VARIANCES OF THE CURRENT POPULATION SURVEY, INCLUDING WITHIN- AND BETWEEN-PSU COMPONENTS AND THE EFFECT OF THE DIFFERENT STAGES OF ESTIMATION

Martha J. Banks and Gary M. Shapiro, U.S. Bureau of the Census

I. Introduction

The important results in this paper are given in the tables in Section IV. The tables contain various results from the Census Bureau variance program used for the Current Population Survey (CPS) that are felt to be useful in planning for and estimating variances for other household surveys.

Section II gives a summary of the sample design and estimation procedures of CPS which are useful in an understanding of the results of this paper. These features, as well as other aspects of the sample not covered here, are described more completely elsewhere/87. This section can be skipped by persons already fully familiar with the survey. Section III describes the variance program now in use.

II. Description of the Current Population Survey (CPS)

A. Nature of the CPS. CPS is a sample survey conducted monthly for the Bureau of Labor Statistics by the Bureau of the Census to obtain national estimates of employment, unemployment, and other characteristics of the labor force. Because of the sample design, the survey can be and is used to produce estimates for a wide variety of other demographic characteristics for the population as a whole, as well as for various subgroups of the population.

B. Sample Design. Since January 1967, the monthly CPS sample has consisted of about 48,000 eligible households in 449 first-stage sampling units (primary sampling units, or PSU's) comprising 863 counties and independent cities. Beginning in July 1969, the overall sampling rate for the survey has been 1 in 1240. The 449 PSU's were selected out of 357 strata. Of the 357 strata, 112 consist of only 1 PSU. Such PSU's are necessarily in sample and are called self-representing (SR). The sampling rate within each of these PSU's is 1 in 1240. The other 245 strata contain more than one PSU each, and the sample PSU's from these strata are called non-self-representing (NSR) since a sample PSU from one of these strata also represents the other PSU's in the same stratum.

The 245 strata were grouped into 122 pairs of strata with one stratum left over. From each pair of strata, one stratum was picked at random, each with equal probability. From each selected stratum one PSU was chosen with probability proportionate to the 1960 populations of the PSU's. The sample size to be taken from the chosen PSU was determined such that the effective sampling rate within the stratum was 1 in 1860 (i.e., 3/2 x 1240).

From each of the remaining 122 strata not selected, two PSU's per stratum were independently chosen with probability proportionate to size. Since the choices were independent, it was possible for the sample PSU's to be either the same or different. In the thirty strata where the two choices were the same, this procedure simply results in twice as large a sample within each twice-chosen PSU as would have been the case if the PSU were chosen only once. The one unpaired stratum left over was handled similarly.

C. Selection of the Sample Within the Sample PSU's. The object of subsampling within each of the sample PSU's was to obtain a self-weighting probability sample of housing units and units in special places. The housing units were selected in segments containing, on the average, about six housing units.

The selection of a sample of segments within each PSU proceeded in several stages as follows:

1. Selection of a sample of the enumeration districts used in the 1960 Census of Population and Housing. The enumeration districts used in the Census were geographic areas, usually with well-defined boundaries and containing, on the average, about 250 dwelling units.
2. Subdivision of each enumeration district into segments.
3. Selection of a sample of segments in each of the selected enumeration districts.

There were two types of sampling used within enumeration districts. List sampling of units enumerated in 1960 was used primarily in urban areas and area sampling was used primarily in rural areas.

A subsample of building permits from a sample of areas where such permits were required and available was used for most of the newly constructed units. Where such permits were either not required or not available, the newly constructed units were picked up in area-sample enumeration districts. In list-sample enumeration districts, however, the newly constructed units were picked up by a successor check, which is described in /87.

