Survey of Income and Program Participation

Composite Estimation For
SIPP Annual Estimates

8911 88

Rameswar P. Chakrabarty
U.S. Bureau of the Census

May 1989

TABLE OF CONTENTS

COMPOSITE ESTIMATION FOR SIPP ANNUAL ESTIMATES

1. INTRODUCTION

Composite Estimation for SIPP - A Preliminary Report [1] has identified the types of composite estimators that can be used in SIPP to improve the reliability of estimates, particularly quarterly estimates. The quarterly reports, Economic Characteristics of Households in the United States, were published for the four quarters of 1984. At present, publication of quarterly reports has been discontinued primarily because economic characteristics of households and persons do not appear to change significantly from quarter to quarter. Annual reports on the economic characteristics of households in the United States are expected to be published in the future. In view of this publication plan, research on the composite estimation for SIPP annual reports was undertaken. This report summarizes results of this research and recommends the composite estimator to be used for SIPP annual reports.

2. ESTIMATION

The analysis of the SIPP data has two major aspects, cross-sectional and longitudinal analyses. Our research is confined to the estimation for cross-sectional reports. Estimation of annual levels (monthly averages for a year) and year to year change in annual levels are of major interest for cross-sectional reports. Composite estimation for quarterly levels and year to year change in quarterly levels was discussed in an earlier report [1] but will not be researched further.

For the current year, the usual ratio estimator uses data only for the current year. This estimator is the result of four steps in the estimation procedure. These steps are (1) computation of simple unbiased estimates, (2) noninterview adjustment, (3) first stage ratio adjustment and (4) second stage ratio adjustment. In this report, the ratio estimator will be called "Simple" (uncomposited) estimator. A composite estimator makes use of data from the current as well as previous occasion(s). Different composite estimators employ different methods of combining data from the current and previous occasion(s).

The SIPP data structure for annual estimates is given in the Attachment II. It will be seen that from 1985 each year has two panels that provide an annual estimate. Some panels do not provide an annual estimate, e.g., 1984 panel provides data only for the first two quarters of 1986, and 1987 panel provides data only for the 4th quarter of 1986. Thus, some monthly data (about 9% to 20%) cannot be used in simple annual estimates and consequently, in the composite estimation based on simple annual estimates. On the other hand, annual estimates obtained by pooling quarterly composite estimates will use almost all monthly data, and therefore, likely to be more efficient. This approach discussed in the earlier report [1] is not viable now since quarterly estimates are no longer published. Therefore, composite estimation of annual levels and year to year change in levels based on simple annual estimates is discussed in the sequel.

Note that composite estimation for annual estimates can be done begining in the year 1986 and 50% of the sample ("designated" housing units) in a year is matched with the previous year. The actual proportion of the sample matched from year to year for households and particularly for persons and families may be somewhat less than that for housing units due to sample attrition resulting from movements of people and changes in families. However, the impact on the efficiency of a composite estimator of the level is generally likely to be negligible due to minor changes in the proportion of sample overlap. This is due to the fact that for estimates of levels, the optimum percent to be matched is no more than 50% (see Cochran,1963). The SIPP panel design is indeed optimal for the composite estimation of annual levels.

From our analysis of the SIPP rotation pattern and data structure (see Attachment II), and research in composite estimation methods we have determined that the AK composite estimator (Huang and Ernst, 1981) currently used in the CPS is applicable only to the CPS-type rotation patterns (4-8-4 or 3-9-3). This estimator cannot be used for the SIPP. The K composite estimator (see Technical Report 40) previously used in the CPS can be used but other minimum variance estimators will be more efficient. Three minimum variance composite estimators are available to improve the reliability of annual estimates. These are given by Cochran (1963), Wolter (1979) and Ernst and Breau (1983). A brief description of these estimators is given in the Attachment I. Relative variances of these estimators compared to the usual ratio estimator are given in Table 1. It can be seen that the composite estimation will reduce the variance of the annual estimate from 1% to 39% when correlation ranges from .2 to .9. (These are assumed correlations, actual correlations from SIPP data are not yet available). The variance of the estimate of change will be reduced by 1% to 59% for correlations ranging from .2 to .9. The correlation between family incomes reported to the IRS on two consecutive years was found to be .80 to .87. (Ponikowski and Tadros, 1984). The correlation between SIPP income data may be somewhat less because of the possibility of a higher response error in SIPP compared to income tax returns. However, when the correlation is .8, the variance of annual income can be reduced by 25% and the variance of the year to year change in income can be reduced by 38% by composite estimation. Thus, substantial gains in reliability of estimates can be achieved by composite estimation. Also, note that the gains in reliability from the composite estimation are much greater in estimating change than current level.

All three estimators are minimum variance estimators and are asymptotically equally efficient. The variances for Cochran and Ernst-Breau estimators given in Table 1 are limiting variances. However, the third or fourth year of composite estimation show that variances from these years are about the same as the limiting variances. The results from Wolter's method are based on assuming that ten years' data will be used in estimation. We have obtained the coefficients of simple annual estimates and variance-covariance matrices for Wolter's method for assumed correlations .2 to .9; the results for correlation .8 are given in Table 2. Wolter's method improves the reliability marginally at middle year(s), for example for the 5th and 6th years in estimation based on ten years' data. Also, note that the reliability of estimates of change is better in Wolter's method because estimates of levels for earlier years are revised using all data up to and including the latest year. The other two methods assume that estimates of levels will not be revised although it is possible to do so. In practice, the revision of estimates published earlier may not be desirable.

Like both the K and AK CPS composite estimators, Cochran and Ernst-Breau estimators require only storage of the previous year's data for computing the current year's estimate. Wolter's method, on the other hand, requires storage of simple estimates (two simple estimates from two panels in each year) of all previous years. This may present a difficult problem in terms of data storage for SIPP because of its numerous data items. Also, Wolter's method is much more computationally complex. For example, with ten years' data, this requires the inversion of 20x20 matrices and becomes a computational burden with a large number of survey characteristics. Data storage requirements and computational burden involved in Wolter's method can be eased by restricting the use of data to fewer years (say 5 years) with some loss in reliability. We have not yet evaluated the loss in efficiency that may result in using 5 years' data compared to all ten years' data. However, since all three estimators are minimum variance estimators, the choice of an estimator must depend on the data storage requirements and computational ease. Cochran's method requires the computation of regression coefficients and regression estimates for all survey items. The Ernst-Breau estimator requires only computation of correlations and thus is computationally easier than the Cochran estimator. We, therefore, recommend that the Ernst-Breau composite estimator be used for annual estimates.

3. ERNST-BREAU COMPOSITE ESTIMATOR

This estimator is described in the Attachment I. The composite estimator of level on the hth (current) year Y_h can be expressed in a recursive form as

$$Y_h=(1-A_h)X_{h,1}+A_hX_{h,2}+Z_1(Y_{h-1}-X_{h-1,1}),$$

where

$Z_1=[1-(1-\rho^2)^{½}]/\rho$,

$A_h=\dfrac{1}{2-\rho Z_1}$,

Y_{h-1} is the composite estimate of level in the h-1 (previous) year,

$X_{h,1}$ is the simple estimate of level for the current year based on the 50% of the sample that is not matched with the previous year,

$X_{h,2}$ is the simple estimate of level for the current year based on the 50% of the sample that is matched with the previous year,

and

$X_{h-1,1}$ is the simple estimate of level for the previous year based on the 50% of the sample that is matched with the current year.

