units. This procedure yields old construction controls and forces the grand totals from the second and third stage to be equal so that raking can be used (cf., Section 7).

Non-sample based mobile home controls have never been used. In telephone conversations with David Fondelier (CSD), it was determined that data on mobile home placements for residential use was not collected prior to 1974. Quarterly data are available from 1974 through 1979 (Construction Reports, Series C20) and monthly data has been collected since 1980. Since the quarterly data did not match the second and fourth stage control cells for the 1978 and 1980 AHS, we were unable to study the effect of using the CSD data for mobile home controls. However, the required information would be available for the 1985 AHS.

6. Changes in the Weighting Control Cells

In this section we compare several alternative sets of control cell configurations for the second and third stages of the weighting procedure. Only the factors currently in use in AHS are considered; a discussion of the research on potential new factors is delayed until Section 8.

For both the second and third stages the definitions of low income, low rent, and low value units used in this section differ from those given in Section 4. The cutoff values listed in Table 1 were changed and given a common value for the 1978 and 1980 AHS. In particular, low income households are defined as those reporting an income of less than $7000, low value units are those whose value, as reported by the occupants, was less than $25000, and rental units with low rent are those whose monthly rent is less than $200. These values were suggested by SMD personnel and are slightly higher than those used in Section 4. As a result, the number of sample units with these characteristics should be larger.

6.1 Second Stage Control Cells

The present second stage control cell configuration is defined by two factors: region of the country and construction type. These were described in Section 5.2 and will not be repeated here.

Four second stage configurations were considered. To obtain the four configurations two levels of each control factor were used; the original set of cells for the factor and a collapsed set of cells. Thus, the configurations form a 2×2 factorial design on the control factors. The levels of the region factors are (i) the original four regions (northeast, north central or midwest, south, and west) and (ii) no regional division of the country. The original set of cells for construction type is used as the first level of this factor. The second level is obtained by collapsing the time of construction portion of the factor, reducing it to three broad categories: old construction (prior to the last census), construction from the last census until the previous AHS, and construction since the previous AHS. As before, old construction is subdivided into occupied and vacant units while the remaining two categories are subdivided into conventional one unit structures, conventional multi-unit structures, and mobile homes. By considering broader time of construction categories we hope to reduce the misclassification rate of the sample units for the current cells, especially for mobile homes. The second stage configurations are listed in Table 5. Configuration A is the one currently in use.

Table 5. Second Stage Control Cell Configurations

Configuration	Factor: Construction Type	Factor: Region
A	original	original
B	collapsed	original
C	original	collapsed
D	collapsed	collapsed

Regardless of the level of the construction type factor there will be cells in the 1980 configuration which do not appear in the 1978 configuration. Thus, the variance of the estimator of change consists of two parts, the variation associated with the change for cells present in both years and the variation associated with the level for cells present in the 1980 configuration but not in the 1978 configuration. The proportion of the total variance in each part may vary with the configuration.

The variance calculations were divided into two parts corresponding to the contributions of the SR and NSR strata (cf., Appendix 1). In our research, the SR strata's contribution to the variance, given by equation (A5), is based on four replications. To obtain the replicates the six panels (first digit of the segment number) were arbitrarily paired into three sets: (1,4), (2,5), and (3,6). For a set of three "levels", a fully balanced design consists of four replicates. The design used in this study is given in Table 6. The weights needed to calculate the NSR strata's contribution to the variance (cf., equation A6) are based on the 1970 Census population estimates for the strata.

In every AHS except the 1981 survey, the second stage sample totals for all new construction cells were used as the independent controls (i.e., the ratio factors were set to 1.0). In 1981,

Table 6. Replicates for the SR Strata's Contribution to the Variance

Replicate	Sign Associated with the Panel Positive	Negative
1	4, 5, 6	1, 2, 3
2	2, 3, 4	1, 5, 6
3	1, 3, 5	2, 4, 6
4	1, 2, 6	3, 4, 5

Survey of Construction (SOC) based controls were used in AHS. For the study of alternative second stage control cell configurations, we used the applicable new construction independent controls from SOC which were used in the 1981 AHS. They are given in a memorandum entitled "AHS-National: Revision of the specifications for weighting the regular year IX (1981) data" (Charles Jones, SMD to Thomas Walsh, DSD, dated July 22, 1982). The old construction controls were obtained as described in Section 5.2 using CPS estimates of total occupied housing units from Table 4. Mobil home controls were taken to be the sample totals before deletion of the C Sample PSUs.

Although independent controls are never completely free of error, usually they are assumed to contain less error than the survey in which they are used. The SOC is subject to many of the same problems as AHS and is probably not significantly less error prone than AHS. Hence, it might be argued that SOC should not be used as a basis for controls. Since there is presently no other reliable source of construction controls, the alternatives to SOC based controls are the use of the AHS sample totals as controls or the elimination of the second stage adjustment altogether. The latter alternative is investigated in Section 6.3.

If the sample totals are used as independent controls for the new construction cells, then the difference between the variances for configuration A (the current procedure) and one of the remaining configurations depends only on old construction. To see this, note that configurations B, C, and D are obtained from configuration A by collapsing sets of control cells. For new construction the estimates will be the same for all configurations since the ratio factors applied to the first stage weights are 1.0 in all cases. For old construction the ratio factors will vary with the configuration, thereby producing second stage weights which depend on the configuration.

Seven of the design SR PSUs are located in more than one region of the country. However, each of these is divided into administrative PSUs which do not cross regional boundaries. We treated the parts of the design PSU in each region as separate PSUs, using the administrative PSUs in that region to form the parts. For example, for the Wilmington SMSA, the New Jersey portion (PSU 108) was treated as one PSU and the Maryland and Delaware portions (PSUs 513 and 553) were treated as a second PSU. The division of these PSUs led to more than 156 SR PSUs in the variance calculations.

For each configuration an estimate of the coefficient of variation for each of the fourteen evaluation variables listed in Section 3.1 was calculated for the entire sample. The CVs for the estimators of the 1980 levels are given in Table 7 and the CVs for the estimators of change are given in Table 8. The CVs were also computed for the SR and NSR strata separately but are not presented in this report.

Table 7. Coefficients of Variation for the Estimators of the 1980 Levels for the Second Stage Cell Configurations

Variable	Configuration A	B	C	D
Coops & Condos	4.9*	5.2	11.3	17.3
1-unit structures	0.4	0.5	1.5	1.7
Mobile homes	1.9	2.7	7.1	10.3
Lacking plumbing	5.5	5.5	13.0	12.9
Vacant rentals	4.2	4.2	9.6	9.6
Black head	2.0	2.1	2.5	2.5
Female head	0.9	0.9	2.2	2.3
Low income	1.0	1.1	2.8	2.8
Low value	2.8	2.8	8.4	8.4
Low rent	1.5	1.6	4.0	4.0
High income-rent ratio	8.2	8.2	18.0	18.0
Urban year round	0.5	0.6	1.7	1.8
Persons 65+ alone	1.6	1.6	4.1	4.0
Old units	0.8	0.8	2.7	2.7

*Entries are percentages.

Table 8. Coefficients of Variation for the Estimators of Change for the Second Stage Cell Configurations

Variable	Configuration A	B	C	D
Coops & condos	26.6*	28.2	61.3	95.4
1-unit structures	7.1	8.3	21.4	22.4
Mobile homes	13.0	17.9	38.0	36.1
Lacking plumbing	37.5	37.2	112.1	112.5
Vacant rentals	17.3	17.0	48.2	47.3
Black head	17.2	17.7	23.8	24.0
Female head	6.2	6.4	15.7	16.1
Low income	15.0	14.6	37.0	36.0
Low value	4.3	4.2	9.5	9.7
Low rent	3.8	3.8	8.9	8.9
High income-rent ratio	31.7	32.2	57.7	57.7
Urban year round	10.1	11.5	37.5	37.6
Persons 65+ alone	37.5	40.2	91.9	95.3
Old units	12.7	12.7	49.8	49.8

*Entries are percentages.

From Tables 7 and 8 it is clear that the CVs for estimators of change are, in general, much larger than those for estimators of level. In addition, configurations C and D have larger CVs than configurations A and B. Therefore, the region of the country should not be eliminated as a second stage factor. A simple count of the smallest CV for each variable (using the unrounded figures rather than the table entries) shows that configuration A yielded the smallest CV for 20 variables, configuration B for 6 variables, and there were two ties. In most cases the differences between CVs for these configurations is small; probably within sampling error.

The results for CVs computed for NSR strata only are similar, although the NSR CVs are larger than their counterparts in Tables 7 and 8. For CVs computed for SR strata only , configurations C and D have the smallest CVs for several variables but configuration A still has the largest number of smallest CVs.

Table 9 contains the sums of the relative variances (square of the CV) of the evaluation variables for each configuration for SR and NSR strata separately and for the entire sample. If we assume that these sums represent the overall performance of the configurations, then the entries for configurations A and B are probably within sampling error for the NSR strata and the entire sample. For the SR strata, no configuration is clearly superior.

An interaction graph for the sum of the relative variances for the estimators of the 1980 levels from the entire sample is shown in Figure 7. The graph for the estimators of change is similar and is not included. Although we do not have sufficient information to determine if there is interaction between the

Table 9. Sums of the Relative Variances of the Evaluation Variables for the Second Stage Cell Configurations

	Configuration			
Grouping	A	B	C	D
SR strata				
Change	2.615*	4.039	1.962	2.754
Level	0.021	0.022	0.023	0.023
NSR strata				
Change	76.341**	99.671	1098.272	819.538
Level	0.111	0.143	2.095	4.190
Entire Sample				
Change	0.590	0.641	3.856	4.440
Level	0.016	0.017	0.090	0.112

* Entries are unitless numbers.

** The large values in this row are due almost entirely to the evaluation variable which estimates the change in the number of cooperatives and condominiums. This is as expected since there are very few such units in NSR strata.

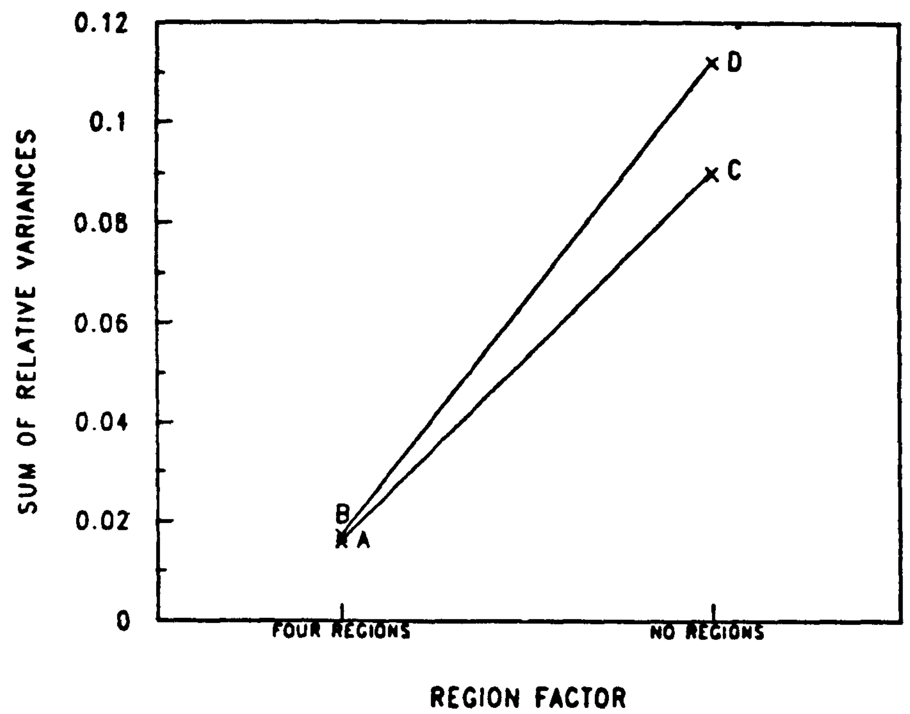


Figure 7. An interaction graph for the sums of the relative variances of the estimators of the 1980 levels in the second stage ratio adjustment

region and construction type factors, there is some indication of it in Figure 7. If the interaction is significant, then the effect of collapsing the construction categories is less when the full set of regions is used as the second factor than when there is no regional breakdown.

The results presented so far indicate that configuration A and B produce essentially the same results after the second stage adjustment. Since the difference between them is the coarseness of the time of construction categories, we investigated the problem of misclassification of units for time of construction by the respondents. This represents only one of many arguments which could be used to choose between configurations A and B.

It is generally acknowledged that respondents are prone to incorrect and inconsistent responses for the age of the unit. For example, recent movers, who comprise between ten and twenty percent of the sample, tend to report the unit as newer than it actually is. Renters also tend to have very poor information on the age of the unit. In the redesign two changes are being made to partially correct this problem. First, the time of construction will only be recorded for the unit's first time in sample. It will then be carried forward for succeeding surveys. Second, when a unit is introduced into the sample, a check will be made to prevent the year of construction from being more recent than the previous survey year unless it is known to be a newly constructed unit. The first change will provide consistency in the longitudinal file and should improve estimates of change. Neither change can prevent a unit from being misclassified.