D. Rotation of the Sample. The rotation system used in the CPS may be described as follows:

1. The entire sample is divided into eight equal, separate, systematic subsamples, referred to as rotation groups. One new rotation group is introducted into the survey each month, and one old one is replaced.
2. Each new rotation group is included in the survey for four months, then is excluded for eight months, then is returned for an additional four months. The chart presented in exhibit A below, shows in a simplified form, how the rotation system operates. Examination of the chart will make clear the important characteristic that in any month, six of the eight rotation groups in sample will have been in the survey for the previous month, i.e., there will always be a 75 percent month-to-month overlap. Also, half of the rotation groups in any month will have been in the survey exactly a year before.

EXHIBIT A CPS ROTATION CHART

1971 March	1	2	3	4			5	6	7	8	
April		2	3	4	5			6	7	8	1
May			3	4	5	6			7	8	1 2
:											
:											
1972 March							5	6	7	8	1 2 3 4

E. Estimation Procedure. To arrive at a final estimate, adjustment for nonresponses is made, two stages of ratio estimation are applied, and the preceding month's data is utilized to form a composite estimate.

1. Adjustment for Nonresponse.-For all units except large special places, strata are combined into 76 groups containing from one to nine strata each. The ratio of the designated sample to the interviewed sample within each of six race-residence categories is used as the adjustment for noninterview.

The basic weight of 1240 (which is the inverse of the overall sampling fraction) for an individual record is then multiplied by the appropriate factor. The adjustment for nonresponse is a little different in the large special places, but the effect is about the same.

2. First Stage Ratio Estimate Adjustment.-The purpose of the first-stage ratio estimate is to reduce the contribution to the variance arising from the sampling of PSU's, i.e., the variance that would still be associated with the estimates if all households in every sample PSU were included in the survey each month.

The first-stage ratios are based on 1960 Census data and are applied only to sample data for the NSR PSU's. For the NSR PSU's in each of the four Census regions, a ratio is computed for each of six race-residence groups as follows:

$$\frac{\text{1960 Census Pop. in Race-Residence Group for NSR Strata in Region}}{\text{Estimates of this Pop. Based on 1960 Census Pop. for Sample PSU's}}$$

The basic weight for all records from NSR strata is then multiplied by the appropriate ratio of this type, on top of the previously applied noninterview factor.

3. Second-Stage Ratio Estimate Adjustment.-The second-stage ratio estimate factor adjusts sample estimates of the U.S. population in a number of age-sex-race groups to independently derived current estimates of the population in each of these groups. These independent estimates are prepared each month by carrying forward the most recent Census data to take account of subsequent aging of the population, mortality, and migration between the U.S. and other countries. The CPS sample returns, after application of the noninterview adjustment and first-stage ratios are, in effect, used only to determine the percentage distribution of the population within each age-sex-race group by employment status and various other characteristics.

4. Composite Estimate.-The last stage in the preparation of estimates is the derivation of a composite estimate. The composite estimate is a weighted average of two estimates for the current month for any particular item. The first of these two estimates is the result of the two stages of ratio estimates described above. The second estimate consists of the composite estimate for the preceding month to which has been added an estimate of the change from the preceding month to the present month, where the estimate of change is based on the six rotation groups common to the two months (about 75 percent of the households in sample in the current month).

For such a composite estimate to be unbiased, the weights for the two components must add to 1. In CPS, the weights used are each $\frac{1}{2}$.

III. Variance Estimation in the CPS

A. Background. In the past decade, the Census Bureau has used several methods of variance computation for the CPS data. The two most reliable of these have been a replication method and a paired difference method based on work by Keyfitz. [2]

The CPS replication variance program used 20 replications. Although some attempt at balancing to improve the reliability of the variance estimates was made, the more effective balancing procedures developed by McCarthy [3] were not used. [8] contains a description of the replication method as used by the Census Bureau.

The Keyfitz method estimates variances analytically. Basically, it ". . . amounts to calculating a linear combination of sample totals for each primary sampling unit, and then estimating the variance of the sum of those linear combinations." [5] Since Keyfitz's original article, Tepping [5] has given a more eloquent and general formulation of the variance method. [4], [6] and [7] contain detailed descriptions of the Keyfitz method as used by the Census Bureau.