Note that the coefficents A_h and Z_1 depend only on the correlation coefficent ρ. The composite estimator becomes the usual uncomposited estimator of the level for the current year when $Z_1=0$ and $A_h=.5$. The values of A_h and Z_1 are given in the Table 3 for correlations ranging from .05 to .95. It is interesting to note that the weight for the previous year's composite estimate (more specifically, for the difference between the composite estimate and matched sample estimate $Y_{h-1}-X_{h-1,1}$) increases as the correlation increases. Also, the weight A_h for the matched half-sample estimate in the current year increases with the correlation. For very low correlation ($\rho< .05$) this estimator reduces for practical purposes to the usual uncomposited estimator of level for the current year (i.e., $Z_1=0$, $A_h=.5$).

The limiting variance of Y_h is

$$V(Y_h) = \sigma^2\left[1-\frac{2}{(2-\rho Z_1)} + \frac{2(1-\rho Z_1)}{(2-\rho Z_1)^2(1-Z_1^2)}\right].$$

Turning to the estimation of year to year change note that theoretically the best (minimum variance) estimate of change can be obtained by recomputing the composite estimates of the earlier years utilizing data for the later years and then taking differences among revised composite estimates of levels. In practice, it may not be feasible to revise the earlier composite estimates because of data storage requirements and additional computing costs. As mentioned earlier, Wolter's method uses this approach. It can be seen from Table 1 that this approach reduces the variance of the change by additional 1% to 9% when the correlation ranges from .2 to .9 compared to Cochran's or Ernst-Breau's method where estimates for earlier years are not revised. Such marginal improvements in estimation of year to year change may not be worth the costs involved in storage of data and computation. In any case, the revision of estimates published earlier may not be desirable. Thus, we assume that the estimates published in earlier years will not be revised. Then, the estimator of change that is consistent with the estimates of levels is simply

$$D_h = Y_h - Y_{h-1}.$$

And its variance (algebraic details omitted) is given by

$$V(D_h) = 2\sigma^2\left[1 - \frac{2}{\theta_1} + \frac{2(1-\rho Z_1)}{\theta_2} - \rho\left(\frac{1}{\theta_1} - \frac{1+Z_1^2}{\theta_2}\right) - \frac{Z_1}{\theta_1} + \frac{2Z_1}{\theta_2}\right],$$

where $Z_1 = [1-(1-\rho^2)^{1/2}]/\rho$,

$\theta_1 = 2 - \rho Z_1$,

and

$\theta_2 = \theta_1^2(1-Z_1^2)$.

4. CORRELATION AND ROBUSTNESS OF THE COMPOSITE ESTIMATOR

As mentioned in the previous section, the optimal coefficients for the Ernst-Breau estimator depend only on the correlation. Year to year correlations for key SIPP characteristics are not available yet. The methodology for computing correlations was given in an earlier memorandum [2]. We have collected available information on year to year correlations from other household surveys. Results are sumarized below:

1. IRS Data

The correlation between family incomes (adjusted gross incomes) reported to the IRS in 1979 and 1980 was found to range between .80 to .87. The same results were also obtained between 1980 and 1981 family incomes. This indicates that the year to year correlation between family incomes is high (Ponikowski and Tadros, 1984). Correlation between family incomes in SIPP may be somewhat less because of the possibility of higher response error in SIPP compared to income tax returns. However, our results suggest that SIPP income data can be considerably improved by composite estimation. Many other SIPP characteristics are related to income and consequently estimates for such characteristics will also be improved by composite estimation.

2. AHS - National Sample

The year to year correlations of housing inventory estimates from AHS National Sample are given in Table 4. It will be seen that year to year correlations of housing inventory characteristics are rather high (.5 to .95). Only for few items like water leakage, breakdowns, leaky roof, etc is the correlation low, about .10-.15. Note that the correlation for income is .5.

3. CPS

The year to year correlations of CPS estimates are given in Table 5. Correlations range from very low to moderate. Out of 100 items, only a few (about 6) items show correlation .5 or above. Many items show correlation between .2 & .5. Several items have correlation close to zero and five items have very low negative (-.02 to -.09) correlation.

4. NCS - National Sample

Estimates of correlation coefficients between the 1975 and 1976 Annual Estimates of victimizations are given in Table 6. It can be seen that correlations range from moderately positive (.539 for household larcenices in metropolitan areas, .512 for total crimes of violence) to moderately negative (-.480 for robberies; inside central cities in metropolitan areas.) The correlations for victimization characteristics may not, however, be relevant to the SIPP.

As mentioned in the previous section, the coefficients Z_1 and A_h in the Ernst-Breau composite estimator depend only on the correlation. As in the CPS, year to year correlations are likely to be very low for some SIPP characteristics and high for others. Thus, the optimal values of the coefficients in the composite estimator will be different for different characteristics. In practice, however, the use of compromise values of Z_1 and A_h for all characteristics makes processing and tabulation of data simpler and estimates in a table retain the simple additive property. In this context, it is important to look at the robustness of the composite estimator with respect to departures from actual correlations. The relative variances of the Ernst-Breau estimator for estimates of level and change are given in Tables 7 and 8. The diagonal (underlined) elements in the Table 7 provide the variance of the composite estimator of the level relative to the uncomposited estimator if actual correlations are used in estimating the level. Off-diagonal elements provide the relative variances that will be obtained by using assumed correlations instead of actual correlations. For example, the relative variance is .928 for actual correlation .5 but it will be .967 if instead .7 is used for correlation. Thus, the reduction in the variance will be 3% rather than 7% but the composite estimator will still be more efficient than the uncomposited estimator. Note that the relative variances ≤ 1 indicate the range of departures from actual correlations for which the composite estimator will be at least as good or better than the uncomposited estimator. The results clearly show that the estimator is remarkably robust for slight departures from actual correlations. It does poorly only in situations where actual correlations are low ($\leq .20$) but high correlations ($\geq .70$) are used in estimation.

The results for relative variances for change (see Table 8) show that the change estimator is even more robust than the level estimator. These results are promising for using a "compromise" value for the correlation in determining the values of the coefficients Z_1 and A_h for the composite estimator. If key SIPP characteristics like income have high correlations and we are willing to accept inefficiencies for some characteristics which may have low correlations, a high value of ρ (say .7) should be used. If all characteristics are more or less equally important and/or if some key characteristics have low correlations, a moderate value of ρ (say .5) should be used. Note that the use of $\rho=.5$ in the composite estimator will be at least as good as the uncomposited estimator for characteristics with low correlation, and much better for characteristics with high correlation. From Table 3 it will be seen that $Z_1=.3$ and $A_h=.5$ for $\rho=.5$, and $Z_1=.4$ and $A_h=.6$ for $\rho=.7$. Also, note that the values of the coefficients are not all that different for $\rho=.5$, .6 or .7.

The use of appropriate pre-determined constant coefficients makes the composite estimator slightly biased but provides approximately minimum variance for many characteristics. This also provides more stable variance estimates by eliminating sampling variances of the coefficients computed from the sample correlation.

We, therefore, recommend using a "compromise composite estimator" for all characteristics. The compromise value of the correlation to be used for the estimator may be decided later after computation of correlations for SIPP characteristics and determination of relative importance of various characteristics.