Intuitively, configuration B should be less affected by misclassification of units than configuration A since it has a coarser time grouping. In an attempt to quantify any differences, units having information on year of construction in both the 1978 and 1980 AHS were cross-classified by construction type as reported in the two surveys. For configuration A, 653 units (1.39% of the units classified) had different time of construction categories in the two surveys while only 354 units (0.75%) had different categories for configuration B. The former rate is 1.8 times as large as the latter, although neither is very large. It does suggest, however, that configuration B is less affected by misclassification than configuration A.

For future surveys it is recommended that the second stage adjustment retain the region factor in its present form and retain the construction type factor, but with the time of construction portion collapsed into the three broad categories used in configuration B. The resulting decrease in the number of second stage control cells, along with the other changes for dealing with the reporting of the time of construction, should reduce the impact of the misclassification of units on survey estimates.

6.2 Third Stage Control Cells

The present third stage control cell configuration is divided into two sections; the first for occupied units and the second for vacant units. The two factors which define the control cells for occupied units are tenure, race, and sex of the householder, and residence status of the unit. Table 10 depicts the current configuration for occupied units. Vacant units are divided by type; year round units versus seasonal and migratory units. Year round units are subdivided into vacants for rent, vacants for sale, and other. Vacant cooperatives and condominiums are generally classified as other.

Table 10. Third Stage Control Cells for Occupied Units

	Residence Status		
	Inside SMSA		Outside SMSA
Tenure, Race, Sex	Central City	Balance	
Owner occupied			
White and other			
Male	×	×	×
Female	×	×	×
Black			
Male	×	×	×
Female	×	×	×
Renter occupied			
White and other			
Male	×	×	×
Female	×	×	×
Black			
Male	×	×	×
Female	×	×	×

The study of alternative third stage cell configurations is divided into two phases. In the first phase, alternatives for occupied units are considered while the control cells for vacant units are unaltered. In the second phase of the study, the optimal configuration for occupied units determined in phase one is fixed and alternative configurations for vacant units are considered. Since occupied and vacant units are adjusted separately using independent control totals from different, but related surveys, a two phase optimization is reasonable.

In the first phase, eight third stage cell configurations were analyzed in the form of a 2×2×2 factorial design. As in the study of the second stage adjustment, the set of original cells and a collapsed set of cells constitute the levels of each design factor. For the residence status factor, the two levels considered are (i) the original cells and (ii) inside an SMSA versus outside an SMSA. The remaining two design factors are sex and race of the householder; no changes in the tenure status categories were considered. Since both race and sex are dichotomous, the levels of each of these factors are (i) the original set of cells and (ii) no subdivision by that factor. Table 11 summarizes the description of the alternative configurations. Configuration A is the one currently in use for occupied units.

Table 11. Definitions of the Third Stage Cell Configurations for Occupied Units

Configuration	Residence	Factor Sex	Race
A	original	original	original
B	collapsed	original	original
C	original	deleted	original
D	collapsed	deleted	original
E	original	original	deleted
F	collapsed	original	deleted
G	original	deleted	deleted
H	collapsed	deleted	deleted

A description of the method used to calculate variance estimates for each evaluation variable is given in Section 6.1. The weights used in these calculations are the second stage weights for the optimal configuration determined in the previous subsection; i.e., for second stage configuration B. The CPS based tenure, race, sex, and residence distribution was converted to a set of control totals using the predicted number of occupied units given in Table 4. The HVS based control totals for vacant units are the same as those used in the 1978 and 1980 AHS.

For each of the eight configurations, an estimate of the CV for each of the fourteen evaluation variables was calculated from the entire sample. The estimated CVs for the estimators of the 1980 levels are given in Table 12 and the estimated CVs for the estimators of change are given in Table 13. The CVs were also computed for the SR and NSR strata separately but are not included.

The results in Tables 12 and 13 are mind-boggling, at best Although each ratio adjustment tends to introduce additional variability in the weights, only those variables which are essentially uncorrelated with the control factors would be expected to have larger CVs. However,the third stage CVs are generally larger than their second stage counterparts. Exceptions are households with black or female heads, which is to be expected since race and sex are third stage control factors. For most variables the NSR strata contribute more variability than the SR strata. This is due, in part, to the reduction in the NSR sample when the data file was constructed for this study (cf., Section 3.1). Thus, for the full sample, the third stage CVs should be much smaller than those given in Tables 12 and 13. However, the results can be used for the purpose of comparing the third stage cell configurations if we assume that the sample reduction alters the magnitude but not the order of the CVs.

Proceeding under this assumption, configurations A and B generally have smaller CVs than the remaining configurations for estimators of level, although in many cases the differences may be within sampling error. However, for estimators of change no single configuration is clearly superior. Table 13 does indicate

Table 12. Coefficients of Variation for the Estimators of the 1980 Levels for the Third Stage Cell Configurations***

Variable	Configuration A	B	C	D	E	F	G	H
Coops & condos	14.2*	14.3	18.1	18.2	14.8	14.8	18.7	18.7
1-unit structures	0.9	0.9*	1.0	1.0	0.9	0.9	1.0	1.0
Mobile homes	9.3	9.2*	10.5	10.3	9.4	9.3	10.7	10.4
Lacking plumbing	10.2	10.1*	11.4	11.4	10.8	10.7	12.2	12.1
Vacant rentals**	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5
Black head	0.3	0.3*	0.4	0.4	2.3	2.3	2.6	2.6
Female head	0.1	0.1*	1.8	1.8	0.1	0.1	1.8	1.8
Low income	2.3*	2.3	2.5	2.4	2.4	2.4	2.5	2.5
Low value	8.2*	8.6	9.8	10.4	8.3	8.7	10.1	10.6
Low rent	3.3*	3.4	3.8	3.9	3.4	3.5	3.9	4.0
High income-rent ratio	18.9	18.8	18.7	18.8	18.8	18.9	18.7*	18.8
Urban year round	1.5*	1.5	1.7	1.7	1.5	1.5	1.7	1.7
Persons 65+ alone	3.9	3.6	4.1	3.8	3.9	3.6*	4.1	3.8
Old units	2.3*	2.3	2.7	2.7	2.3	2.4	2.7	2.7

* The configuration with the smallest CV for that evaluation variable (using unrounded figures).

** The CVs for vacant rentals are identical since all phase one configurations have the same control cells for vacant units.

*** Entries are percentages.

Table 13. Coefficients of Variation for the Estimators of Change for the Third Stage Cell Configurations***

Variable	Configuration A	B	C	D	E	F	G	H
Coops and condos	36.1	36.9	34.4	35.6	36.0	36.7	34.1*	35.3
1-unit structures	14.3	14.4	13.2	14.2	14.2	14.2	13.4	13.2*
Mobile homes	34.7	34.2	34.1	33.9*	34.9	34.3	34.3	34.0
Lacking plumbing	165.5	158.1	125.0	122.1	125.3	124.0	93.9	93.0*
Vacant rentals**	697.9	697.9	697.9	697.9	697.9	697.9	697.9	697.9
Black head	4.7	4.5	3.9	3.8*	31.0	35.5	24.9	26.9
Female head	3.2	3.1*	13.1	13.4	3.5	3.4	13.3	13.5
Low income	28.8*	29.1	32.4	33.4	29.4	29.9	33.2	34.5
Low value	11.9	11.7	12.5	12.3	11.9	11.6*	12.3	12.0
Low rent	8.0	7.9	6.7*	6.7	8.3	8.2	6.9	6.9
High income-rent ratio	70.6	69.8	66.6*	69.4	70.2	70.2	66.7	69.8
Urban year round	33.3	34.2	28.6	28.9	33.8	35.2	28.6	29.1
Persons 65+ alone	806.1	710.6	93.3	90.4*	407.3	402.8	93.5	91.8
Old units	39.2	38.8	36.4	35.7	39.2	38.4	36.2	35.5

* The configuration with the smallest CV for that evaluation variable (using unrounded figures).

** The CVs for vacant rentals are identical since all phase one configurations have the same control cells for vacant units.

*** Entries are percentages.

that reductions in the number of tenure, race, sex control cells lead to smaller CVs for many of the evaluation variables. These conclusions are reinforced by the results presented in Table 14.

An interaction graph is shown in Figure 8 for the sums of the relative variances of the estimates of change in levels. Since the plot for each residence category exhibits the same pattern, there is no three way interaction between sex, race, and residence. Although we do not have sufficient information to perform a significance test, there is an indication of a sex-race interaction; deletion of the race of the head of household causes a much larger reduction in the sum of the relative variances when the sex of the head is included than when it is deleted. To the

Table 14. Sums of the Relative Variances of the Evaluation Variables for the Third Stage Cell Configurations

Grouping	Configuration A	B	C	D
SR strata				
Change	4.389	4.405	4.953	4.930
Level	0.022	0.022	0.023	0.024
NSR strata				
Change	2316.071	2067.514	445.443	491.490
Level	3.320	3.327	5.054	5.106
All strata				
Change	117.564	102.833	52.193	52.118
Level	0.090	0.090	0.111	0.112

Grouping	Configuration E	F	G	H
SR strata				
Change	4.080	4.007	4.920	4.830
Level	0.022	0.023	0.024	0.025
NSR strata				
Change	1819.018	1821.749	432.341	492.264
Level	3.510	3.519	5.327	5.383
All strata				
Change	68.105	67.746	51.583	51.601
Level	0.093	0.094	0.116	0.117

* Entries are unitless numbers.

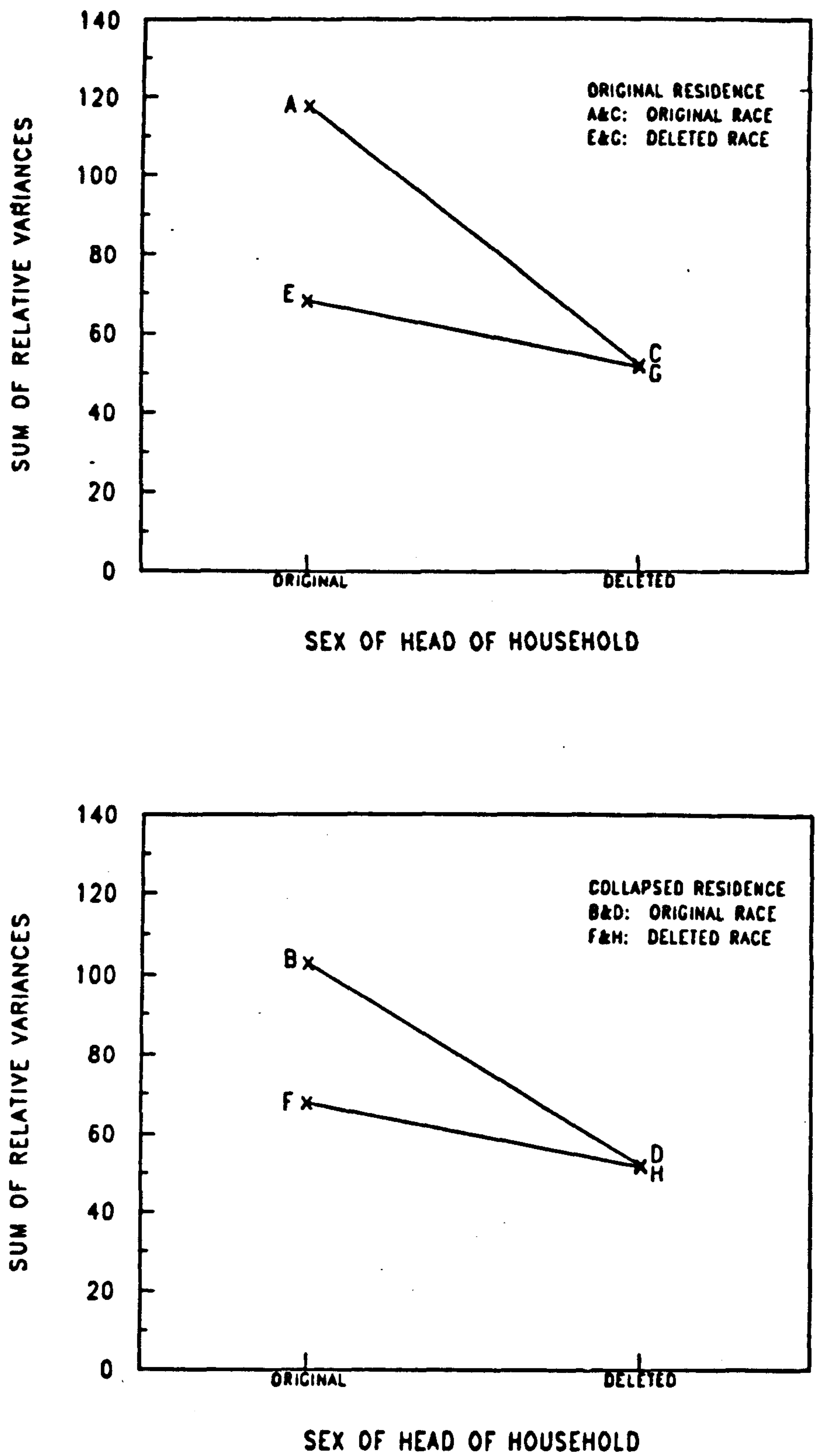


Figure 8. An interaction graph for the sum of the relative variances of the estimators of change in the third stage ratio adjustment

extent that the sums of the relative variances represent the overall performance of the configuration with respect to estimation of change, the sex of the head of household factor should be deleted from the third stage control cell definition; i.e., choose from among configurations C, D, G, and H. The additional reduction obtained by deleting the race of the head and collapsing the residence factor (i.e., choosing configuration H) is probably not statistically significant.