B. Basic Theory. In the Keyfitz formulation, the basic idea is that

$E(x_1 - x_2)^2 = VAR(x_1 + x_2)$ provided that $E(x_1) = E(x_2)$.

Key theorems in [4] give simple expressions for the relvariance of a ratio and of a sum of ratios, e.g.,

$$V^2_{\frac{x}{y}} = \sum_s \left(\frac{x_{s1} - x_{s2}}{E(x)} - \frac{y_{s1} - y_{s2}}{E(y)} \right)^2$$

where x_{si} is an estimate for the i^{th} half of the s^{th} "Keyfitz cluster." (The meaning of "Keyfitz cluster" is clarified below.)

In the Tepping formulation the Taylor series approximation, to terms of the first degree, is written out for the estimate of interest. The variance of the Taylor approximation is then computed directly.

C. Treatment of SR PSU's. Variances are computed differently for SR and NSR PSU's. The SR PSU's are collapsed into 18 clusters. Subcluster 1 of a Keyfitz cluster (the x_{1s} in the above formula is an estimate for sub-cluster 1 of the s^{th} Keyfitz cluster) consists of four of the eight rotation groups, and subcluster 2 consists of the other four. The Keyfitz method is applied four different times for different combinations of the rotation groups. The four resulting variance estimates are then averaged, giving a more reliable result than if only one combination had been used.

D. Treatment of NSR PSU's. For NSR PSU's two techniques are applied. One technique is the one described above, with each subcluster consisting of four rotation groups. This technique, applied to SR PSU's, gives an estimate of total SR variance, but applied to NSR PSU's, it gives an estimate of the NSR within-PSU variance only.

For estimating the total NSR variance, there are 123 Keyfitz clusters, each consisting of one of the pairs of strata as described in II.B. For each pair of strata, there are three sample PSU's,

two from one stratum (denoted by A1 and A2) and one from the second stratum (denoted by B). For variance purposes, A1 and B can be thought of as representing only the PSU's in their respective strata; while A2 can be thought of as representing all PSU's in both strata. Thus, for all PSU's, there is a between-PSU-within-stratum variance component, but for only 1/3 of the PSU's there is also a between-stratum component. In order to reflect this in an unbiased fashion, a weighted average of two variance estimates is formed [6]. The first is of the form $(A_1 - A_2)^2$, and includes an unbiased estimate of the between-PSU variance but no between-stratum variance.

The second is of the form $\left(\frac{A_1 + A_2}{2} - B\right)^2$, and includes a between-stratum variance as well. Weights of 7/12 and 1/9, respectively, were derived for these two terms to produce the desired unbiased variance estimate.

E. Census Computer Program. The computer program is written to estimate variances for 45 simple totals (such as total unemployed persons). (The figure of 45 is an upper limit determined by computer storage space considerations.) However, the program can also compute a limited number (10) of covariances, so that with a little arithmetic, variances of simple ratios (such as the unemployment rate) can also be estimated. Estimates of within-PSU variance, between-PSU variance, and between-stratum variance are computed as well as total variance. Separate variance estimates are also produced for each of the several estimators for the unbiased estimate (includes noninterview adjustment), the first-stage ratio estimate, the second-stage ratio estimate, the first- and second-stage combined ratio estimate, and the composite estimate. Further, variances of both monthly level and month-to-month change are produced. (The latter is produced only for the first-and second-stage combined ratio estimate and for the composite estimate.) Also, for the unbiased estimate and the first-stage ratio estimate, the variances for estimates for SR PSU's only and for NSR PSU's only are given.

IV. Data from Keyfitz Variance Program

A. Introduction. Now we get to the heart of the paper: Presentation of tables. All of the tables contain actual data. Nothing completely new and unexpected is presented. Most tables substantiate theoretical work for which there previously has been little or no empirical verification.

It should be noted that the number of digits shown in the tables are not an indication of the reliability of the estimates. In general, the last digit or two are of doubtful significance, but the figures were left unrounded so that the reader can manipulate them as desired before rounding.