5. Summary and Recommendations

1. The SIPP panel design with 50% of the sample matched from one year to the next is optimal for composite estimation of annual levels.

2. Composite estimation can be utilized from 1986 to improve the reliability of estimates of annual levels and change.

3. Three minimum variance composite estimators, Cochran (1983), Wolter

(1979) and Ernst-Breau (1983) are applicable but based on data storage requirements and computational simplicity the Ernst-Breau Estimator is recommended.

4. The coefficients of the Ernst-Breau Estimator depend only on the correlation and the estimator is quite robust with respect to departures from true correlations.

5. Ernst-Breau Estimator with a "compromise" value of the correlation for all characteristics is recommended for use.

6. A "compromise estimator" makes the processing and tabulation of data simpler and estimates in a table retain the simple additive property.

7. An additional benefit of a "compromise estimator" is that variance estimates are likely to be more stable.

8. The coefficients that may be used for the "compromise estimator" are indicated but should be decided after the computation of year to year correlations of various SIPP characteristics and determination of their relative importance.

9. Further Research Needed

   1. Compute year to year correlations of various SIPP characteristics.

   2. Determine the relative importance of various SIPP

characteristics in terms of publications and anticipated use for policy decisions and research.

3. Determine the exact coefficients to be used in the "compromise composite estimator."

References

1. Chakrabarty, R.P. (1986), "Composite Estimation for SIPP - A Preliminary Report." SIPP Working Paper Series No. 8610, U.S. Bureau of the Census, Washington, D.C.

2. Chakrabarty, R.P. (1986), "Computation of Year to Year Correlations of Key SIPP Characteristics for Composite Estimation Research." Census Bureau Memorandum, August 1, 1986, Statistical Research Division. Bureau of the Census, Washington, D.C.

3. Cochran, W.G. (1963), Sampling Techniques, John Wiley and Sons, New York.

4. Ernst, L., and Breau, P.(1983), unpublished manuscript, Bureau of the Census, Washington, D.C.

5. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953), Sample Survey Methods and Theory Vol.II. John Wiley & Sons, New York.

6. Huang, E.T. and Ernst, L.R. (1981), "Comparison of An Alternative Estimator to the Current Composite Estimator in CPS", Proceedings of the American Statistical Association, Survey Research Methods Section, 303-308.

7. Ponikowski, C. and Tadros, W. (1984), "Correlation Between Year to Year Incomes" Memorandum for Documentation, May 9, 1984, Statistical Methods Division, Bureau of the Census, Washington, D.C.

8. U.S. Dept of Commerce, Bureau of the Census, (1978), "The Current Population Survey: Design & Methodology" Technical Paper 40. Washington, D.C.

9. Wolter, K.M. (1979), "Composite Estimation in Finite Populations", Journal of the American Statistical Association 74, 604-613.

Table 1. The Relative Variance of Composite Estimators for Annual Levels and Change

Correlation	Composite Estimators		Relative Variance of Level	Relative Variance of Change
.2	Cochran	K = .4949 3/	.990	.989
	Wolter*		.990 1/	.983 1/
			.990 2/	.977 2/
.3	Cochran	K = .4882 3/	.976	.972
	Wolter*		.976 1/	.959 1/
			.954 2/	.950 2/
.4	Cochran	K = .4782 3/	.956	.946
	Wolter*		.956 1/	.922 1/
			.905 2/	.907 2/
.5	Cochran	K = .4641 3/	.928	.906
	Wolter*		.928 1/	.867 1/
			.866 2/	.845 2/
.6	Cochran	K = .4444 3/	.889	.847
	Wolter*		.889 1/	.790 1/
			.800 2/	.762 2/
.7	Cochran	K = .4166 3/	.833	.758
	Wolter*		.833 1/	.682 1/
			.714 2/	.650 2/
.8	Cochran	K = .3750 3/	.750	.625
	Wolter*		.750 1/	.531 1/
			.600 2/	.500 2/
.9	Cochran	K = .3036 3/	.607	.412
	Wolter*		.607 1/	.317 1/
			.456 2/	.297 2/

Note: The proportion of the sample (housing units) matched between two consecutive years is .5. The results from the Ernst-Breau Estimator are the same as those from the Cochran Estimator.

* Assumes ten years' data will be used and estimates for earlier years will be revised using later years' data.

1/ Relative variance for end (1, 10) years.

2/ Relative variance for middle (5th and 6th) years.

3/ Optimum value of K.

Table 2. Optimum Coefficients for Wolter s Estimators of Annual Levels $(\rho = .80)$.

Estimator										
$\hat{X}_{10}$	.3750004	.6249996	-.3124989	.3124989	-.1562477	.1562477	-.0781203	.0781203	-.0390530	.0390530
$\hat{X}_9$	.1875008	-.1875008	.4687519	.5312481	-.2656210	.2656210	-.1328044	.1328044	-.0663901	.0663901
$\hat{X}_8$	.0937515	-.0937515	.2343786	-.2343786	.4721951	.5078049	-.2538908	.2538908	-.1269222	.1269222
$\hat{X}_7$	.0468779	-.0468779	.1171947	-.1171947	.2461088	-.2461088	.4980774	.5019226	-.2509153	.2509153
$\hat{X}_6$	.0234432	-.0234432	.0586081	-.0586081	.1230769	-.1230769	.2490843	-.2490843	.4996338	.5003662
$\hat{X}_5$	.0117302	-.0117302	.0293255	-.0293255	.0615835	-.0615835	.1246333	-.1246333	.2499998	-.2499998
$\hat{X}_4$	.0058823	-.0058823	.0147057	-.0147057	.0308819	-.0308819	.0624991	-.0624991	.1253658	-.1253658
$\hat{X}_3$	.0029755	-.0029755	.0074387	-.0074387	.0156212	-.0156212	.0316143	-.0316143	.0634146	-.0634146
$\hat{X}_2$	.0015564	-.0015564	.0038910	-.0038910	.0081711	-.0081711	.0165367	-.0165367	.0331707	-.0331707
$\hat{X}_1$	.0009155	-.0009155	.0022888	-.0022888	.0048065	-.0048065	.0097275	-.0097275	.0195122	-.0195122

Estimator										
$\hat{X}_{10}$	-.0195122	.0195122	-.0097275	.0097275	-.0048065	.0048065	-.0022888	.0022888	-.0009155	.0009155
$\hat{X}_9$	-.0331707	.0331707	-.0165367	.0165367	-.0081711	.0081711	-.0038910	.0038910	-.0015564	.0015564
$\hat{X}_8$	-.0634146	.0634146	-.0316143	.0316143	-.0156212	.0156212	-.0074387	.0074387	-.0029755	.0029755
$\hat{X}_7$	-.1253658	.1253658	-.0624991	.0624991	-.0308819	.0308819	-.0147057	.0147057	-.0058823	.0058823
$\hat{X}_6$	-.2499998	.2499998	-.1246333	.1246333	-.0615835	.0615835	-.0293255	.0293255	-.0117302	.0117302
$\hat{X}_5$	.5003662	.4996338	-.2490843	.2490843	-.1230769	.1230769	-.0586081	.0586081	-.0234432	.0234432
$\hat{X}_4$	.2509153	-.2509153	.5019226	.4980774	-.2461088	.2461088	-.1171947	.1171947	-.0468779	.0468779
$\hat{X}_3$	.1269222	-.1269222	.2538908	-.2538908	.5078049	.4921951	-.2343786	.2343786	-.0937515	.0937515
$\hat{X}_2$	.0663901	-.0663901	.1328044	-.1328044	.2656210	-.2656210	.5312481	.4687519	-.1875008	.1875008
$\hat{X}_1$	.0390530	-.0390530	.0781203	-.0781203	.1562477	-.1562477	.3124989	-.3124989	.6249996	.3750004