Since the emphasis in AHS publications appears to be on estimated levels rather than on change, it is recommended that the current tenure, race, sex control cells continue to be used. In the future if more emphasis is placed on estimating change, then serious consideration should be given to reducing the number of these cells by deleting the sex of the head of household factor. For a given set of tenure, race, sex control cells, the pair of configurations generated by the two levels of the residence status factor give essentially the same results. Thus, based only on the evidence presented here, no recommendation can be made for or against the use of the collapsed residence cells.

In the second phase of this study configuration A is used to define the control cells for occupied units. Four different configurations are considered for the vacant unit control cells. They form a 2×2 factorial design. For the vacancy type factor the two levels considered are (i) the original set of four cells and (ii) year round versus seasonal and migratory vacants. The second design factor was obtained by instituting a residence status for vacants in the form of inside an SMSA versus outside an SMSA. In the original third stage configuration vacants are not subdivided by residence status. Table 15 summarizes the configurations for the second phase.

Table 15. Definitions of the Third Stage Cell Configurations for Vacant Units

	Factor	
Configuration	Vacancy type	Residence
A	original	absent
I	original	present
J	collapsed	absent
K	collapsed	present

For configurations I and K, the bivariate distribution for vacant units is based on information obtained from Paul Harple (HOUS). The marginal distributions for vacancy type agree with those used in the 1978 and 1980 AHS. The residence factor was not expanded to the full set of three cells used for occupied units since the necessary information from the fourth quarter 1978 HVS was not readily available.

Ten of the fourteen evaluation variables listed in Section 3.1 deal only with occupied units and are not affected by changes in the control cells for vacant units. Thus, the comparison of the alternative configurations depends only on the following four variables:

1. the number of cooperatives and condominiums,
5. the number of vacant units for rent,
12. the number of urban year round units,
14. the number of units built before 1949.

In the course of this phase of the third stage research a problem was discovered with the seasonal and migratory vacants. A count of these units in our data file yielded a much smaller number than the corresponding count from the AHS weighting program output. Three factors contribute to the discrepancy; we were unable to determine the vacancy status of some units and did not include them in the analysis, some units were lost when the C Sample was removed from our data file, and a small number of units may have been misclassified under year round vacants as other vacants. If seasonal and migratory vacants are not evenly distributed over the NSR PSUs, then the adjustment to the weights for units in NSR PSUs to compensate for the deletion of the C Sample did not completely accomplish its purpose. In any case, the result of the loss of these units led to smaller sample totals and larger ratio adjustment factors than expected.

For each of the four vacancy configurations an estimate of the CV was calculated for each of the evaluation variables from the entire sample. The estimated CVs for the four variables which are affected by changes in the vacant unit control cells are presented in Table 16.

Table 16. Coefficients of Variation for the Evaluation Variables Affected by Changes in Vacant Unit Control Cells

	Configuration			
Variable	A	I	J	K
1980 levels				
Coops & condos	14.2*	14.2	14.2	14.2
Vacant rentals	6.5	5.7	29.8	21.1
Urban year round	1.5	1.3	1.7	1.4
Old units	2.3	2.5	2.3	2.6
Changes				
Coops & condos	36.1	36.6	36.5	36.9
Vacant rentals	697.9	584.0	62.0	75.5
Urban year round	33.3	35.4	28.6	33.0
Old units	39.2	42.5	39.0	38.2

*Entries are percentages.

From Table 16 the only substantial difference among the four configurations is in the CVs for vacant units for rent. Collapsing the vacancy type to year round versus seasonal and migratory vacants dramatically reduces the CVs for the estimator of change and increases them for the estimator of level. In comparison, the effect of a residence factor within each design level of vacancy type is negligible. Since the enormous reduction for the change estimator is probably more important than the increase for the estimator of level, configuration J is recommended for use in place of the current configuration A.

To summarize the results of the third stage research, it is recommended that future surveys retain the current set of control cells for occupied units and reduce the vacant unit control cells to a classification of year round versus seasonal and mirgatory vacants. The use of a residence factor for vacant units does not appear to be warranted.

6.3 Elimination of the Second Stage Adjustment

In Section 6.1 the problem of obtaining reliable construction control totals was briefly discussed. In light of these difficulties, the possibility of eliminating the second stage adjustment is investigated.

The two alternative weighting schemes to be compared are (i) the application of the second and third stage ratio adjustments to the first stage weights using the optimal configurations determined in the preceding subsections and (ii) the application of the "third stage" (CPS-HVS based) ratio adjustment to the first stage weights using the optimal third stage configuration. The results of the variance calculations for the first alternative have already been presented in Section 6.2. The second alternative requires the calculation of a new set of ratio factors for the CPS-HVS control cells to be applied to the first stage weights before variance computations can be performed.

The estimated CVs for each of the fourteen evaluation variables were calculated for the second alternative and are presented in Tables 17 and 18 for estimates of the 1980 levels and changes in the levels, respectively. For convenience, the CVs for the first alternative are also included in the table.

The results in Tables 17 and 18 indicate that, in general, the corresponding CVs for the two alternatives are approximately the same. The exceptions for estimators of change are the number of occupied mobile homes, the number of units with persons 65 or older who live alone, and the number of units built before 1949. For the estimators of 1980 levels, only the estimator of the number of vacant units for rent has a substantially larger CV when the second stage adjustment is omitted. Among these exceptions, only the larger CV for the number of units with persons 65 or older living along is surprising. The remaining variables are construction related and their CVs would be expected to increase.

The large increases in the CVs for the three change estimators listed above are all due to large differences in the estimates themselves; the difference in the corresponding estimated standard errors are relatively small. That is, elimination of the second stage adjustment led to significantly smaller estimates of change for these variables. In contrast, the increase in the magnitude of the CV for the number of vacant units for rent in the 1980 AHS is the result of an increase in the estimated standard error of the estimator.

Table 17. Coefficients of Variation for the Estimators of 1980 Levels for the Optimal Third Stage Configuration with and without the Second Stage Adjustment

Variable	Second Stage Included	Second Stage Excluded	Ratio of CVs*
Coops & condos	14.2**	14.8**	1.0
1-unit structure	0.9	0.9	1.0
Mobile homes	9.3	9.6	1.0
Lacking plumbing	10.2	9.7	1.0
Vacant rentals	29.8	40.5	1.4
Black head	0.3	0.2	0.7
Female head	0.1	0.1	0.9
Low income	2.3	2.3	1.0
Low value	8.2	7.7	0.9
Low rent	3.3	3.2	1.0
High income-rent ratio	18.9	19.1	1.0
Urban year round	1.7	2.1	1.2
Persons 65+ alone	3.9	3.8	1.0
Old units	2.3	2.6	1.1

* Ratio = (CV with second stage excluded)/(CV with second stage included).

** Entries are percentages.

Table 18. Coefficients of Variation for the Estimators of Change for the Optimal Third Stage Configuration with and without the Second Stage Adjustment

Variable	Second Stage Included	Second Stage Excluded	Ratio of CVs*
Coops & condos	36.5**	36.6**	1.0
1-unit structure	14.3	12.8	0.9
Mobile homes	34.7	111.9	3.2
Lacking plumbing	165.5	157.6	1.0
Vacant rentals	62.0	65.7	1.1
Black head	4.7	3.8	0.8
Female head	3.2	2.8	0.9
Low income	28.8	29.9	1.0
Low value	11.9	11.4	1.0
Low rent	8.0	7.9	1.0
High income-rent ratio	70.6	65.0	0.9
Urban year round	28.6	26.0	0.9
Persons 65+ alone	806.1	3335.8	4.1
Old units	39.0	69.9	1.8

* Ratio = (CV with second stage excluded)/(CV with second stage included).

** Entries are percentages.

Since very few CVs increase substantially when the second stage adjustment is omitted, it could be argued that it need not be included in the weighting scheme. However, care must be taken since its elimination precludes the use of raking on the second and third stage control cell configurations (cf., Section 7). It is possible that, after raking, the CVs listed in Tables 17 and 18 for the first alternative (using the second and third stage adjustments) may be reduced sufficiently to change the overall conclusion concerning the comparison of the two alternatives. If the reductions after raking are less dramatic, then elimination of the second stage adjustment may still be feasible but less attractive. In either case, it is recommended that more extensive research on this question be undertaken.

7. Raking in the AHS Estimation Procedure

Iterative proportional fitting or raking, as it is more commonly known, is a method of adjusting a table of non-negative real numbers so the resulting table entries sum to prespecified sets of marginal totals. The technique is used in sampling to produce survey estimates which are consistent with independent estimates of the same parameters. A byproduct of raking is a reduction in variance estimates.

For the simplest case of a two way table, raking consists of alternately scaling the rows and columns until the table entries sum to within a specified tolerance of the given marginal totals. If all table entries are positive, the procedure will converge

(R. Sinkhorn (1967). Diagonal equivalence to matrices with prescribed row and column sums. Amer. Math. Monthly 74, 402-405). For tables containing zero entries, raking will not always converge. There are known sets of conditions under which raking applied to a table with zero entries will converge (e.g., J. Fagan and B. Greenberg (1984). Making tables additive in the presence of zeros. Proc. of ASA Section on Survey Research Methods, 195-200).

In this section we show how the current AHS estimation procedure fits into the raking format and investigate potential changes and improvements in its application.

7.1 Raking in AHS

In current AHS estimation procedure the second and third stage post-stratification ratio adjustments are repeated as fourth and fifth stages, respectively. These four stages can be viewed as the first two cycles of a raking procedure. To see this, form a two way table whose columns are defined by the construction control cells (second stage) and whose rows are defined by the CPS-HVS control cells (third stage). Each sampled unit can be cross-classified into one cell of the table. The table to be raked has entries which are the sums of the first stage weights of all units classified in each cell. Thus, the AHS weighting scheme can be thought of as a series of ratio adjustments, the last of which is the third stage, followed by raking applied to the second and third stages.

The prespecified marginal totals for the raking procedure are the independent second and third stage control totals. Since the second stage occupied old construction controls are obtained by subtraction and the third stage vacant controls are obtained by multiplying the second stage sample total for vacants by the HVS vacancy distribution, the row and column independent controls will sum to the same grand totals, thereby achieving a basic consistency necessary to apply raking.

Each cycle in the raking procedure for this table consists of two steps; a column adjustment followed by a row adjustment. The adjustments made in the first raking cycle correspond to the second and third stage ratio adjustments. The second raking cycle yields the fourth and fifth stages of the estimation procedure (assuming SOC controls are used). Thus, the last four stages in the weighting procedure can be thought of as the first two cycles of raking for the given table.

The zero entries in the table to be raked can be either sampling or structural zeros. Examples of structural zeros occur for cells defined by old construction vacant units (second stage) and each tenure, race, sex, residence combination for occupied units (third stage). Such cells should have zero entries in the raked table. On the other hand, it is possible that a particular AHS sample may not contain any units in a given cell even though such units exist in the population. For example, sampling zeros may occur in cells defined by mobile homes in a region (second stage) and some tenure, race, sex, residence combinations (third stage). Such cells may have nonzero entries in the raked table.

Among the raking related questions which need to be investigated for AHS are the following.

1. Since the table will necessarily contain zero entries, will the raking procedure converge?
2. How many raking cycles should be used in AHS?
3. What is the effect of reversing the order of the steps within each raking cycle?

Each of these questions will be addressed in turn.

7.2 A Full Application of Raking

Although the current AHS weighting scheme contains the first two cycles of a raking procedure, it may be advantageous to incorporate additional cycles. An SMD memorandum entitled "CPS redesign: Second stage estimation - study plan for empirical research on cell definition, raking, and control of bias" (John Bushery and Robert Wilkinson to Lawrence Cahoon, dated February 15, 1984) indicates that after the first cycle in raking, additional cycles have little practical effect on the variance and bias of ratio estimates and are used primarily to force agreement between marginal sample and control totals. The redesigned CPS weighting procedure uses six cycles for this purpose. In light of the CPS experience, it may be necessary to include additional raking cycles in the AHS weighting scheme to achieve reasonable agreement of sample and control totals.

Additional raking cycles can be included in a weighting scheme with only a minimal increase in cost. Once the table to be raked is constructed, the intermediate steps in the raking procedure provide the necessary information to form the ratio adjustment factors for each additinal stage in the weighting scheme. The entire set of factors can be applied to the first stage weights in one pass through the data file.

In our study of raking in AHS, the two way table to be raked is formed using the optimal second and third stage control cell configurations determined in Sections 6.1 and 6.2, respectively. That is, the columns are defined using second stage configuration B and the rows are defined using third stage configuration J. Thus, the table has 26 rows and 32 columns. Only the 1980 AHS sample is analyzed.