B. Tables 1 and 2. Both Tables 1 and 2 record the design effects (Deff's)[1] for each of the items in the variance program for unbiased estimates and the first- and second-stage combined ratio estimate (i.e., noncomposite estimate). The figures are ratios of the actual monthly CPS variances (using an annual average of the monthly data) divided by the variances appropriate for a simple random sample of persons and an unbiased estimate $\left(\frac{pq}{n}\right)$. Here, p represents the proportion computed from the sums of the twelve monthly CPS estimates of totals. In Table 2, for a characteristic like "Unemployed males of Negro and other races," the denominator of p is males of Negro and other races, 16 and over (14 and over for 1965 and 1966). For the same characteristic (and all other characteristics) in Table 1, the denominator of p is Total Persons, 16 and over. Only those characteristics which are subsets of age-sex-race groups for which independent control totals are used in the ratio estimation are included in Table 2.

In making comparisons between Deff's in Tables 1 and 2, there are small increases in Table 2 for relatively rare characteristics like "Males who are agriculture employed", but rather large increases in Table 2 for the more frequent characteristics like "Females in civilian labor force". These more frequent characteristics are the ones most helped by the second-stage ratio estimate factors, as indicated by the unusually small Deff's for the first- and second-stage combined ratio estimate in Table 1. In contrast, the Deff's in Table 2 for these more frequent characteristics are more in line with the Deff's for other types of characteristics in Table 1.

Table 2 Deff's are primarily applicable under two circumstances. One circumstance is if one is drawing a sample from a universe consisting of only a restricted age-sex-race group (e.g., males 16-19). The second is if you are actually interested in the percentage of persons in a restricted age-sex-race group. (In this case, however, only the Deff's for the first and second-stage combined estimate are applicable.)

In comparing the unbiased estimate and ratio estimate, note that without a single exception, the ratio estimate reduces the variance. In general, as would be expected, the characteristics possessed by a relatively large percentage of the population are helped most by the ratio estimate. The effect of the ratio estimation is rather dramatic for these "large" characteristics. For the unbiased estimate, these "large" characteristics have among the largest design effects, while for the ratio estimate they have the lowest Deff's. The one characteristic possessed by a small percent of the population that is very much helped is "Employed persons of Negro and other races", but this characteristic constitutes a large percentage of total persons of Negro and other races in certain age-sex categories for which independent controls are used.

For the unbiased estimate, only rural population and agriculture employed characteristics have Deff's as high as the "large" characteristics. The Deff's for these rural items are significantly reduced by the ratio estimate, but they still remain relatively high. This obviously reflects the highly clustered nature of these populations.

1/ Within the Census Bureau, the term "Factor over Random" is used instead of design effects, since it is the factor that expresses the amount of variance over and above simple random sampling variance.

Since 1967 is the year that the present sample design was instituted, one other thing to look for in Table 1 is the difference between the Deff's for 1967 through 1969, and those for 1965 and 1966. The only really significant differences occur for agriculture employed and rural farm items and for the items "At school" and "Self-employed". For all of these items, the Deff's are smaller for the recent years for the unbiased estimate, but interestingly enough, not for the ratio estimate. (In fact, for "At school", the Deff's are actually larger for the ratio estimate for recent years.)

An important change which was made in the method of estimating variances for 1967 is almost certainly responsible for the differences for the agriculture employed and rural farm items. Beginning in 1967, the program provided an unbiased estimate of variance, whereas previously the program used a collapsed-stratum estimate which is upward-biased because it contained a between-stratum component. A change in the sample design in 1967 permitted this improvement. Table 3 can be used to estimate the magnitude of this bias (see Section IV.C). Its elimination is probably the reason for the large reductions among agriculture employed and rural farm characteristics, since these characteristics have very large between-PSU components of the total variance.

This change is also undoubtedly partially responsible for the difference for the "Self-employed" item, since about 25 percent of the self-employed are in agriculture.