Table 2a. Variance - Covariance Matrix of Annual Estimates
(Relative Variance - Covariance)

.37500045	.18750076	.09375145	.04687787	.02344323	.01173020	.00588226	.00297546	.00155640	.00091553
.18750076	.31875129	.15937747	.07969238	.03985349	.01994133	.00999985	.00505829	.00264587	.00155640
.09375145	.15937747	.30469222	.15235308	.07619049	.03812314	.01911736	.00967026	.00505829	.00297546
.04687787	.07969238	.15235308	.30119033	.15062274	.07536652	.03779355	.01911736	.00999985	.00588226
.02344323	.03985349	.07619049	.15062274	.30036635	.15029315	.07536652	.03812314	.01994133	.01173020
.01173020	.01994133	.03812314	.07536652	.15029315	.30036635	.15062274	.07619049	.03985349	.02344323
.00588226	.00999985	.01911736	.03779355	.07536652	.15062274	.30119033	.15235308	.07969238	.04687787
.00297546	.00505829	.00967026	.01911736	.03812314	.07619049	.15235308	.30469222	.15937747	.09375145
.00155640	.00264587	.00505829	.00999985	.01994133	.03985349	.07969238	.15937747	.31875129	.18750076
.00091553	.00155640	.00297546	.00588226	.01173020	.02344323	.04687787	.09375145	.18750076	.37500045

$\times 2$.

Table 3. Values of Coefficeints Z_1 and A_h for the Ernst-Breau Estimator

Correlation Coefficient ρ	z_1	A_h
.05	.0250	.5003
.10	.0501	.5013
.20	.1010	.5051
.30	.1535	.5118
.40	.2087	.5218
.50	.2679	.5359
.60	.3333	.5555
.70	.4084	.5834
.80	.5000	.6250
.90	.6268	.6964
.95	.7239	.7620

Table 4. Year to Year Correlation of Housing Inventory Estimates from AHS National Sample

Characteristic	Correlation Coefficient
1. All Units	.95
2. New Construction	.60
3. Mobile Homes	.90
4. Year Round Vacants	.80
5. In Central City	.95
6. Not in Central City	.95
7. Inside SMSA	.95
8. Outside SMSA	.95
9. Negro	.70
10. Spanish	.70
11. White	.50
12. Recent Movers	.20
13. Occupied Last Winter	.90
14. Male Head	.90
15. Female Head	.50
16. Year Built	.90
17. Plumbing Facilities	
a. With all	.90
b. Lacking some or all	.50
18. Kitchen Facilities	
a. For exclusive use of household	.90
b. Also used by another household	.50
c. No complete kitchen facilities	.50
19. Air Conditioning	.70
20. Sewage Disposal	.90
21. Garbage Collection	.80
22. Source of Water	.70
23. Cooking Fuel	.90
24. Heating Fuel	.70
25. Basement	
a. Water leakage	.15
b. No water leakage	.30
c. Other	.90
26. Heating Equipment	.60
27. Additional Heat	
a. Additional heat source	.60
b. No additional heat source	.15
28. Units in Structure	.70
29. Number of Rooms	.60
30. Number of Bathrooms	.70
31. Bedrooms	
a. None	.80
b. 1 Bedroom	.55
c. 2 Bedrooms	.45
d. 3+ Bedrooms	.85

Table 4 (cont'd)

Characteristics	Correlation Coefficient
32. Number of Automobiles	.50
33. Number of Persons	.50
34. Subfamilies	
a. Total or White	.50
b. Negro	.60
c. Spanish	.50
35. Persons per Room	
a. Total or White	.50
b. Negro	.50
c. Spanish	.60
36. Inclusion in Rent	.40
37. Mortgage	.70
38. Value	
a. Less than $20,000	.30
b. $20,000 to $40,000	.40
c. Greater than $40,000	.50
39. Contract Rent	.50
40. Gross Rent	.50
41. Income	.50
42. Fuse or Switch Blowout	
a. Yes, fuse blowout	.23
b. No, fuse blowout	.60
43. Water Supply Breakdown	
a. No breakdowns	.80
b. Completely without water	.10
44. Sewage Disposal Breakdown	
a. No breakdowns	.90
b. With breakdowns	.00
45. Flush Toilet Breakdown	
a. No breakdowns	.60
b. With breakdowns	.22
46. Heating Equipment Breakdown	
a. No breakdowns	.60
b. With breakdowns	.22
47. Other Breakdowns	
a. No breakdowns	.60
b. With breakdowns	.10
48. Items that can change in a year (leaky roof, mice or rats, holes in floor, etc.)	
a. With problems	.15
b. Without problems	.30
49. Opinions	
a. Excellent	.35
b. Good	.30
c. Fair	.30
d. Poor	.25
50. Public Housing	.40

Source: Memorandum from Jones to Young, October 14, 1976, Bureau of the Census.

Table 5. Year to Year Correlation of CPS Estimates

(1974-1975)

	Characteristics 1/	Correlation Coefficient
1.	Total Population	0.98
2.	Nonwhite Population	1.00
3.	Rural Nonfarm Population	0.64
4.	Total Rural Farm Population	0.42
5.	Central City Population	0.72
6.	SMSA Population	0.71
7.	Nonwhite SMSA Population	0.62
8.	Household Heads	0.46
9.	Total Not in Labor Force	0.34
10.	Not in Labor Force - Keeping House	0.35
11.	Not in Labor Force - Male, ages 20+	0.29
12.	Not in Labor Force - Nonwhite Male, ages 16-64 In School	0.19
13.	Not in Labor Force - Nonwhite Female, Ages 25-64	0.23
14.	Not in Labor Force - Nonwhite	0.30
15.	Agriculture Employed - Total	0.43
16.	Agriculture Employed - Male	0.44
17.	Agriculture Employed - Working 35+ Hours Per Week	0.43
18.	Agriculture Employed - Self-Employed	0.36
19.	Agriculture Employed - Wage and Salary	0.39
20.	Agriculture Employed - Female	
21.	Agriculture Employed - Working 1-34 Hours Per Week, Economic Reasons	0.06
22.	Agriculture Employed - Teenager (16-19)	0.24
23.	Agriculture Employed - Nonwhite	0.58
24.	Agriculture Employed - Unpaid Family Workers	0.44
25.	Agriculture Employed - Female, Ages 20+ Farmers and Farm Managers	0.11
26.	Agriculture Employed - Nonwhite, Self-Employed	0.03
27.	Civilian Labor Force - Total	0.34
28.	Nonagriculture Employed - Total	0.32
29.	Nonagriculture Employed - Wage and Salary	0.31
30.	Nonagriculture Employed - Worker 35+ Hours Per Week	0.26
31.	Nonagriculture Employed - Male	0.31
32.	Nonagriculture Employed - Female	0.28
33.	Nonagriculture Employed - Blue Collar	0.41
34.	Nonagriculture Employed - Wage and Salary Workers, Manufacturing	0.49
35.	Nonagriculture Employed - Male, Ages 20+ White Collar	0.50
36.	Nonagriculture Employed - Female, Ages 25+ Full-time, Worker 40 Hours Per Week	0.22

1/ All characteristics refer to the population 16+ years, unless otherwise indicated.