Calculations were performed using a general purpose raking program written in Fortran by Jim Fagan (SRD). The program was modified to obtain additional intermediate output for analysis purposes. Documentation of the computational procedures used in the program are given in a (draft) report entitled "Algorithms for making tables additive: Raking, maximum likelihood, and minimum chi-square" (Jim Fagan and Brian Greenberg, dated June 16, 1985). The program treats all zeros as structural. However, a sampling zero can be replaced by a small positive value (e.g., a value less that the smallest first stage weight in the sample) if it is desirable to allow the corresponding entry in the raked table to be positive.

For our study the raking algorithm is said to have converged if the absolute value of the differences between the marginal sample and control totals are all less than a prespecified tolerance. Equivalently, we require the absolute difference between the sample and control totals for all cells in the second and third stage configurations to be less than the tolerance.

Defining convergence in terms of the ratio adjustment factors was originally considered and then rejected. That is, if the absolute value of the difference between each ratio factor and 1.0 is less than a prespecified tolerance, then convergence is said to be attained. However, such a convergence criterion can be misleading because the ratio depends on the magnitude of the terms which comprise it as well as on their difference. Table 19 illustrates this point. Since the primary purpose of raking in the weighting scheme is to minimize the magnitude of the discrepancies between the sample and control totals, the definition of convergence should not be given in terms of the behavior of the ratio factors.

Table 19. An Example of the Effect of the Magnitude of the Terms on the Ratio

Total Sample	Control	Difference	Ratio
150000	150100	100	1.0007
15000	15100	100	1.0067

Several tolerance levels were considered. The number of raking cycles required to achieve convergence for each tolerance level is given in Table 20. Although the exact number of cycles is dependent on the algorithm and the sample (which determines the table entries), valid comparisons can be made among the tolerance levels. There is usually very little difference in the number of cycles required for convergence by different raking algorithms.

As expected, the number of cycles required for convergence decreases as the tolerance increases. However, the large number of cycles needed for convergence may be surprising. This is discussed below in more detail.

Two approaches can be taken to the selection of the number of raking cycles to be incorporated into the AHS weighting

Table 20. The Number of Cycles Required for Convergence for Various Tolerance Levels

Tolerance	Number of Cycles
1.0	54
10.0	44
100.0	34
1000.0	25
10000.0	15
100000.0	5

scheme. First, a tolerance level can be chosen according to some criterion. Although this would ensure an upper bound on the differences between sample and control totals, the number of cycles would probably change slightly for each survey year. This may be difficult to explain to users but should not cause any serious programming problems. The second approach is to select the number of cycles to be included in the weighting scheme. The sample-control differences cannot be predetermined but a rough estimate of their magnitude could be obtained.

In either approach it must be realized that the sample-control differences are measured at the control cell level and not at the survey estimate level. For example, if a tolerance of 1000.0 is chosen and a particular survey estimate is obtained by summing ten cells, then the estimate could differ from the control by as much as 10000.0. Thus, care must be exercised in the use of the tolerance level to make decisions about the raking procedure.

The raking procedure was examined more thoroughly for a tolerance of 100.0. As a byproduct, the behavior for tolerances greater than this is readily available. A contour plot of the absolute differences between the control and sample totals is shown in Figure 9. The coding for the contours is given in Table 21. Cycle 0 represents the original disposition of the table to be raked. A positive contour value indicates that the sample total is less than the corresponding control total.

Table 21. Coding of the Coutour Values for Figure 9.

Range of the Absolute Difference*	Code (Unsigned)
0 to 100	0
100 to 200	1
200 to 500	2
500 to 1000	3
1000 to 10000	4
10000 to 100000	5
1000000 to 1000000	6
greater than 1000000	7

*Difference = (Control) - (Sample).

From Figure 9, the second stage cell sample totals converge to the corresponding controls faster than do the third stage cells. Among the second stage cells, old construction cells required the largest number of cycles to attain convergence while mobile homes constructed since the previous (1978) AHS needed the least. This is due, in part, to the relative magnitude of the totals in these cells (cf., the example presented in Table 19). However, old construction cells also had the largest initial differences between the sample and control totals. The results for the second stage also illustrate the fact that during raking a particular cell may have a small sample-control difference for an early iteration, only to have the absolute difference increase and then decrease before stabilizing.

For the third stage, the two vacant unit cells required the largest number of cycles to achieve convergence and, in fact, Construction Controls (Second Stage Cells)

NE	Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3
0	6	-6	5	4	5	5	5	-2
1	5	6	4	5	-4	4	4	0
2	5	4	4	5	0	4	4	0
3	5	-5	4	4	3	0	2	0
4	5	-5	4	4	3	-2	0	0
5	5	-5	4	3	3	-2	0	0
6	5	-5	4	2	3	-2	0	0
7	5	-5	3	2	3	-2	0	0
8	5	-5	3	1	2	-2	0	0
9	5	-5	2	1	2	-1	0	0
10	5	-5	2	0	2	-1	0	0
11	5	-5	2	0	2	-1	0	0
12	5	-5	1	0	1	-1	0	0
13	5	-5	1	0	1	0	0	0
14	4	-4	1	0	1	0	0	0
15	4	-4	0	0	0	0	0	0
16	4	-4	0	0	0	0	0	0
17	4	-4	0	0	0	0	0	0
18	4	-4	0	0	0	0	0	0
19	4	-4	0	0	0	0	0	0
20	4	-4	0	0	0	0	0	0
21	4	-4	0	0	0	0	0	0
22	4	-4	0	0	0	0	0	0
23	3	-3	0	0	0	0	0	0
24	3	-3	0	0	0	0	0	0
25	3	-3	0	0	0	0	0	0
26	2	-2	0	0	0	0	0	0
27	2	-2	0	0	0	0	0	0
28	2	-2	0	0	0	0	0	0
29	2	-2	0	0	0	0	0	0
30	1	-1	0	0	0	0	0	0
31	1	-1	0	0	0	0	0	0
32	1	-1	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0

NC	Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3
0	-6	-6	6	6	3	6	5	-4
1	-5	6	-5	5	-4	4	5	-2
2	5	5	4	5	-2	4	4	0
3	5	-5	4	4	4	0	0	0
4	5	-5	4	4	4	-3	-3	0
5	5	-5	4	2	4	-3	-4	0
6	5	-5	4	-1	4	-3	-3	0
7	5	-5	4	-2	4	-2	-3	0
8	5	-5	4	-2	3	-2	-3	0
9	5	-5	4	-2	3	-2	-3	0
10	5	-5	4	-2	3	-2	-2	0
11	5	-5	4	-1	2	-2	-2	0
12	5	-5	4	-1	2	-1	-2	0
13	5	-6	4	-1	2	-1	-2	0
14	4	-4	3	-1	2	-1	-1	0
15	4	-4	3	0	1	0	-1	0
16	4	-4	3	0	1	0	-1	0
17	4	-4	2	0	1	0	0	0
18	4	-4	2	0	0	0	0	0
19	4	-4	2	0	0	0	0	0
20	4	-4	2	0	0	0	0	0
21	4	-4	1	0	0	0	0	0
22	4	-4	1	0	0	0	0	0
23	3	-3	0	0	0	0	0	0
24	3	-3	0	0	0	0	0	0
25	3	-3	0	0	0	0	0	0
26	2	-2	0	0	0	0	0	0
27	2	-2	0	0	0	0	0	0
28	2	-2	0	0	0	0	0	0
29	2	-2	0	0	0	0	0	0
30	1	-1	0	0	0	0	0	0
31	1	-1	0	0	0	0	0	0
32	1	-1	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0

S	Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3
0	-7	-6	-5	6	6	6	5	-5
1	-6	5	-5	6	-5	5	5	-4
2	5	-6	4	5	-4	4	4	0
3	5	-6	5	4	4	-1	-3	1
4	5	-6	5	-4	4	-4	-4	1
5	5	-5	4	-4	4	-4	-4	1
6	5	-5	4	-4	4	-4	-4	1
7	5	-5	4	-4	4	-4	-4	1
8	5	-5	4	-4	4	-4	-4	0
9	5	-5	4	-4	4	-4	-4	0
10	5	-5	4	-4	4	-3	-4	0
11	5	-5	4	-4	4	-3	-3	0
12	5	-5	4	-4	4	-3	-3	0
13	5	-5	4	-4	3	-2	-3	0
14	4	-5	4	-4	3	-2	-2	0
15	4	-4	3	-3	3	-2	-2	0
16	4	-4	3	-3	2	-2	-2	0
17	4	-4	3	-3	2	-1	-2	0
18	4	-4	2	-2	2	-1	-1	0
19	4	-4	2	-2	2	-1	-1	0
20	4	-4	2	-2	1	0	-1	0
21	4	-4	2	-2	1	0	0	0
22	4	-4	1	-1	1	0	0	0
23	3	-4	1	-1	0	0	0	0
24	3	-4	1	0	0	0	0	0
25	3	-3	0	0	0	0	0	0
26	2	-3	0	0	0	0	0	0
27	2	-3	0	0	0	0	0	0
28	2	-2	0	0	0	0	0	0
29	2	-2	0	0	0	0	0	0
30	1	-2	0	0	0	0	0	0
31	1	-1	0	0	0	0	0	0
32	1	-1	0	0	0	0	0	0
33	0	-1	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0

W	Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3
0	7	-6	5	6	4	6	4	-4
1	6	-7	4	6	-4	5	5	-2
2	5	-6	5	5	-3	4	4	0
3	5	-6	4	4	4	-2	-2	0
4	5	-5	4	1	4	-4	-4	0
5	5	-5	4	-4	4	-4	-4	0
6	5	-5	4	-4	4	-4	-4	0
7	5	-5	4	-4	4	-4	-4	0
8	5	-5	4	-4	4	-4	-4	0
9	5	-5	4	-4	3	-3	-3	0
10	5	-5	4	-4	3	-3	-3	0
11	5	-5	4	-3	3	-3	-3	0
12	5	-5	3	-3	2	-2	-3	0
13	5	-4	3	-3	2	-2	-2	0
14	4	-4	3	-2	2	-2	-2	0
15	4	-4	2	-2	2	-2	-2	0
16	4	-4	2	-2	1	-1	-1	0
17	4	-4	2	-2	1	-1	-1	0
18	4	-4	2	-1	1	-1	-1	0
19	4	-4	1	-1	0	0	0	0
20	4	-4	1	-1	0	0	0	0
21	4	-4	1	0	0	0	0	0
22	4	-4	0	0	0	0	0	0
23	3	-3	0	0	0	0	0	0
24	3	-3	0	0	0	0	0	0
25	3	-3	0	0	0	0	0	0
26	2	-2	0	0	0	0	0	0
27	2	-2	0	0	0	0	0	0
28	2	-2	0	0	0	0	0	0
29	2	-2	0	0	0	0	0	0
30	1	-1	0	0	0	0	0	0
31	1	-1	0	0	0	0	0	0
32	1	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0

Figure 9. A contour plot of the absolute differences between the sample and control totals

CPS-HVS Controls (Third Stage Cells)

	Central City								Balance SMSA								Non-SMSA								Vacant	
	Owner				Renter				Owner				Renter				Owner				Renter					
Cycle	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
0	6	5	6	-5	-6	6	6	5	7	5	-5	-5	-5	-5	-4	5	6	-5	5	5	-6	6	5	5	-7	7
1	6	-5	6	-5	-6	-5	6	5	6	-5	-4	-5	-6	-6	-5	4	6	4	6	5	-6	5	5	5	-7	7
2	-4	-4	4	4	-5	-5	-5	-5	-5	-5	4	2	-6	-5	-4	-4	5	5	4	4	-5	-5	4	3	-6	6
3	-5	-5	-4	-4	-5	-5	-4	-5	-5	-5	-4	-3	-5	-5	-4	-4	-5	-5	-4	-3	-5	-5	-4	-4	5	6
4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	6
5	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	6
6	-5	-4	-4	-4	-5	-5	-4	-4	-5	-5	-4	-3	-5	-4	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	6
7	-5	-4	-4	-4	-5	-4	-4	-4	-5	-5	-4	-3	-5	-4	-3	-3	-5	-5	-4	-4	-5	-4	-4	-3	6	5
8	-5	-4	-4	-4	-4	-4	-4	-4	-5	-4	-4	-3	-4	-4	-3	-3	-5	-4	-4	-3	-4	-4	-3	-3	6	5
9	-5	-4	-4	-4	-4	-4	-4	-4	-5	-4	-4	-2	-4	-4	-3	-3	-5	-4	-4	-3	-4	-4	-3	-3	5	5
10	-5	-4	-4	-4	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-2	-2	-5	-4	-4	-3	-4	-4	-3	-3	5	5
11	-4	-4	-4	-3	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-3	-2	5	5
12	-4	-4	-4	-3	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-2	-2	5	5
13	-4	-4	-4	-3	-4	-4	-3	-4	-5	-4	-2	-1	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-2	-2	5	5
14	-4	-4	-3	-2	-4	-4	-3	-3	-4	-4	-2	-1	-4	-4	-1	-1	-4	-4	-2	-2	-4	-3	-2	-1	5	5
15	-4	-4	-3	-2	-4	-4	-3	-3	-4	-4	-2	-1	-4	-3	-1	-1	-4	-4	-2	-1	-4	-3	-1	-1	5	5
16	-4	-3	-2	-2	-4	-4	-2	-3	-4	-4	-2	0	-4	-3	-1	-1	-4	-4	-2	-1	-4	-3	-1	-1	5	5
17	-4	-3	-2	-2	-4	-3	-2	-2	-4	-3	-1	0	-4	-3	0	0	-4	-4	-2	-1	-4	-2	-1	0	5	4
18	-4	-3	-2	-1	-3	-3	-2	-2	-4	-3	-1	0	-3	-2	0	0	-4	-3	-1	0	-3	-2	0	0	4	4
19	-4	-2	-2	-1	-3	-3	-2	-2	-4	-3	-1	0	-3	-2	0	0	-4	-3	-1	0	-3	-2	0	0	4	4
20	-4	-2	-1	-1	-3	-2	-1	-2	-4	-2	0	0	-3	-2	0	0	-4	-3	-1	0	-2	-2	0	0	4	4
21	-3	-2	-1	0	-2	-2	-1	-1	-4	-2	0	0	-2	-2	0	0	-4	-2	0	0	-2	-1	0	0	4	4
22	-3	-2	-1	0	-2	-2	-1	-1	-4	-2	0	0	-2	-1	0	0	-4	-2	0	0	-2	-1	0	0	4	4
23	-3	-1	0	0	-2	-2	0	-1	-4	-2	0	0	-2	-1	0	0	-3	-2	0	0	-2	-1	0	0	4	4
24	-2	-1	0	0	-2	-1	0	0	-3	-1	0	0	-1	-1	0	0	-3	-1	0	0	-1	0	0	0	4	4
25	-2	-1	0	0	-1	-1	0	0	-3	-1	0	0	-1	0	0	0	-3	-1	0	0	-1	0	0	0	4	4
26	-2	0	0	0	-1	-1	0	0	-3	-1	0	0	-1	0	0	0	-2	-1	0	0	-1	0	0	0	4	4
27	-2	0	0	0	-1	C	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	4	3
28	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	3	3
29	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	3	3
30	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	3	2
31	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	2	2
32	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	2	2
33	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1