A change in the population base used for labor force data is probably responsible for the difference for the "At school" item. The tabulation change is that all characteristics after 1966 are tabulated for the civilian noninstitutional population 16 and over; whereas in the previous years, 14 and 15 year olds were also included in the tabulations. Obviously, this cange can be expected to have important effects for this item, while at the same time affecting other items only slightly.

C. Table 3. Table 3 gives the ratio of the total between-variance (sum of the between-PSU and "between-stratum variance") to the total variance, and the ratio of the between-stratum variance to the total variance. See Section III. D. for the meaning of "between-stratum variance."

Since both of these between-variance estimates are derived by the subtraction of one variance estimate from another, they are not very reliable. For this reason the ratios of between to total variance vary so from year to year that even the 3-year average shown in the table is only a crude measurement device, as the negative ratios indicate. It can be clearly seen, however, that the overwhelming component of variance is within-PSU variance rather than between-PSU or between-stratum variance.

Items involving agriculture employed and those employed as wage and salary workers have relatively high total between variance; while for unemployment items, it is relatively low. The average ratio of total between variance to total variance is .114 for agriculture employed, .121 for wage and salary workers, and .011 for the unemployed.

It is rather surprising to see the results for the rural population. One would expect these items to have high positive ratios rather than the largest of all the negative ratios. Although the ratios for all three years are negative, only the ratio for one year is a large negative.

When a collapsed-stratum variance estimation procedure is used for a sample design where there is only one sample PSU in each stratum, the variance estimate is an overestimate due to the inclusion of a "between-stratum variance." The data in Table 3 can be used to estimate the relative magnitude of this bias in the variance estimate for the CPS design that was in effect prior to 1967. In that design the strata were the same as the present strata, but there was only one sample PSU per stratum.

The magnitudes of the between-PSU-within-stratum variance and the "between-stratum variance" (as calculated by a collapsed-stratum procedure) for the previous design are approximated by multiplying the corresponding estimates from Table 3 by 1.5 and 9.0 respectively.

Thus, in the former design, for the "Average of all items", the ratio of between-PSU variance to total variance can be estimated as .159 $/.159 = (1.5)(.051 - .011) + 9(.011)/$, with the (.099) term representing the bias of the between-PSU variance estimate. Note that in this case, the bias represents about 60 percent of the between-PSU variance estimate.

D. Table 4. The composite estimate is the weighted average of two estimates for the same characteristic, as explained in Section II.E.4. Table 4 presents the ratio of the variance of the composite estimate to the variance of the first- and second-stage combined ratio estimate (i.e., the noncomposite estimate) for both estimates of monthly level and of monthly change.

In general, the composite estimate reduces the variances somewhat, but this is not always the case. For unemployment items, "Part-time usually full time" items, and for the item "With a job, not at work", the use of the composite estimate increases the variance.

A composite estimate that weighted the two component estimates differently would be better for these items. In 1963, Gurney /1/ determined the optimal weights for various items. She estimated, for example, that for unemployment data, weighting the noncomposite estimate by .7 and the other estimate by .3 would result in a variance lower than either the present composite estimate or the noncomposite estimate. In considering the estimation procedure for CPS, it was decided to use a single pair of weights that would be reasonably good for all items, rather than different weights for different items, in order to minimize complications.

For each characteristic, the use of the composite estimate helps the estimate of monthly change more (or hurts it less) than it does the estimate of monthly level. This is as expected, since the composite estimate makes use of the previous month's estimate. On the average, the variance on the estimate of monthly level is reduced by 4 percent, while the variance on monthly change is reduced by 19 percent.

E. Table 5. Ofentimes, in order to increase reliability, several months' worth of data are accumulated and averaged. The first four columns of Table 5 present the reduction in variance obtained by using 3, 6, 9 or 12 consecutive months' worth of data rather than a single month's data. It is also frequently necessary to estimate the variance of the estimated difference between two months' data. The ratios presented in the last two columns are of the form:

$\frac{VAR(x - y)}{\frac{1}{2}(VAR(x) + VAR(y))}$. For the first of these columns, x and y represent estimates for adjacent months; while for the last column, they represent estimates for two months a year apart. The data in this table results from a special computer program, as well as the regular Keyfitz program.