Table 5 continued

Characteristics	Correlation Coefficient
37. Nonagriculture Employed - Wage and Salary Workers, Retail Trade	0.29
38. Civilian Labor Force - Part-time, Ages 20-64	0.24
39. Nonagriculture Employed - Worked 1-34 Hours Per Week, Usually Full-time	0.04
40. Civilian Labor Force - Teenage (16-19)	0.19
41. Nonagriuculture Employed Teenage (16-19)	0.12
42. Nonagriculture Employed - Self-Employed	0.37
43. Nonagriculture Employed - Worked 1-14 Hours Per Week	0.11
44. Nonagriculture Employed - Wage and Salary Workers, Construction	0.26
45. Nonagriculture Employed - Employed, School is Main Activity, Ages 16-21	0.23
46. Nonagriculture Employed - Worked 1-34 Hours Per Week, Economic Reasons	0.11
47. Nonagriculture Employed - With a Job, Not at Work	0.03
48. Nonagriculture Employed - Worked 1-34 Hours Per Week, Usulaly Full-time, Economic Reasons	0.09
49. Nonagriculture Employed - Live on a Farm	0.28
50. Nonagriculture Employed - With a Job, Not at Work, Salary paid	-0.02
51. Nonagriculture Employed - Wage and Salary, Private Household Workers	0.27
52. Nonagriculture Employed - With a Job, Not at Work, On Vacation	0.00
53. Nonagriculture Employed - Female, Ages 20+, Drivers and Deliverymen	0.30
54. Nonagriculture Employed - Male, Ages 45-54, Wage and Salary, Private Household Workers	0.03
55. Civialian Labor Force - Nonwhite	0.30
56. Nonagriculture Employed - Nonwhite	0.30
57. Nonagriculture Employed - Nonwhite, Female	0.23
58. Nonagriculture Employed - Nonwhite, Blue Collar	0.29
59. Nonagriculture Employed - Nonwhite, Male, Full-time, Worked Equal to or Less than 40 Hours Per Week	0.27
60. Nonagriculture Employed - Nonwhite, Full-time, Worked 41+ Hours Per Week	0.16
61. Nonagriculture Employed - Nonwhite, Male, Ages 35-44	0.11
62. Nonagriculture Employed - Nonwhite, Male, Service Workers	0.30
63. Employed - Nonwhite, Female, Ages 45-54	0.38
64. Employed - Nonwhite, Worked 1-34 Hours Per Week, Economic Reasons	0.19
65. Nonagriculture Employed - Nonwhite, Worked 1-34 Hours Per Week, Economic Reasons	0.21

Table 5 continued

Characteristics		Correlation Coefficeint
66.	Nonagriculture Employed - Nonwhite, Employed as Managers, Officials, Proprietors	0.29
67.	Unemployed - Total	0.11
68.	Unemployed - Full-time Labor Force, Ages 20-64	0.05
69.	Unemployed - 5+ weeks	0.06
70.	Unemployed - Because Lost Last Job	0.13
71.	Unemployed - Male	0.13
72.	Unemployed - White Male	
73.	Unemployed - Household Heads	0.12
74.	Unemployed - Female	0.12
75.	Unemployed - White Female, Teenage (16-19)	-0.03
76.	Unemployed - White Female	0.13
77.	Unemployed - Had Been Wage and Salary, Manufacturing	0.16
78.	Unemployed - Teenage (16-19)	0.08
79.	Unemployed - Male Unemployed Less Than 5 Weeks	0.01
80.	Unemployed - Had Been Craftsmen and Kindred Workers	0.07
81.	Unemployed - Female, Full-time Labor Force, Reentered after 5 years	-0.09
82.	Unemployed - Because Left Last Job	-0.08
83.	Unemployed - Part-time Labor Force Ages 20-64	0.05
84.	Unemployed - No Previous Full-time Work	0.04
85.	Unemployed - White Female, Ages 20-64, Never Married	0.09
86.	Unemployed - Vietnam Era Veterans, Male, Ages 20-24	0.01
87.	Unemployed - Part-time Labor Force, Male Ages 20+	0.05
88.	Unemployed - Less Than 5 Weekds, Industry was Public Administration	-0.08
89.	Unemployed - Nonwhite	0.10
90.	Unemployed - Nonwhite, 5+ weeks	0.08
91.	Unemployed - Nonwhite, Because Lost Last Job	0.04
92.	Unemployed - Nonwhite, Male	0.04
93.	Unemployed - Nonwhite, Female	0.06
94.	Unemployed - Nonwhite, teenage, (16-19)	0.02
95.	Unemployed - Nonwhite, Female, 5-14 Weeks	0.04
96.	Unemployed - Nonwhite, 27+ Weeks	0.07
97.	Unemployed - Nonwhite, Teenage Female (16-19)	0.01
98.	Unemployed - Nonwhite, Teenage, Male (16-19)	0.02
99.	Unemployed - Nonwhite, Male, Ages 20-64, Divorced, Widowed or Separated	0.12
100.	Unemployed - Nonwhite, Because Left Last job	0.03

Source: Train, G., Cahoon, L., and Makens, P., "The Current Population Survey - Variances, Inter-Relations and Design Effects," presented at the annual meeting of the American Statistical Association, 1978.

Table 6. Estimated Correlation Coefficients
Between the 1975 & 1976 Annual Estimates of Victimizations
(NCS - National Sample)

Characteristics	Correlation Coefficient
Total crimes against persons	.512
Total crimes of violence	.363
Total rapes	.062
Total robberies	-.303
Total robb. & attempted robb. w/injury	-.021
Total robb. & attempted robb. w/injury from serious assault	.096
Total robb. & attempted robb. w/injury from minor assault	-.110
Total robb. & attempted robb. w/o injury	-.293
Total assaults	.394
Total aggravated assaults	.287
Total aggravated assaults w/injury	.373
Total attempted assaults w/weapon, aggravated	.114
Total simple assaults	.353
Total simple assaults w/injury	.193
Total attempted assaults w/o weapon, simple	.367
Total crimes of theft	.424
Total personal larcenies w/contact	-.071

Table 6. (continued)

Characteristics	Correlation Coefficient
Total purse snatching	-.020
Total pocket pickings	-.040
Total personal larcenies w/o contact	.410
Crimes of violence; White	.352
Robberies; White	-.223
Assaults, White	.330
Aggravated assaults; White	.192
Simple assaults; White	.271
Crimes of theft; White	.392
Crimes of violence; Black	.118
Robberies; Black	-.046
Assaults; Black	.041
Crimes of theft; Black	.317
Crimes of violence; male	.309
Robberies, male	-.081
Assaults; male	.040
Crimes of theft; male	.286
Crimes of violence; female	.151
Robberies; female	-.010
Assaults; female	.421
Crimes of theft; female	.274
Crimes of violence; age 12-19	.157
Robberies; age 12-19	.193
Assaults; age 12-19	.139
Crimes of theft; age 12-19	.305