Figure 9. (continued)

controlled convergence for the entire table. The large number of cycles need by the white, male owner cell in each residence category is due to the magnitude of the totals and their initial difference. The sample totals for the vacant unit cells converged to their controls from below. This is undoubtedly related to the data file problem with these cells described in Section 6.2. This behavior of the vacant unit cells forces other third stage cells to compensate by converging to their controls from above. However, the fact that this occurred for all occupied cells is somewhat surprising. It is doubtful that these overall trends will exist in general for AHS data.

If Figure 9 indicates typical raking behavior for AHS, then at least ten cycles should be included in the weighting scheme. At that point for the 1980 AHS sample, six cells have converged and thirteen have sample-control differences in excess of 10000.0. If fifteen cycles are used, then thirteen cells have converged and only two have differences exceeding 10000.0. Since these two cells are the third stage vacant unit cells, a smaller number of cycles may suffice.

In summary, it is clear that more than two raking cycles need to be included in the AHS weighting scheme. If the fixed number of iterations approach is chosen, then a minimum of ten raking cycles is recommended.

7.3 Reversal of the Order of the Second and Third Stage Adjustments

In raking there is perfect agreement between the marginal sample and control totals for the margin which is adjusted last in each cycle. For AHS this means that the sample and control totals for the CPS-HVS control cells (third stage) are equal at the end of each raking cycle. At present the quality of the CPS-HVS controls is superior to that of the construction controls. Hence, it is reasonable that these controls should form the final adjustment in the weighting scheme. However, at some point in the future it may be more important to end the weighting scheme with a construction control adjustment.

Assuming that both construction and CPS-HVS controls are included in the weighting scheme, construction controls will form the final adjustment if the order of the present second and third stages are reversed. The effect of this reversal on the raking procedure is investigated in this subsection.

Since the raking program which was used first adjusts the columns of a two way table, the table to be raked is the transpose of the table described in Section 7.2. It contains 26 columns defined by the CPS-HVS controls and 32 rows defined by the construction controls. As before, a tolerance of 100.0 is used. A contour plot of the absolute differences is shown in Figure 10. The contours are defined as in the previous subsection (cf., Table 21).

From Figure 10 convergence was attained after 38 cycles, four more than for the original table in Section 7.2. Comparing Figures 9 and 10, the vacant unit cells for the CPS-HVS controls required several more cycles to achieve convergence. Changes for the remaining cells were negligible. After ten cycles, the same cells as before had absolute differences exceeding 10000.0. Thus, the reversal of the order of the present second and third

CPS-HVS Controls (New Second Stage Cells)

	Central City								Balance SMSA								Non-SMSA								Vacant	
	Owner				Renter				Owner				Renter				Owner				Renter					
Cycle	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
0	6	5	6	-5	-6	6	6	5	7	5	-5	-5	-5	-5	-4	5	6	-5	5	5	-6	6	5	5	-7	7
1	-6	5	5	5	-6	-6	5	5	-6	-5	5	4	-6	-6	-5	-5	6	6	5	5	-4	-5	5	5	-6	6
2	-5	-5	-4	-4	-5	-5	-5	-5	-6	-5	-2	-3	-5	-5	-4	-4	-4	-4	4	4	-5	-5	4	2	6	6
3	-5	-5	-4	-4	-5	-5	-4	-5	-6	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	6	6
4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	6	6
5	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	6
6	-5	-4	-4	-4	-5	-5	-4	-4	-5	-5	-4	-3	-5	-4	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	6
7	-5	-4	-4	-4	-5	-4	-4	-4	-5	-5	-4	-3	-5	-4	-4	-4	-5	-5	-4	-4	-5	-4	-4	-4	6	5
8	-5	-4	-4	-4	-5	-4	-4	-4	-5	-4	-4	-3	-4	-4	-3	-3	-5	-4	-4	-3	-4	-4	-4	-3	6	5
9	-5	-4	-4	-4	-4	-4	-4	-4	-5	-4	-4	-2	-4	-4	-3	-3	-5	-4	-4	-3	-4	-4	-3	-3	5	5
10	-5	-4	-4	-4	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-3	-3	-5	-4	-4	-3	-4	-4	-3	-3	5	5
11	-4	-4	-4	-3	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-3	-2	5	5
12	-4	-4	-4	-3	-4	-4	-4	-4	-5	-4	-3	-2	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-2	-2	5	5
13	-4	-4	-4	-3	-4	-4	-3	-4	-5	-4	-2	-1	-4	-4	-2	-2	-5	-4	-3	-2	-4	-4	-2	-2	5	5
14	-4	-4	-3	-2	-4	-4	-3	-4	-5	-4	-2	-1	-4	-4	-2	-1	-4	-4	-2	-2	-4	-3	-2	-2	5	5
15	-4	-4	-3	-2	-4	-4	-3	-3	-4	-4	-2	-1	-4	-4	-1	-1	-4	-4	-2	-1	-4	-3	-2	-1	5	5
16	-4	-3	-3	-2	-4	-4	-2	-3	-4	-4	-2	0	-4	-3	-1	-1	-4	-4	-2	-1	-4	-3	-1	-1	5	5
17	-4	-3	-2	-2	-4	-3	-2	-3	-4	-4	-1	0	-4	-3	-1	0	-4	-4	-2	-1	-4	-2	-1	-1	5	4
18	-4	-3	-2	-1	-3	-3	-2	-2	-4	-3	-1	0	-3	-2	0	0	-4	-3	-1	0	-3	-2	-1	0	5	4
19	-4	-2	-2	-1	-3	-3	-2	-2	-4	-3	-1	0	-3	-2	0	0	-4	-3	-1	0	-3	-2	0	0	4	4
20	-4	-2	-2	-1	-3	-2	-1	-2	-4	-3	0	0	-3	-2	0	0	-4	-3	-1	0	-3	-2	0	0	4	4
21	-3	-2	-1	0	-2	-2	-1	-1	-4	-2	0	0	-2	-2	0	0	-4	-2	0	0	-2	-1	0	0	4	4
22	-3	-2	-1	0	-2	-2	-1	-1	-4	-2	0	0	-2	-1	0	0	-4	-2	0	0	-2	-1	0	0	4	4
23	-3	-1	-1	0	-2	-2	0	-1	-4	-2	0	0	-2	-1	0	0	-4	-2	0	0	-2	-1	0	0	4	4
24	-2	-1	0	0	-2	-1	0	0	-3	-2	0	0	-2	-1	0	0	-3	-2	0	0	-2	0	0	0	4	4
25	-2	-1	0	0	-1	-1	0	0	-3	-1	0	0	-1	0	0	0	-3	-1	0	0	-1	0	0	0	4	4
26	-2	0	0	0	-1	-1	0	0	-3	-1	0	0	-1	0	0	0	-3	-1	0	0	-1	0	0	0	4	4
27	-2	0	0	0	-1	0	0	0	-2	-1	0	0	-1	0	0	0	-2	-1	0	0	-1	0	0	0	4	3
23	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	4	3
29	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	3	3
30	-1	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	0	0	0	3	2
31	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	2	2
32	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	2	2
33	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 10. A contour plot of the absolute differences between the sample and control totals when the order of the second and third stages are reversed

Construction Controls (New Third Stage Cells)

	NE								NC								S								W							
	Old		Recent			New			Old		Recent			New			Old		Recent			New			Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3
0	6	-6	5	4	5	5	5	-2	-6	-6	6	6	3	6	5	-4	-7	-6	-5	6	6	6	5	-5	7	-6	5	6	4	6	4	-4
1	5	6	5	5	4	5	5	-3	-7	6	5	6	-5	6	5	-4	-7	-5	-6	6	6	6	5	-5	7	-7	-4	6	-4	6	5	-4
2	6	-5	3	5	-4	3	4	0	5	4	-4	5	-4	4	4	-2	-6	-6	-5	5	-5	3	5	-3	6	-6	5	5	-4	4	4	-1
3	6	-5	4	4	3	-2	2	0	5	-5	4	5	4	-2	-2	0	5	-6	4	5	4	-4	-4	1	6	-6	5	5	4	-4	-3	0
4	5	-5	4	4	4	-2	0	0	5	-5	5	4	4	-3	-4	0	5	-6	5	-4	4	-4	-4	2	5	-6	4	4	4	-4	-4	0
5	5	-5	4	4	4	-2	0	0	5	-5	4	3	4	-3	-4	0	5	-6	5	-4	4	-4	-4	2	5	-5	4	-4	4	-4	-4	0
6	5	-5	4	3	3	-7	0	0	5	-5	4	0	4	-3	-4	0	5	-5	5	-4	4	-4	-4	1	5	-5	4	-4	4	-4	-4	0
7	5	-5	4	2	3	-2	0	0	5	-5	4	-2	4	-3	-4	0	5	-5	4	-4	4	-4	-4	1	5	-5	4	-4	4	-4	-4	0
8	5	-5	3	2	3	-2	0	0	5	-5	4	-2	4	-3	-3	0	5	-5	4	-4	4	-4	-4	1	5	-5	4	-4	4	-4	-4	0
9	5	-5	3	1	2	-2	0	0	5	-5	4	-2	4	-2	-3	0	5	-5	4	-4	4	-4	-4	0	5	-5	4	-4	4	-4	-4	0
10	5	-5	2	1	2	-2	0	0	5	-5	4	-2	3	-2	-3	0	5	-5	4	-4	4	-4	-4	0	5	-5	4	-4	3	-4	-4	0
11	5	-5	2	0	2	-1	0	0	5	-5	4	-2	3	-2	-2	0	5	-5	4	-4	4	-3	-4	0	5	-5	4	-4	3	-3	-3	0
12	5	-5	2	0	2	-1	0	0	5	-5	4	-1	3	-2	-2	0	5	-5	4	-4	4	-3	-4	0	5	-5	4	-3	3	-3	-3	0
13	5	-5	2	0	1	-1	0	0	5	-5	4	-1	2	-1	-2	0	5	-5	4	-4	4	-3	-3	0	5	-5	4	-3	2	-3	-3	0
14	5	-5	1	0	1	0	0	0	5	-4	4	-1	2	-1	-2	0	5	-5	4	-4	3	-2	-3	0	5	-4	3	-3	2	-2	-2	0
15	4	-4	1	0	1	0	0	0	4	-4	3	-1	2	-1	-1	0	4	-5	4	-4	3	-2	-3	0	4	-4	3	-2	2	-2	-2	0
16	4	-4	1	0	0	0	0	0	4	-4	3	0	2	0	-1	0	4	-4	3	-3	3	-2	-2	0	4	-4	2	-2	2	-2	-2	0
17	4	-4	0	0	0	0	0	0	4	-4	3	0	1	0	-1	0	4	-4	3	-3	2	-2	-2	0	4	-4	2	-2	1	-2	-2	0
18	4	-4	0	0	0	0	0	0	4	-4	2	0	1	0	0	0	4	-4	3	-3	2	-1	-2	0	4	-4	2	-2	1	-1	-1	0
19	4	-4	0	0	0	0	0	0	4	-4	2	0	0	0	0	0	4	-4	2	-2	2	-1	-2	0	4	-4	2	-1	1	-1	-1	0
20	4	-4	0	0	0	0	0	0	4	-4	2	0	0	0	0	0	4	-4	2	-2	2	-1	-1	0	4	-4	1	-1	0	-1	-1	0
21	4	-4	0	0	0	0	0	0	4	-4	2	0	0	0	0	0	4	-4	2	-2	1	0	-1	0	4	-4	1	-1	0	0	0	0
22	4	-4	0	0	0	0	0	0	4	-4	1	0	0	0	0	0	4	-4	2	-2	1	0	-1	0	4	-4	1	0	0	0	0	0
23	4	-4	0	0	0	0	0	0	4	-4	1	0	0	0	0	0	4	-4	1	-1	1	0	0	0	4	-4	0	0	0	0	0	0
24	4	-4	0	0	0	0	0	0	4	-3	1	0	0	0	0	0	4	-4	1	-1	0	0	0	0	4	-3	0	0	0	0	0	0
25	3	-3	0	0	0	0	0	0	3	-3	0	0	0	0	0	0	3	-4	1	-1	0	0	0	0	3	-3	0	0	0	0	0	0
26	3	-3	0	0	0	0	0	0	3	-3	0	0	0	0	0	0	3	-3	0	0	0	0	0	0	3	-3	0	0	0	0	0	0
27	3	-3	0	0	0	0	0	0	3	-2	0	0	0	0	0	0	3	-3	0	0	0	0	0	0	3	-2	0	0	0	0	0	0
28	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	2	-3	0	0	0	0	0	0	2	-2	0	0	0	0	0	0
29	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0
30	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	2	-2	0	0	0	0	0	0
31	1	-1	0	0	0	0	0	0	2	-1	0	0	0	0	0	0	2	-2	0	0	0	0	0	0	1	-1	0	0	0	0	0	0
32	1	-1	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	1	-2	0	0	0	0	0	0	1	-1	0	0	0	0	0	0
33	1	-1	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	1	-1	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 10 (continued)

stage adjustments has essentially no effect on the raking procedure. The recommendation to include ten or more raking cycles in the AHS weighting scheme is unchanged.