It should be remembered that for CPS, each rotation group is in sample four months, excluded eight months, and then returned for four (see Exhibit A in Section II.D). This results in a 75 percent monthly overlap of rotation groups and a 50 percent overlap of rotation groups for year-apart data. The figures in the table would, of course, be different with a different rotation scheme.

Looking at the "average" part of the table (the first four columns) there are wide differences among items. Items for which the correlation over time is low, such as unemployment items, are helped considerably by multiple-month data; while items for which the correlation over time is high, such as agriculture employed and rural items, are helped only a little. The CPS sample design is such that when a segment drops out of sample, it is replaced by another segment which is in close geographic proximity. Thus, there is a correlation not only between identical rotation groups but between non-identical rotation groups. For example, if all the people in a segment are rural farm, then all the people in the replacement segment are also likely to be rural farm. Because of the rotation scheme, a hypothetical item with perfect correlation between identical rotation groups ($\rho_s = 1.00$) and no correlation between different rotation groups ($\rho_s = 0.00$) would cause the 3-month variance to be reduced to .78 of the monthly variance. (See appendix for the calculation of .78) Since, for example, the agriculture-employed figure is .83, this means that the correlation between non-identical rotation groups must be at least .23, as can be calculated from formula 5 in the appendix.

The fifth column of the table gives a ratio of the variance for a difference of two adjacent months to the variance of monthly level. Note that for a hypothetical item with perfect correlation between identical rotation groups and no correlation between non-identicals, the ratio would be .50. If there were no correlation at all, the ratio would be 2.00. The last column gives a similar ratio for two months a year apart, instead of two adjacent months. For this ratio, perfect correlation between identicals and no correlation between non-identicals would yield a ratio of 1.00, and no correlation at all would yield a ratio of 2.00.

For the "Average of all items", the ratio for adjacent months is about 45 percent larger than the ratio for data a year apart. This is as expected, since not only are there fewer identical rotation groups for the latter, but the passage of time usually reduces the correlation both between identicals and between non-identicals. There seems to be one major exception to this in the table - a person's likelihood of being on vacation. Thus, "With a job, not a work" is the only item that does not have a higher ratio for data a year apart than for adjacent months.

As an example of how to use these columns, a good approximation to the variance between adjacent months' unemployment levels can be obtained from a single month's variance by multiplying by 1.47. Use a factor of 1.56 to estimate the variance of yearly differences in teenage employment.

G. Table 6. The Keyfitz and replication methods of estimating variances were referred to in Section III. The sixth table uses 1964 data to compare the relvariances (variances divided by the squares of the estimates) calculated by the replication method and by the Keyfitz method. This is shown for each of the items the two computer programs have in common. Variances for both the first- and second-stage combined ratio estimate and the composite estimate are compared.

The two methods appear to give consistent results, as can be seen from columns 3 and 8. However, the replication relvariances vary more from month to month than do the relvariances computed by the Keyfitz method. Columns 4, 5, 9 and 10 give one-sixth of the range of the monthly figures. This approximation to the standard error of the relvariance estimates shows, as expected, that the Keyfitz method provides much more reliable estimates than does the replication method (20 replications). Had the McCarthy method /3/ of choosing the half-sample been used, the replication method would compare more favorably.

Prior to 1968, there was a minor error in the computation of the composite estimate. This has been taken into account by a slight adjustment in columns 6 and 9.

V. Acknowledgements

The authors wish to thank Robert H. Hanson and Joseph Waksberg for their many valuable comments on the preparation of this paper.