Table 6 (continued)

Characteristics	Correlation Coefficient
Crimes of violence age 20-34	.189
Robberies; age 20-34	-.119
Assaults; age 20-34	.193
Crimes of theft; age 20-34	.136
Crimes of violence; age 65+	-.017
Robberies; age 35-64	-.211
Assaults; age 35-64	.095
Crimes of theft; age 35-64	.160
Crimes of violence; never married	.271
Robberies never married	.026
Assaults; never married	.156
Crimes of theft; never married	.296
Crimes of violence; married	.322
Robberies; married	.029
Assaults married	.305
Crimes of theft; married	.137
Crimes of violence; male; never married	.112
Robberies; male; never married	-.046
Assaults; male; never married	.067
Crimes of theft; male never married	.163
Crimes of violence; female; married	.110
Robberies; female; married	-.145
Assaults; female; married	.166
Crimes of violence; widowed	.463

Table 6 (continued)

Characteristics	Correlation Coefficient
Crimes of violence; family income ≤ $7,500	.244
Robberies; family income ≤ $7,500	.051
Assaults; family inc. ≤ $7,500	.233
Crimes of theft; family inc. ≤ $7,500	.543
Crimes of violence; family inc. $7,500-$14,999	.055
Robberies, family inc. $7,500-$14,999	-.138
Assaults; family inc. $7,500-$14,999	.192
Crimes of theft, family inc. $7.500-$14,999	.146
Crimes of violence; family inc. $15,000+	.160
Robberies; family inc. $15,000+	-.065
Assaults; family inc. $15,000+	.078
Crimes of theft; family inc. $15,000+	.310
Crimes of violence; family inc. < $15,000	.219
Robberies; family inc. < $15,000	-.195
Assaults; famuly inc. < $15,000	.301
Crimes of violence; separated or divorced	-.050
Crimes of violence; inside central cities in metropolitan areas	.168
Robberies; inside central cities in metropolitan areas	-.480
Assaults; inside central cities in metropolitan areas	.322
Crimes of theft; inside central cities in metropolitan areas	.440

Table 6 (continued)

Characteristics	Correlation Coefficient
Crimes of violence; outside central cities in metropolitan areas	.238
Robberies; outside central cities in metropolitan areas	-.095
Assaults; outside central cities in metropolitan areas	.186
Crimes of theft; outside central cities in metropolitan areas	.339
Crimes of violence; in nonmetropolitan areas	.488
Robberies; in nonmetropolitan areas	-.015
Assaults; in nonmetropolitan areas	.481
Crimes of theft; in nonmetropolitan areas	.600
Crimes of violence; in metropolitan areas	.329
Robberies; in metropolitan areas	-.350
Assaults; in metropolitan areas	.389
Crimes of theft; in metropolitan areas	.438
Total crimes against households-property	.392
Total burglaries	.128
Total burglaries - forcible entry	.023
Total burglaries - unlawful entry	.173
Total burglaries - attempted forcible entry	.223
Total household larcenies	.384
Total household larcenies - completed	.303

Table 6 (continued)

Characteristics	Correlation Coefficient
Total household larcenies - attempted	.099
Total household larcenies; ≤ $50	.210
Total household larcenies; $50+	.228
Total motor vehicle thefts	.057
Total motor vehicle thefts - completed	-.026
Total motor vehicle thefts - attempted	.097
Burglaries; White	.143
Household larcenies; White	.336
Motor vehicle thefts; White	-.029
Burglaries; Black	-.018
Household larcenies; Black	.272
Motor vehicle thefts; Black	.173
Burglaries; owned/bought tenure	.164
Household larcenies; owned/bought tenure	.237
Motor vehicle thefts; owned/bought tenure	-.009
Burglaries; rented tenure	-.039
Household larcenies; rented tenure	.037
Motor vehicle thefts; rented tenure	.093
Burglaries; age of head 12-34	.013
Household larcenies; age of head 12-34	.365
Motor vehicle thefts; age of head 12-34	-.038
Burglaries; age of head 35+	.304
Household larcenies; age of head 35+	.304
Motor vehicle thefts; age of head 35+	-.011

Table 6 (continued)

Characteristics	Correlation Coefficient
Burglaries; inside central cities in metropolitan areas	.317
Household larcenies; inside central cities in metropolitan areas	.418
Motor vehicle thefts; inside central cities in metropolitan areas	.022
Burglaries; outside central cities in metropolitan areas	.165
Household larcenies; outside central cities in metropolitan areas	.349
Motor vehicle thefts; outside central cities in metropolitan areas	.082
Burglaries in nonmetropolitan areas	.339
Household larcenies in metropolitan areas	.539
Motor vehicle thefts in nonmetropolitan areas	.145
Burglaries; family inc. ≤ \$7,500	.226
Household larcenies; family inc. ≤ \$7,500	.424
Motor vehicle thefts; family inc. ≤ \$7,500	-.002
Burglaries; family inc. ≤ \$15,000	.168
Household larcenies; family inc. ≤ \$15,000	.281
Motor vehicle thefts; family inc. ≤ \$15,000	.056
Burglaries; family inc. \$15,000+	.117
Household larcenies; family inc. \$15,000+	.184
Motor vehicle thefts; family inc. \$15,000+	-.029

Source: Memorandum from Assanua to Schooley, April 17, 1980, Bureau of the Census.

Table 7. The Relative Variance of the Composite Estimator of Annual Level

Assumed Correlation	Actual Correlation .10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.10	.997	.992	.987	.982	.977	.972	.967	.962	.957	.955
.20	1.000	.990	.979	.969	.959	.948	.938	.927	.917	.912
.30	1.009	.993	.976	.960	.943	.927	.911	.894	.878	.869
.40	1.028	1.004	.980	.956	.933	.909	.885	.861	.838	.826
.50	1.061	1.028	.995	.961	.928	.895	.862	.829	.796	.779
.60	1.120	1.074	1.028	.981	.935	.889	.843	.796	.750	.727
.70	1.234	1.167	1.100	1.033	.967	.900	.833	.767	.700	.666
.80	1.479	1.375	1.271	1.167	1.062	.958	.854	.750	.646	.594
.90	2.209	2.009	1.809	1.609	1.408	1.208	1.008	.807	.607	.507
.95	3.479	3.126	2.773	2.419	2.066	1.713	1.359	1.006	.653	.476

Table 8. The Relative Variance of the Composite Estimator of Annual Change

Assumed Correlation	Actual Correlation .10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.10	.997	.992	.986	.979	.971	.962	.952	.940	.926	.918
.20	.999	.989	.977	.964	.949	.932	.912	.889	.862	.847
.30	1.005	.989	.972	.953	.932	.907	.878	.845	.805	.783
.40	1.014	.994	.971	.946	.917	.885	.847	.803	.751	.722
.50	1.027	1.002	.974	.942	.906	.865	.818	.762	.697	.660
.60	1.046	1.015	.980	.941	.897	.847	.788	.720	.640	.594
.70	1.075	1.037	.994	.947	.892	.830	.758	.675	.576	.520
.80	1.124	1.076	1.023	.964	.896	.818	.729	.625	.502	.432
.90	1.224	1.162	1.093	1.015	.926	.825	.709	.573	.412	.320
.95	1.336	1.261	1.178	1.084	.978	.857	.717	.553	.361	.250

Attachment I

COMPOSITE ESTIMATORS

A. Composite Estimation in the CPS

1. The K Composite Estimator

"The Current Population Survey - Design and Methodology," Technical Paper 40 (1978). U.S. Dept of Commerce, Bureau of the Census.