7.4 Variances and Raking

The evidence presented in Section 7.2 indicates that more than two raking cycles should be included in the AHS weighting scheme. This conclusion is based solely on the objective of forcing agreement between the sample and control totals for each adjustment cell. In this section we study the effect of additional raking cycles on the variances of the evaluation variables.

Variance estimates for each of the evaluation variables for the 1980 levels were computed after five and ten complete cycles of raking. The corresponding CVs were calculated and are presented in Table 22. For purposes of comparison the CVs after the third stage ratio adjustment (the end of the first raking cycle) are included.

From Table 22 only the CVs of the number of vacant units for rent and the number of units built before 1949 change significantly. The reduction in the former CV and the dramatic increase in the latter are due primarily to problems with the seasonal and migratory vacant units in the data file (cf., Section 6.2). The estimates of the number of cooperatives and condominiums and the number of urban year round units are affected to a much lesser extent. This is as expected given the nature of the problem units and the evaluation variables. For the remaining variables, the slight differences in the CVs across cycles are due to minor

Table 22. Coefficients of Variation for the Estimators of the 1980 Levels after Various Numbers of Cycles in the Raking

Variables	Number of Cycles 1*	5	10
Coops & condos	14.2**	15.0	15.0
1-unit structures	0.9	1.0	1.0
Mobile homes	9.3	10.0	10.0
Lacking plumbing	10.2	10.8	10.8
Vacant rentals	29.8	12.7	13.1
Black head	0.3	0.3	0.3
Female head	0.1	0.1	0.1
Low income	2.3	2.5	2.4
Low value	8.2	8.8	8.8
Low rent	3.3	3.4	3.4
High income-rent ratio	18.9	18.6	18.6
Urban year round	1.7	1.3	1.3
Persons 65+ alone	3.9	4.0	4.0
Old units	2.3	95.6	122.1

* The first cycle represents the second and third stage ratio adjustments.
** Entries are percentages.

variations in both the estimated levels and standard errors. That is, for twelve of the fourteen variables there is no significant change in either of the components of the CV.

The results presented in Table 22 indicate that additional raking cycles beyond the first iteration do not provide significant reductions in the variances of the evaluation variables. Similar results were obtained in research on the CPS redesign (cf., an internal SMD memorandum entitled "CPS redesign: Second stage estimation - Study plan for empirical research on cell definitions, raking, and control of bias" from John Bushery and Robert Wilkinson to Lawrence Cahoon, dated February 15, 1984). Thus, variance considerations should not play a role in determining the number of raking cycles to be included in the AHS weighting procedure.

8. New Control Factors

Two new factors for defining control cells were investigated. Both relate to the socioeconomic status of households. The first factor attempts to measure the quality of the housing stock while the second is based on the household's income relative to the cost of living for the area in which the unit is located. The housing quality factor may best be utilized as a replacement for the third stage residence factor. The income factor may be better suited for use as an additional control factor.

In this section the construction of each of the factors is discussed and its potential for use is assessed. No numerical work was carried out to evaluate these factors since many of the variables required for the research are not available on the AHS Longitudinal File.

8.1 Housing Quality Factor

The current third stage residence control factor divides the country into three broad categories based on SMSA status; i.e., units are grouped according to political and geographical boundaries. An alternative grouping of the housing inventory could be based on a measure of the "quality of the housing stock". For example, within an SMSA, this factor would group exclusive areas of the central city with similar type suburbs rather than with other central city areas which might have low income and poor housing quality. The housing quality factor would have a small number of broad categories and classification of sampled units would be a function of several variables, each attempting to measure a different aspect of the quality of the housing stock. Such a factor could prove to be important for estimating the amount of substandard housing in the inventory.

In the construction of a housing quality factor there are three important considerations.

1. How should the concept of "housing quality" be defined?
2. Should the factor be defined and applied at a single unit or a neighborhood level?
3. Should the factor be based on an evaluation of physical characteristics of the unit or neighborhood by the interviewer or on perhaps less tangible variables by the occupants?

Although data in both AHS and the Census are collected at the housing unit level rather than at some aggregate level, a housing quality factor envisioned as a replacement for the current residence factor is more naturally defined for some larger area which we shall refer to as a neighborhood. Classification of a sampled unit should be based on a combination of the interviewer's assessment of the physical structure and the occupants' evaluation of the surrounding area's suitability for residential use.

In the past the Census Bureau has used several different definitions of housing quality. The first attempt to measure housing quality was in the 1940 Census where the concept of "state of repairs" was used as an indicator. Each unit was classified as needing major repairs or not needing major repairs, depending on the physical condition of the structure. Classification did not depend on the type of structure; e.g., a tarpaper shack may have been classified as not needing major repairs. The results proved to be unsatisfactory and led to the development of the concept of "condition of structure". In the 1950 Census the categories dilapidated and not dilapidated were used to measure the condition of the structure. A dilapidated unit had one or more serious deficiencies, was of inadequate original construction to provide adequate shelter, or endangered the safety of the occupants. A series of minor deficiencies could lead to classification as delapidated. In the 1960 Census the definition of condition of structure was essentially unchanged but the classification was expanded to three categories: sound, deteriorating, and dilapidated. Since then the classification of units by the condition of the structure has been discontinued.

In each of the 1940-1960 Censuses enumerators were required to make an overall judgment according to specified criteria. The judgment was to be made for each unit separately, disregarding the neighborhood, the age of the structure, and the characteristics of its occupants.

A more widely accepted concept of housing quality is that of "substandard housing" which incorporates both the structural condition of the unit and the availability of specific plumbing facilities. Units are classified as substandard or standard. Although this classification is not used explicitly in Census Bureau publications, tabulations are produced to which the labels standard and substandard could be directly applied. The classification is widely accepted since it can be used nationwide and it embodies the criteria of hazards to health, safety, and welfare.

Intuitively it would seem to be more appropriate to base the classification of units or neighborhoods on an objective evaluation by a trained interviewer than on a more subjective evaluation by the occupants. In a study based on 1976 AHS data (Robert Marans, "The determinants of neighborhood quality: An analysis of the 1976 Annual Housing Survey", AHS Studies No. 3, 1979), it was found that the primary contributor to respondents' evaluation of the quality of their neighborhood is their (subjective) feeling about conditions such as general neighborhood upkeep, street noise, and crime. However, the consistency of the interviewers' ratings are open to question, especially when an overall rating is required. In a Census Bureau study of the 1960 Census statistics on housing quality (Working Paper No. 25 entitled "Measuring the quality of housing: An appraisal of Census statistics and methods", 1967), the overall quality of the data was found to be poor. In fact, small area statistical analyses tended to provide more accurate indicators of the structural condition of housing on a unit basis than did enumerator ratings in some areas of the country.

In a more recent Texas study (William Schucany, et al. (1979), "Analysis of the reliability of a new scale for housing quality", Journal of Statistical Planning and Inference 3, 305-313), the use of a booklet of photographs to illustrate the rating scale for each component of a housing quality index enabled enumerators to provide much more reliable data. The results obtained from componentwise evaluation by enumerators followed by the use of an index compared favorably with assessments made by experienced city housing inspectors. The photographs apparently enabled the enumerators to make better judgments on fragments of the total appearance than on the overall quality of the unit, and the index successfully converted the components into an overall rating.

The data collected on sample units would be used to determine neighborhood quality which, in turn, would be used to determine the control factor cell for all units in the neighborhood. Two studies provide evidence that the sample unit to neighborhood link can be made. In Working Paper 25, it was found that, despite the overall poor quality of the data, there was a high degree of correlation for Census tracts in large cities between objective measures of quality and the proportion of units with low structural quality as determined by enumerators. The correlations varied directly with the size of the area. In a more recent study based on a sample of 1976 AHS data (William Bielby, "Evaluating measures of neighborhood quality in the Annual Housing Survey", AHS Studies No. 2, 1979), it was concluded that responses to neighborhood quality items in AHS tended to vary systematically by neighborhoods and can be used to reflect variation among neighborhoods.

Any definition of the concept of "neighborhood quality" should contain two major components. The first should deal with the amount of substandard housing in the neighborhood. The second component should incorporate other non-housing variables in a role analogous to that played by plumbing facilities in the definition of substandard housing. Examples of such variables currently available in AHS include the level of street repair, street traffic, neighborhood crime, litter, abandoned structures, industrial activity, and noise and odors. The first component is determined by the physical condition of the individual units while the second is a function of the general living environment. The latter can be summarized by an index of neighborhood non-housing conditions. Statistically, the index could be defined as the first principal component from a principal component analysis of a selected set of variables currently included in AHS.

Assuming that a housing quality factor is defined at a neighborhood level, independent control totals are required at the same level of aggregation. One possible source of controls is adjusted Census tabulations. Counts obtained from the most recent Census could be adjusted to account for changes in the housing inventory over time. A similar idea is used to obtain independent estimates for CPS age, race, sex cells, but the methodology would necessarily be different. An example of an approach which may successfully provide independent estimates is described in a report by John Weicher et al. entitled "National housing needs and quality changes during the 1980's" (AHS Study No. 10, 1980). The methodology would require refinement and testing before it could be applied to AHS.

In summary, a housing quality control factor is feasible. It should be possible to obtain independent controls by adjusting or projecting counts from the previous Census. More research involving housing subject matter specialists would be required before such a factor could be incorporated into the AHS estimation procedure.

8.2 Relative Income Factor

Since many of the tabulations produced from AHS involve financial characteristics, household income may be a useful control factor. The current AHS questionnaire contains detailed items on income, both sources and amounts, so that total household income is available. The income reference period is the twelve months preceding the interview.

Total household income is, to some extent, a function of the geographical location of the sampled unit. Individuals with similar jobs in different parts of the country may have widely varying incomes, depending on the local economy. From another viewpoint, the income of a "middle class" household in rural Texas undoubtedly provides a lower standard of living in Washington, D.C. Hence, total household income, while accessible, will not by itself be an adequate basis for a control factor. Some adjustment must be made.

Two types of modifications can be studied. First, a direct adjustment can be made by replacing total income by income relative to a measure of the cost of living of the area or to a typical income based on household characteristics such as size and composition. The second alternative is an indirect adjustment; a cross-classification which would be used in an additional ratio adjustment stage in the weighting scheme.

There are many approaches which might be used to_determine a cost of living factor. Among them are the use of

1. an index such as the Consumer Price Index (CPI),
2. a measure of the poverty rate for the area,
3. a measure of "typical" household income for the area.

The CPI measures average changes in the prices of goods and services, both overall and for selected categories of items. It is computed at the national and regional levels as well as for selected SMSAs. However, the BLS Handbook of Methods (Chapter 13) indicates that rural households are not included in the CPI. In addition, geographical area indices do not measure relative prices or living costs between areas. Hence, the CPI does not appear to be useful in adjusting income.

The CPS produces poverty levels and rates by various demographic characteristics as well as by region and residence status (cf., CPS Reports, Series P-60). These could be used as a measure of the cost of living for an area since poverty thresholds only vary by family size, sex of the head, number of children, and farm-nonfarm status. However, care must be exercised in the use of poverty information since the rates are computed on a person basis rather than on a household basis; i.e., the poverty rate is the ratio of the number of persons below the appropriate threshold to the population of the area. An area with a few large households below the poverty threshold and an area with many small households in poverty could have identical rates.