The K composite estimator of level (mean or total of a certain labor force characteristic) in the current month h, y_h' is given by

$$y_h' = (1-K)\, y_h + K\,(y_{h-1}' + d_{h,\,h-1});\; 0 < K < 1,$$

where

$$d_{h,\,h-1} = y_h'' - y_{h-1}'',$$

y_{h-1}' is the composite estimator of level for the preceding month h-1,

y_h is the second-stage ratio estimator of level for the current month h,

y_h'' and y_{h-1}'' are the second-stage ratio estimators for months h and h-1 based on the six rotation groups that are in sample for both months h and h-1.

The variance of this estimator is given in Hansen, Hurwitz and Madow (1953), Sample Survey Methods and Theory Vol. II. John Wiley & Sons New York. This estimator with K = .5 was used previously in the CPS.

2. The AK Composite Estimator

Huang and Ernst (1981) "Comparison of an Alternative Estimator to the Current Composite Estimator in CPS," Proceedings of the American Statistical Association, Survey Research Methods Section, 303-308.

Under the 4-8-4 rotation pattern, let y_{hi} be the second stage ratio estimator of the level (mean or total) in the current month h based on the rotation group i (which is in its i-th time in the sample), i = 1, 2, ...8. Assume the following:

(1) $V(y_{hi}) = \sigma^2$ for all h, i

(2) Estimators derived from different rotation groups of a given month are uncorrelated; that is $Cov\ (y_{hi}, y_{hj}) = 0\ ,\ i \neq j = 1, 2, \ldots 8$

(3) Estimators derived from the overlapping rotation groups are covariance stationary.

The AK composite estimator is a generalization of the K composite estimator previously used in CPS and is given by

$$y'_h = (1/8)\{(1-K+A)\ (y_{h1} + y_{h5}) + (1-K-A/3) \\ (y_{h2} + y_{h3} + y_{h4} + y_{h6} + y_{h7} + y_{h8})\} \\ + K\ (y'_{h-1} + d_{h,h-1}), \quad 0 \le A \le 1,\ 0 \le K \le 1,$$

where $$d_{h,h-1} = (1/6)\ (y_{h2} + y_{h3} + y_{h4} + y_{h6} + y_{h7} + y_{h8} \\ -(y_{h-1,1} + y_{h-1,2} + y_{h-1,3} + y_{h-1,5} \\ + y_{h-1,6} + y_{h-1,7})).$$

Note that when A = 0, K = 0 the AK composite estimator reduces to the simple average of the estimates from the eight rotation groups. When A = 0 and K = .5 the AK composite estimator reduces to the K composite estimator previously used in the CPS.

Thus, the AK composite estimator has the potential of assigning more weight to the rotation groups which have been in the sample for the 1st and 5th time, and less weight to the rest of the rotation groups than the corresponding weights in the K composite estimator.

The variance of the AK composite estimator is given in Huang and Ernst (1981). The

AK composite estimator was found to be more efficient than the K composite estimator for monthly level, month to month change and annual average for both 4-8-4 and 3-9-3 rotation patterns. The AK composite estimator with A = .2 and K = .4 is currently being used in the CPS.

B. Minimum Variance Composite Estimators

1. Wolter (1979) "Composite Estimation in Finite Populations", Journal of the American Statistical Association, 74, 604-613.

This procedure is applicable to any rotation pattern. Illustrated here for first quarter estimates one year apart considering only two years' data.

For 1987 and 1986 1st quarters, the vector of simple estimators is

$$Y = (y_{t\ t}, y_{t\ t-1}, y_{t\ t-2}, y_{t-1\ t-1}, y_{t-1\ t-2}, y_{t-1\ t-3})',$$

where y_{ij} is the first quarter simple estimate (monthly average) for the i-th year from the j-th year panel, (t = 87, t-1 = 86. etc.).

Let $\beta = (x_t, x_{t-1})'$

be the vector of unknown parameters (true monthly averages for the quarter).

Assume the general linear model

$$Y = X\beta + e,$$

the error vector e satisfies

$$E(e) = 0,\ E(e e') = V,$$

V is the variance - covariance matrix of the vector of simple estimators and must be nonsingular. We have verified algebraically that V is nonsingular for annual estimates, and for all quarterly estimates considering sample overlap one year apart.

Then

$$\hat{\beta} = PY,$$

where

$P = (X' V^{-1} X)^{-1} X' V^{-1}$, V^{-1} is the inverse of matrix V,

and $Var(\hat{\beta}) = (X' V^{-1} X)^{-1} = C$ say.

The design matrix X and variance - covariance matrix V in this case are

$$X = \begin{vmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{vmatrix} \qquad V = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \rho & 0 & 0 \\ 0 & 0 & 1 & 0 & \rho & 0 \\ 0 & \rho & 0 & 1 & 0 & 0 \\ 0 & 0 & \rho & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

Where $\sigma^2 = 1$ (assumed) and ρ is the correlation coefficient.

The matrix P is composed of the coefficients that are applied to the simple estimators in forming the minimum variance linear unbiased (MVLU) estimators.

Thus, $\hat{X}_t = P_1Y$, $\hat{X}_{t-1} = P_2Y$,

where P_i is the i-th row of P.

The estimator of change $X_t - X_{t-1}$ is simply $\hat{X}_t - \hat{X}_{t-1} = P_1Y - P_2Y$.

$$Var(\hat{X}_t - \hat{X}_{t-1}) = C_{11} + C_{22} - 2C_{12},$$

where C_{ij} is the ij element of C matrix.

It is important to note that this estimator uses all available data up to and including the year t. When data from succeeding years 1988, 1989 etc. are available, estimates for 1987, 1986 etc. also change.

Advantage: Applicable to any rotation pattern and data structure. It requires only specifications for design and variance-covariance matrices X and V.

Disadvantages:

(1) Requires storage of simple estimates for all previous years.

(2) Composite estimates for earlier years are changed.

(3) Becomes a computational burden with large number of survey characteristics as number of years increases because it requires inversion of matrices V and $(X' V^{-1} X)$.

2. Cochran (1963) Sampling Techniques, John Wiley & Sons, New York.

Assume sample size on each occasion is n. On occasion h, the proportion of the sample matched with occasion (h-1) is λ . The variance σ^2 and the correlation ρ between item values on the same units on two successive occasions are assumed same throughout. A part of the sample on occasion h may also be matched with the sample on occasion (h-2), but this is ignored. The two estimators of mean $\bar{Y}_h$ on the h-th occasion are $\bar{y}_{hu}$ from the unmatched proportion of the sample, and $\bar{y}'_{hm} = \bar{y}_{hm} + b\,(\bar{y}'_{h-1} - \bar{y}_{h-1,m})$ from the matched portion of the sample where b is the regression coefficient and $\bar{Y}'_{h-1}$ is the composite estimator on occasion h-1.