The use of some measure of "typical" household income may be the most promising of the three methods. A measure of central tendency such as the median income for households of a particular size and composition could be used as a benchmark for adjustment. Other percentiles could be used in place of the median.

Since a direct adjustment based on some measure of the cost of living has many problems, the inclusion of an additional ratio adjustment stage in the weighting procedure may be a more reasonable alternative. An examination of the poverty thresholds reveals a potential second factor to be used with an income factor to define a cross-classification. The additional factor could have categories defined by the size of the household, sex of the head, and a residence status breakdown (e.g., urban-rural or farm-nonfarm). Such a factor which follows the major divisions used for poverty determination should have an effect similar to a direct adjustment to income based on the cost of living.

There are many sources of independent controls for an income related factor. Both CPS and SIPP collect income data which is independent of AHS and could be used as controls. It may also be possible to use income distributions obtained from the Internal Revenue Service as controls. Regardless of the source, any differences in the definition of household income and in the time period in which the income is obtained must be taken into account.

In summary, an income related control factor may be both useful and feasible. However, much more research must be undertaken before such a factor could be incorporated into the AHS weighting scheme.

9. Conclusions, Recommendations, and Future Research

The research described in this report deals with the redesign of the AHS weighting procedure beyond the first stage ratio adjustment. The principal topics investigated are

1. the stability of the estimates of selected characteristics over time,
2. the improvement of the current independent control totals,
3. alternative second and third stage control cell configurations for the current control factors,
4. the use of raking,
5. new control factors.

The conclusions and recommendations for each topic are summarized below, followed by a brief discussion of items for future research.

In our research, the estimated coefficients of variation (CVs) are used to mesure stability over time. The data series includes all ten AHS. Taking into account the various sample reductions and changes in population characteristics which heavily influenced some evaluation variables, the estimates from AHS as represented by the fourteen evaluation variables are generally stable over time. However, several variables do have consistently large CVs and an attempt should be made to remedy this problem. Since estimators of change tend to have larger CVs than the corresponding estimators of level, the problem is more serious for them.

The research on the improvement of the current independent control counts produced two important recommendations, the first dealing with mobile home controls and the second with the estimation of the number of occupied units. Until recently there have been no independent control counts available for mobile home cells in the second stage ratio adjustment. Since controls are now available, it is recommended that they be used in future AHS. Initially a simple average of several quarters of control data should be sufficient.

The regression procedure currently used to obtain an estimate of the total number of occupied units from CPS for use in the third stage ratio adjustment should be changed. The present procedure does not take into account the correlation structure in the data series. A time series model such as that described in Section 5.1 is recommended as a replacement.

The comparisons of alternative sets of control cell configurations for the second and third stages are based on the estimated CVs of the evaluation variables and on sums of their relative variances. Four sets of second stage cell configurations were considered. Based on these comparisons, it is recommended that the current regional division be retained and the time of construction cells be collapsed into three broad categories; construction prior to the last Census, construction from the last Census to the previous AHS, and construction since the last AHS. The present housing structure type subdivisions within each time category should be retained.

The research into alternative third stage control cell configurations was carried out in two phases; changes in the configuration for occupied units were investigated first and the optimal configuration for occupied units was used when vacant unit cell configurations were studied. In all cases the second stage configuration recommended above was used. Unfortunately, the research involving the vacant units is clouded by problems with our handling of seasonal and migratory vacant units in the longitudinal file (cf., Section 6.2). We do believe, however, that comparisons among the various configurations for vacant units are valid. The actual numerical estimates of level and change are suspect.

Eight occupied unit and four vacant unit configurations were considered. Based on these comparisons, it is recommended that the current third stage control cell configuration for occupied units be retained. It is also recommended that vacant units be divided into only two cells; year round vacants and seasonal and migratory vacants.

Since the independent control totals for the second stage construction factor are not as reliable as controls should be, the possibility of eliminating the second stage ratio adjustment was investigated. In general the estimated CVs for the evaluation variables under the optimal third stage configuration did not change substantially when the second stage was eliminated. However, more extensive research should be conducted before recommending removal of the present second stage adjustment from the weighting scheme.

The present AHS weighting scheme contains two cycles of a raking procedure. Each cycle consists of two steps corresponding to a reapplication of the second and third stage ratio adjustments. The convergence criterion is based on the difference between the sample and control totals for each second and third stage cell. Using this criterion it is recommended that additional cycles be included in the weighting scheme. Reasonable agreement is attained after ten cycles. These results are not affected by reversing the order of the steps within each cycle. Estimated variances of the evaluation variables are not significantly reduced after the first few cycles.

Initial research into two new control factors was conducted. The first factor investigated is based on the quality of housing, both for individual units and at the neighborhood level. The primary problems with such a factor are creating a workable definition of substandard housing with which reliable data can be obtained and finding a source of independent control counts. The second factor is based on household income. The most promising approach is the creation of a new ratio adjustment stage which consists of a cross-classification defined by an income factor and a demographic type factor that serves to distinguish differences caused by variations in household composition and costs of living. Both of these new factors merit additional research.

The major recommendations discussed above can be summarized as follows:

1. a time series model should be used to obtain the total number of occupied units,
2. the number of categories in the second stage consturction factor and the vacant unit portion of the third stage configuration should be reduced,
3. additional raking cycles should be included in the weighting procedure,
4. additional research should be conducted on ways of improving construction controls and/or replacing the construction factor by an income related factor.

Several topics for future research have been uncovered in the course of our investigations. The two most important are listed in item four above; determine the feasibility of new control factors and improve the independent controls for the construction factor. The first topic is discussed in Section 8. The second offers several interesting possibilities. Should the construction factor be eliminated? If not, how can the Survey of Construction (SOC) be improved to provide better controls? Can SOC and AHS data be combined to produce better controls than can be obtained from SOC alone? This last question is part of the more general problem of combining information from several surveys.

Acknowledgments

Many individuals have provided assistance during this research. In addition to those cited in the text of this report, we would like to thank Dick Bartlett and Peggy Goldsmith (DSD) for providing the longitudinal data file and related documentation, Dennis Schwanz and Jim Hartman (SMD) for assistance in shaping the research plan and for patiently answering, often more than once, many detailed questions about AHS, Larry Ernst (SRD) for many helpful discussions, Diane Bailie (SRD) for carrying out the computer work required for the raking research and for constructing the figures in this report, and finally, Alice Bell (SRD) for typing (and retyping and retyping...) this report.

Appendix 1

We shall formulate the Taylor series method of variance estimation in terms which are sufficiently general to cover all applications in this report. The derivations are not intended to be mathematically rigorous. To establish notation, for control cells $i = 1, \ldots, I$, define

x_i = sample total for the characteristic of interest (here the evaluation variable) for cell i before ratio adjustment,

y_i = sample total for all sampled units for cell i before ratio adjustment,

X_i = population total for the characteristic of interest for cell i,

Y_i = population total for cell i.

Let r_i be the ratio adjustment factor for cell i; i.e., $r_i = Y_i/y_i$. Let $f_i = X_i/Y_i$. Then the ratio estimator for the level of the characteristic of interest after adjustment to the control totals is given by

$$\hat{x}_R = \sum_i r_i x_i$$

$$= \sum_i Y_i\left(\frac{x_i}{y_i}\right) \quad . \qquad (A1)$$

The ratio estimator $\hat{x}_R$ can be approximated by expanding each cell ratio x_i/y_i in (A1) in a Taylor series about the corresponding population ratio X_i/Y_i. The approximate mean squared error of $\hat{x}_R$ can be obtained from the resulting expansion.

In general, if $f(x,y) = x/y$, then the first order Taylor series expansion about X/Y is given by

$$\frac{x}{y} = \frac{X}{Y} + (x-X)\frac{1}{Y} - (y-Y)\frac{X}{Y^2} + o(|x-X|,|y-Y|),$$

so that, assuming $Y \neq 0$ and y is bounded away from zero, as $x \to X$ and $y \to Y$,

$$\frac{x}{y} \doteq \frac{X}{Y} + (x-X)\frac{1}{Y} - (y-Y)\frac{X}{Y^2} \quad . \qquad \text{(A2)}$$

Applying (A2) to the ratio estimator $\hat{x}_R$ under the assumption that x and y are random variables and that the cell sample sizes are sufficiently large to insure the validity of the Taylor series expansion, we have

$$\begin{aligned}
\hat{x}_R - X &= \sum_i \left[Y_i\left(\frac{x_i}{y_i}\right) - X_i\right] \\
&= \sum_i Y_i\left[\frac{x_i}{y_i} - \frac{X_i}{Y_i}\right] \\
&\doteq \sum_i Y_i\left[(x_i - X_i)\frac{1}{Y_i} - (y_i - Y_i)\frac{X_i}{Y_i^2}\right] \\
&= \sum_i (x_i - f_i y_i) \\
&= \sum_i z_i \quad ,
\end{aligned}$$

where $\qquad z_i = x_i - f_i y_i \quad .$

Hence, we have

$$\begin{aligned}
(\hat{x}_R - X)^2 &\doteq \left(\sum_i z_i\right)^2 \\
&= \sum_i z_i^2 + 2\sum_{i<j}\sum z_i z_j \quad .
\end{aligned}$$

Although expectations cannot be taken across the approximation sign, it can be shown rigorously that the aysmptotic mean square error of $\hat{x}_R$ is given by

$$MSE(\hat{x}_R) = E[(\hat{x}_R - X)^2]$$

$$= \sum_i E[z_i^2] + 2 \sum_{i<j}\sum E(z_i z_j) \quad . \qquad (A3)$$

If we assume that x_i and y_i are unbiased estimators of the parameters X_i and Y_i, respectively, then $E[z_i] = 0$ and $E[z_i^2] = Var(z_i)$. If z_i and z_j are uncorrelated for $i \neq j$, as would be the case if the cells were formed from stratification rather than post-stratification, then the second term on the right hand side of (A3) would be zero. For sufficiently large sample sizes, this assumption may be reasonable (cf., W. Cochran (1977), Sampling Techniques, third edition, section 5A9). It should be noted that since the estimates from (A3) will be used to compare competing control cell configurations, we need only assume that ignoring the crossproduct terms in (A3) will not change the relative ordering of the configurations. In any case, the second term will be dropped so that (A3) reduces to

$$MSE(\hat{x}_R) = \sum_i Var(z_i)$$

$$= \sum_i Var(z_i^{SR}) + \sum_i Var(z_i^{NSR}) \ , \qquad (A4)$$

where z_i^{SR} and z_i^{NSR} represent the contributions of the SR and NSR strata to the variance from the i-th control cell, respectively.

Following the method developed in Technical Paper 40

(Appendix K) for CPS, the contribution to the variance by the SR strata can be estimated by forming an average from replicate half samples. The estimator is given by

$$\hat{Var}(z_i^{SR}) = \frac{1}{R} \sum_{r=1}^{R} \sum_{s=1}^{SR} (\hat{z}_{irs1} - \hat{z}_{irs2})^2 , \qquad (A5)$$

where $\hat{z}_{irsh}$ is the value of z_i for the s-th SR stratum obtained from the h-th half sample in the r-th replicate with X_i replaced by x_i; i.e.,

$$\hat{z}_{irsh} = x_{irsh} - (\frac{x_i}{Y_i}) y_{irsh} .$$

Using a collapsed strata grouping of the NSR strata, the variance contribution from the NSR strata can be estimated by

$$\hat{Var}(z_i^{NSR}) = 4 \sum_{s=1}^{NSR} (w_{s2}\hat{z}_{is1} - w_{s1} \hat{z}_{is2})^2 , \qquad (A6)$$

where $\hat{z}_{ish}$ is the z_i value from the h-th sampled PSU in the s-th collapsed NSR stratum and w_{sh} is the proportion of the s-th collapsed stratum size associated with the original stratum from which the h-th PSU was selected.

The estimator of change in the level of the characteristic of interest is obtained similarly. Letting $\Delta\hat{x}_R = \hat{x}_R^{80} - \hat{x}_R^{78}$ represent the estimator of change, we have

$$\Delta\hat{x}_R - \Delta X = (\hat{x}_R^{80} - X^{80}) - (\hat{x}_R^{78} - X^{78})$$

$$= \sum_i [(x_i^{80} - f_i^{80} y_i^{80}) - (x_i^{78} - f_i^{78} y_i^{78})]$$

$$= \sum_i \left(z_i^{80} - z_i^{78}\right)$$

$$= \sum_i \Delta z_i \; . \qquad \text{(A7)}$$

The remainder of the derivation of the approximation for $MSE(\Delta\hat{x}_R)$ proceeds as above with z_i replaced by Δz_i.

Appendix 2

In response to comments from SMD personnel on an earlier draft of Section 7 of this report, the convergence criterion for the raking algorithm was changed and the programs were rerun using the new criterion. In Section 7.2 the algorithm was said to have converged if the absolute value of the differences between the marginal control and sample totals are all less than the selected tolerance. In this appendix, the convergence criterion is given in terms of the absolute differences between the ratio of the marginal control to sample totals and 1.0.

Several tolerance levels were considered. The number of raking cycles required to achieve convergence for each tolerance is given in Table 23. As expected, the number of cycles increases as the tolerance decreases. When compared to the middle portion of Table 20 (p. 54), there is essentially no difference in the number of cycles required for convergence using the two criteria.