Then the composite estimator for the h-th occasion is

$$\bar{y}'_h = K\,\bar{y}_{hu} + (1-K)\,\bar{y}'_{hm},$$

and its variance is

$$V(\bar{y}'_h) = \frac{\sigma^2}{n}\,g_h,$$

where

$$g_h = \frac{K^2}{1-\lambda} + \frac{(1-K)^2(1-\rho^2)}{\lambda} + \rho^2(1-K)^2\,g_{h-1}.$$

The asymptotic value of $V(\bar{y}'_h)$ as $h \to \infty$ is

$$V(\bar{y}'_\infty) = \frac{\sigma^2}{n}\,g_\infty ,$$

where

$$g_\infty = \frac{\lambda K^2 + (1-\lambda)(1-K)^2(1-\rho^2)}{\lambda(1-\lambda)[1-\rho^2(1-K)^2]} .$$

The value of K which minimizes the limiting variance $V(\bar{y}'_\infty)$ is

$$K_{opt} = \frac{\sqrt{(1-\rho^2)}\,\left[\sqrt{(1-\rho^2) + 4\,\lambda(1-\lambda)\rho^2} - \sqrt{(1-\rho^2)}\right]}{2\lambda\rho^2} .$$

The estimator of change $\bar{Y}_h - \bar{Y}_{h-1}$ is simply $\bar{y}'_h - \bar{y}'_{h-1}$.

And its variance is

$$V(\bar{y}'_h - \bar{y}'_{h-1}) = \frac{\sigma^2}{n}[g_h + g_{h-1}\{1-2\rho(1-K)\}].$$

Advantages: (1) Requires storage of the composite estimate for the previous year only.

(2) Composite estimates for earlier years are not changed.

(3) It is also the minimum variance estimator.

(4) Computationally much simpler than the Wolter Estimator.

3. Ernst-Breau (1983). Unpublished Manuscript, Bureau of the Census, Washington, D.C.

This composite estimator was developed specifically for the CPS Supplements that are repeated every year like the March Supplement and thus assumed 50% overlap between samples one year apart and no overlap between samples two years apart.

The estimator of the current yearly level is

$$Y_1 = (1-A_1)X_{1,1} + A_1X_{1,2} - A_2X_{2,1} + A_2X_{2,2} - A_3X_{3,1}$$

$$+ A_3X_{3,2} - \ldots\ldots\ldots\ldots - A_nX_{n,1} + A_nX_{n,2},$$

where $X_{1,1}$ is the simple estimate of yearly level for the current year based on the 50% of the sample that is not matched with the previous year,

$X_{1,2}$ is the simple estimate of yearly level for the current year based on the 50% of the sample that is matched with the previous year, and

$X_{2,1}$ is the simple estimate of yearly level for the previous year based on the 50% of the sample that is matched with the current year, etc.

Assume $V(X_{i,j}) = \sigma^2$, for all i,j,

$Cov\ (X_{i,2}, X_{i+1,1}) = \rho\sigma^2$, for all i,

and all other covariances are zero.

Then coefficients A_i's are given by

$$A_i = C_1 Z_1^{i-1} + C_2 Z_2^{i-1},$$

where $Z_1 = [1 - \sqrt{(1-\rho^2)}]/\rho,$

$$Z_2 = [1 + \sqrt{(1-\rho^2)}]/\rho,$$

$$C_1 = (2 - \rho Z_1)/[(2 - \rho Z_1)^2 - Z_1^{2n-2} (2 - \rho Z_2)^2],$$

$$C_2 = -Z_1^{2n-2} (2 - \rho Z_2)/[(2 - \rho Z_1)^2 - Z_1^{2n-2}(2 - \rho Z_2)^2].$$

$$\lim_{n \to \infty} c_1 = 1 / (2 - \rho Z_1),$$

$$\lim_{n \to \infty} C_2 = 0,$$

Therefore, $\lim_{n \to \infty} Ai = Z_1^{i-1} / (2 - \rho Z_1).$

Finally, the limiting variance (algebraic details omitted) is

$$V(Y_1) = \sigma^2 \{1 - \frac{2}{(2-\rho Z_1)} + \frac{2(1-\rho Z_1)}{(2-\rho Z_1)^2(1-Z_1^2)}\}.$$

Advantages:

(1) Requires storage of the composite estimate for the previous year only.

(2) Composite estimates for earlier years are not changed.

(3) It is also the minimum variance estimator.

(4) Computationally much simpler than the Wolter Estimator, and slightly simpler than the Cochran Estimator.

Attachment II.

Data Structure for Annual Estimates

Annual Estimate

1984 Estimate

1984 Panel

Wave	Rotation
2	1 2 3
3	4 1 2 3
4	4 1 2 3
5	4 1 2 3

1985 Panel

Wave	Rot	Data for
1	2	OND 1/
	3	ND
	4	D

(Insufficient) 2/

1985 Estimate

1984 Panel

Wave	Rotation
5	1 2 3
6	4 1 2 3
7	4 1 2 3
8	4 1 2 3

1985 Panel

Wave	Rot
1	2 3 4 1
2	2 3 4
3	1 2 3 4
4	1 2 3 4

1986 Panel

Wave	Rot	Data for
1	2	OND 1/
	3	ND
	4	D

(Insufficient) 2/

1986 Estimate

1984 Panel

Wave	Rotation		Data For
8	1		J 1/
	2		JF
	3		JFM
9	4		JFMA
		1	FMAM
		2	MAMJ
		3	AMJJ

(Insufficient) 2/

1985 Panel

Wave	Rot
4	2 3 4
5	1 2 3 4
6	1 2 3 4
7	1 2 3 4

1986 Panel

Wave	Rot
1	2 3 4 1
3	2 3 4 1
3	2 3 4
4	1 2 3 4

1987 Panel

Wave	Rot	Data for
1	2	OND 1/
	3	ND
	4	D

(Insufficient) 2/

1/ Letters J, F, ... D stand for months January, February ... December.

2/ This panel does not provide an annual estimate for the year e.g., 1985 Panel provides data only for the fourth quarter of 1984.

Attachment II (cont.)

Data Structure for Annual Estimates

1987 Estimate

1985 Panel

Wave	Rot	Data for
7	2	J [1]
	3	JF
	4	JFM
8	1	JFMA
	2	FMAM
	3	MAMJ
	4	AMJJ

(Insufficient) [2]

1986 Panel

Wave	Rot
4	2 3 4
5	1 2 3 4
6	1 2 3 4
7	1 2 3 4

1987 Panel

Wave	Rot
1	2 3 4 1
2	2 3 4
3	1 2 3 4
4	1 2 3 4

1988 Panel

Wave	Rot	Data for
1	2	OND [1]
	3	ND
	4	D

(Insufficient) [2]

1988 Estimate

1986 Panel

Wave	Rot	Data for
7	2	J [1]
	3	JF
	4	JFM
8	1	JFMA
	2	FMAM
	3	MAMJ

(Insufficient) [2]

1987 Panel

Wave	Rot
4	2 3 4
5	1 2 3 4
6	1 2 3 4
7	1 2 3 4

1988 Panel

Wave	Rot
1	2 3 4 1
2	2 3 4
3	1 2 3 4
4	1 2 3 4

1989 Panel

Wave	Rot	Data for
1	2	OND [1]
	3	ND
	4	D

(Insufficient) [2]

[1] Letters J, F, ...D stand for months January, February, ...December.

[2] This panel does not provide an annual estimate for the year e.g., 1985 Panel provides data only for the two quarters of 1987.