Table 23. The Number of Cycles Required for Convergence for Various Tolerance Levels

Tolerance	Number of Cycles
0.0001	33
0.001	24
0.01	14

The convergence of the raking procedure was examined more thoroughly for a tolerance level of 0.001. A contour plot of the differences between the ratios of the marginal control to sample total and 1.0 is shown in Figure 11. Cycle 0 represents the disposition of the table before raking. The coding for the contours is given in Table 24. A positive value in the plot indicates a ratio greater than 1.0.

Table 24. Coding for the Contour Values in Figure 11

Range of the Absolute Difference*	Code (Unsigned)
0 to 0.001	0
0.001 to 0.005	1
0.005 to 0.01	2
0.01 to 0.05	3
0.05 to 0.1	4
0.1 to 0.5	5
0.5 to 1.0	6
1.0 to 1.5	7
1.5 to 2.0	8
greater than 2.0	9

*Difference = (control total/sample total) - 1.0.

In Figure 11 the problem with the vacant units in the data file becomes evident. The cells involving these units are the last to converge in both the second and third stage configurations. Without these cells it appears that ten to twelve

Construction Controls (Second Stage Cells)

	NE								NC								S								W							
	Old		Recent			New			Old		Recent			New			Old		Recent			New			Old		Recent			New		
Cycle	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	3
0	3	-5	3	1	4	6	7	-4	-4	-5	4	4	1	7	6	-5	-5	-5	-3	5	5	6	5	-5	5	-5	3	5	3	7	4	-5
1	1	5	1	4	-2	3	4	-3	-2	5	-2	4	-3	4	5	-3	-3	4	-2	4	-3	4	5	-3	3	-6	1	4	-3	4	5	-3
2	2	1	1	3	0	3	3	0	1	3	1	3	0	3	3	-1	1	-4	1	3	-1	3	3	-1	2	-5	1	3	-1	3	3	0
3	2	-3	1	2	1	0	1	1	1	-3	1	2	1	0	0	1	1	-4	1	1	1	0	-1	1	2	-5	1	1	1	-1	-1	1
4	1	-4	1	1	1	-1	0	1	1	-3	1	1	1	-1	-2	1	1	-4	1	-1	1	-1	-3	1	1	-4	1	0	1	-2	-3	1
5	1	-3	1	0	1	-1	-1	1	1	-3	1	0	1	-1	-2	1	1	-4	1	-1	1	-1	-3	1	1	-4	1	-1	1	-2	-3	1
6	1	-3	0	0	1	-1	-1	1	1	-3	1	0	1	-1	-2	1	1	-3	1	-1	1	-1	-3	1	1	-4	1	-1	1	-2	-3	1
7	1	-3	0	0	1	-1	-1	1	1	-3	1	0	1	-1	-2	1	1	-3	1	-1	1	-1	-2	1	1	-3	1	-1	1	-2	-2	1
8	1	-3	0	0	1	-1	-1	1	1	-3	1	0	1	-1	-2	1	1	-3	1	-1	1	-1	-2	1	1	-3	0	0	1	-1	-2	1
9	1	-3	0	0	1	-1	-1	1	1	-3	1	0	1	-1	-1	1	1	-3	0	-1	1	-1	-2	1	1	-3	0	0	1	-1	-2	1
10	1	-3	0	0	1	-1	0	1	1	-3	0	0	1	-1	-1	1	1	-3	0	-1	1	-1	-1	1	1	-3	0	0	1	-1	-1	1
11	1	-3	0	0	1	-1	0	1	1	-3	0	0	1	-1	-1	1	1	-3	0	0	1	-1	-1	1	1	-3	0	0	1	-1	-1	1
12	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	-1	0	0	-3	0	0	0	-1	-1	0	0	-3	0	0	0	-1	-1	0
13	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	-1	0	0	-2	0	0	0	0	-1	0	0	-3	0	0	0	-1	-1	0
14	0	-2	0	0	0	0	0	0	0	-2	0	0	0	0	-1	0	0	-2	0	0	0	0	-1	0	0	-2	0	0	0	-1	-1	0
15	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0	0	-2	0	0	0	0	-1	0	0	-2	0	0	0	-1	-1	0
16	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0	0	-2	0	0	0	0	-1	0
17	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
18	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
19	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
20	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
21	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0

Figure 11. A contour plot of the absolute differences between the ratio of the control to marginal totals and 1.0

CPS-HVS Controls (Third Stage Cells)

	Central City								Balance SMSA								Non-SMSA								Vacant	
	Owner				Renter				Owner				Renter				Owner				Renter					
Cycle	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
0	5	3	4	-4	-3	4	5	3	4	3	-3	-4	-3	-3	-3	4	3	-3	5	4	-4	4	5	4	-5	9
1	4	-2	5	-3	-5	-3	5	3	3	-3	-3	-4	-4	-4	-4	2	3	0	5	5	-4	3	5	5	-5	9
2	-1	-1	1	2	-3	-3	-2	-2	-1	-1	2	1	-3	-3	-3	-3	2	2	3	3	-2	-3	2	1	-4	6
3	-1	-1	-1	-1	-2	-2	-2	-2	-1	-1	-1	-1	-2	-2	-2	-2	-1	-1	-1	-1	-2	-2	-1	-1	3	5
4	-1	-2	-1	-2	-2	-2	-1	-2	-1	-1	-1	-1	-1	-2	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	3	5
5	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	4
6	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	4
7	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	3
8	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	3
9	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	3
10	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	3
11	-1	-1	-1	-1	0	0	0	-1	-1	-1	-1	-1	0	0	0	0	-1	-1	-1	-1	-1	0	-1	-1	3	3
12	0	0	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 11. (continued)

cycles would be needed to attain convergence. Thus, the basic conclusion reached in Section 7 is unaltered; additional raking cycles should be incorporated into the AHS weighting procedure.

An argument against the use of the ratio of the marginal control to sample total to define the convergence criterion was given in Section 7.2. The argument is based on the fact that, for a fixed difference, the ratio depends on the magnitude of its components as well as on their difference. From another perspective, given a fixed ratio (or equivalently, a fixed tolerance level), the difference between the components of the ratio must increase as the value of the components increases. This can have undesirable effects on estimates constructed from such data.

As an example of the problem, suppose we want to compare the number of owner occupied units inside an SMSA but not in the central city for blacks and whites. The estimates in Current Housing Reports, Series H-150-83 for 1983 (Table A1) are 21260 and 989 for whites and blacks, respectively (All estimates are in thousands.). For purposes of illustration, suppose that these are the true values. If a tolerance of 0.01 is used, the sample total for whites could differ from the true value by as much as 212.6. This is approximately 21.5% of the black total. For a tolerance of 0.005, the difference could be as large as 106.3, or approximately 10.7% of the black total. Sampling variability and nonsampling errors could, in effect, increase these percentages. In light of this, the comparison could be almost meaningless.

Although the example given above technically could not occur since it involves third stage control cells which are part of the final ratio adjustment, the point is clear. There is a danger that large differences allowed by a convergence criterion based on ratios could make comparisons involving both large and small cells meaningless. From Figure 11, many cells have differences in the 0.005 to 0.01 range after three raking cycles. Thus, the example also reinforces the conclusion that additional raking cycles need to be included in the weighting scheme.

Appendix 3

In his review of the final draft of this report, Gary Shapiro (SMD) commented that the use of CVs to measure the relative precision of estimators of change for the evaluation variables may be misleading. If the estimated change in level is small, then the CV will be sensitive to small perturbations in the estimated change. Since even moderate changes in these CVs could influence the overall conclusions of the comparison study of alternative control cell configurations (especially for the third stage), he suggests that the estimator of change in the denominator of the CV be replaced by the estimator of the level for one of the AHS involved in the change calculation. This modification provides a more stable measure of the relative precision of the estimated change which is similar to a CV. Formally, the estimated relative precision (ERP) of the estimator of the change in level is defined to be

$$ERP(\hat{\Delta}) = 100 \cdot se(\hat{\Delta}) \, / \, E(\hat{\theta}_{80}) \quad , \tag{A8}$$

where

$\hat{\theta}_i$ is the estimator of the level for year i (i = 78, 80),

$\hat{\Delta} = \hat{\theta}_{80} - \hat{\theta}_{78}$.

The ERPs as defined in (A8) were computed for all second and third stage cell configurations included in the research described in Section 6. The results for the four second stage configurations defined in Section 6.1 (cf., Table 5, p. 27) are presented in Table 25. Tables 26 and 27 give the estimated ERPs for the eight third stage occupied unit control cell configurations (cf., Table 11, p. 36) and the four third stage vacant unit control cell configurations (cf., Table 15, p. 43), respectively.

Table 25. Estimated Relative Precision for the Estimators of Change for the Second Stage Control Cell Configurations

	Configuration			
Variable	A	B	C	D
Coops & condos	3.5*	3.6	8.2	12.0
1-unit structures	0.2	0.2	0.6	0.7
Mobile homes	1.5	2.1	4.4	4.2
Lacking plumbing	2.3	2.3	5.0	5.1
Vacant rentals	3.7	3.7	10.4	10.4
Black head	0.9	0.9	1.1	1.1
Female head	0.7	0.7	1.7	1.7
Low income	1.0	1.0	2.4	2.4
Low value	2.5	2.5	5.4	5.5
Low rent	1.3	1.3	3.1	3.1
High income-rent ratio	10.5	10.6	20.5	20.5
Urban year round	0.2	0.2	0.7	0.7
Persons 65+ alone	1.2	1.2	3.3	3.2
Old units	0.3	0.3	1.1	1.1

* Entries are percentages.

Table 26. Estimated Relative Precision for the Estimators of Change for the Third Stage Occupied Unit Control Cell Configurations**

Variable	Configuration A	B	C	D	E	F	G	H
Coops & condos	4.8*	4.9	5.0	5.2	4.8	4.9	5.0	5.1
1-unit structures	0.6	0.6	0.6*	0.6	0.6	0.6	0.6	0.6
Mobile homes	4.2	4.2*	4.2	4.2	4.2	4.2	4.2	4.2
Lacking plumbing	6.0	6.0	5.0	4.9*	6.2	6.2	5.1	5.1
Vacant rentals***	6.3	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Black head	0.2	0.2	0.2	0.2*	1.5	1.6	1.4	1.5
Female head	0.2	0.2*	1.4	1.4	0.2	0.2	1.4	1.4
Low income	2.6	2.7	2.2*	2.3	2.7	2.8	2.3	2.4
Low value	6.6	6.6	6.8	6.8	6.8	6.5	6.7	6.5*
Low rent	3.0	3.0	2.5*	2.5	3.1	3.1	2.6	2.6
High income-rent ratio	21.4	21.4	21.2*	22.2	21.4	21.6	21.2	22.3
Urban year round	0.6	0.6	0.5	0.6	0.6	0.6	0.5*	0.5*
Persons 65+ alone	3.0	2.9*	3.4	3.3	3.0	3.0	3.4	3.3
Old units	1.1	1.1	0.9	0.9	1.1	1.1	0.9*	0.9

* The configuration with the smallest ERP for that evaluation variable (using unrounded figures).
** Entries are percentages.
*** The ERPs for vacant rentals are identical since no occupied units are used in the calculation.

Table 27. Estimated Relative Precision for the Estimators of Change for the Third Stage Vacant Unit Control Cell Configurations

	Configuration			
Variable	A	I	J	K
Coops & condos	4.8*	4.8	4.8	4.8
Vacant rentals	6.3	6.0	11.6	10.9
Urban year round	0.6	0.6	0.6	0.6
Old units	1.0	1.2	1.0	1.0

* Entries are percentages.

As expected, the estimated ERPs in Tables 25, 26, and 27 are generally smaller than the corresponding estimated CVs in Tables 8, 13, and 16, respectively. However, the conclusions are essentially the same. From Table 25, second stage configurations A and B have smaller ERPs than do configurations C and D for each evaluation variable. The choice betwen configurations A and B again is not determined by the estimates of relative precision. As in Section 6.2, configuration B is used as the basis for the third stage calculations.

The ERPs given in Table 26 for the third stage occupied unit control cell configurations do not provide a clearcut candidate for the best configuration; the same conclusion was drawn from Table 13 using CVs. As before, the current third stage configuration will be used as the basis for the vacant unit control cell configuration comparisons.

The results in Table 27 differ from those in Table 16 for the estimate of change in the number of vacant units for rent. Consistency is maintained for the remaining variables. However, the conclusion based on the ERPs in Table 27 supports rather than contradicts the results obtained from the estimators of the 1980 levels in Table 16; viz., the only substantial difference between the four configurations is in the number of vacant units for rent. Configurations A and I have the best relative precisions.

In summary, the use of ERPs instead of CVs for the estimators of change in levels does not affect the conclusions for the second stage and the third stage occupied unit control cell configurations. That is, it is recommended that the second stage breakdown of the time of construction factor be reduced to the three broad categories defined on page 26 and that the present housing structure subcategories be retained. The current third stage occupied unit control cells should also be retained. The selection of a third stage vacant unit control cell configuration, although not clearcut in Section 6.2, is even more dependent on a criterion other than the relative precision of the evaluation variables.