TABLE 1 Design Effects for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Characteristic	Percent of Population possessing characteristic					Design Effects [1]									
	Population 14+		Population 16+			Unbiased Estimate					1st & 2nd Stage Combined Est.				
	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969
Civilian Labor Force															
Total	57	57	60	60	60	7.65	9.29	9.49	9.87	10.09	1.16	1.06	.97	1.03	1.15
Females	20	21	22	22	23	2.54	2.60	2.63	2.72	2.79	1.12	1.07	.94	.93	1.01
Under 20 yr.	5.4	5.9	5.1	5.1	5.2	1.71	1.90	1.66	1.75	1.97	.82	.79	.62	.67	.69
Employed															
Negro and Other Races	5.8	5.9	6.2	5.8	6.2	6.03	6.30	5.25	5.90	6.23	.49	.54	.48	.53	.59
Males	35	35	36	36	36	3.50	4.48	4.19	4.23	4.23	.32	.32	.28	.28	.31
Working 1-14 hours	3.4	3.4	3.0	3.1	3.0	1.73	1.65	1.41	1.62	1.58	1.45	1.31	1.21	1.39	1.29
Part-Time (1-34 hr.), usually full-time	3.4	3.0	3.6	4.5	4.1	1.97	1.75	1.64	1.85	2.10	1.69	1.51	1.40	1.53	1.78
Part-Time (1-34 hr.), usually full-time, part-time for economic reasons	.85	.72	.92	.78	.77	1.81	1.45	1.42	1.90	1.49	1.63	1.32	1.33	1.77	1.39
Self-employed	6.4	6.1	5.6	5.4	5.3	2.51	2.38	1.85	1.83	1.86	1.46	1.38	1.38	1.45	1.41
Nonag. Employed															
Total	51	52	54	55	55	6.63	8.25	8.45	8.81	9.15	1.27	1.18	1.16	1.17	1.23
Male	32	32	34	34	34	3.44	4.39	4.26	4.27	4.33	.48	.48	.45	.43	.44
Female	18	19	21	21	21	2.35	2.48	2.50	2.62	2.69	1.11	1.10	.95	.97	1.01
Working 35 hr.+	39	39	41	40	40	4.54	5.65	6.00	5.66	5.62	1.07	1.10	1.13	1.12	1.17
With a job, not at work	2.5	2.6	2.9	3.1	3.2	1.44	1.60	1.58	1.66	1.75	1.25	1.32	1.31	1.36	1.44
Ag. Employed															
Total	3.5	3.1	3.0	2.9	2.6	6.81	6.46	4.30	4.43	4.13	3.21	3.14	3.12	3.30	3.04
Male	2.9	2.6	2.4	2.4	2.2	5.18	5.06	3.30	3.47	3.23	2.42	2.46	2.48	2.68	2.45
Female	.64	.57	.52	.51	.47	3.21	3.17	2.51	2.55	2.35	2.28	2.34	2.09	2.21	2.04
Working 35 hr.+	2.3	2.1	2.0	1.9	1.7	5.43	5.22	3.58	3.45	3.37	2.72	2.69	2.77	2.65	2.60
Unpaid family workers and self-employed	2.3	2.1	2.0	1.9	1.8	6.63	6.48	3.95	3.91	3.73	2.90	3.02	2.69	2.69	2.53
Employed as Wage and Salary Worker															
In nonag.	45	47	50	50	50	5.80	7.14	7.45	8.09	8.42	1.32	1.27	1.16	1.24	1.39
In ag.	1.2	1.0	1.0	1.0	.87	4.08	3.45	3.16	3.61	2.99	3.16	2.82	2.95	3.38	2.86
In durables or nondurables (mfg.)	14	15	16	16	15	2.78	3.57	3.13	3.41	3.42	1.71	2.10	1.73	1.88	2.05
In durables	8.1	8.6	9.3	9.2	9.2	2.49	3.34	2.68	2.72	2.67	1.92	2.46	1.89	1.84	1.82
In construction	2.9	2.9	3.0	3.0	3.0	1.56	1.66	1.54	1.61	1.84	1.28	1.30	1.27	1.31	1.46
In retail trade	6.8	6.9	7.5	7.5	7.5	1.65	1.77	1.89	1.78	2.08	1.22	1.21	1.33	1.32	1.52
In service industry, including private household workers	13	14				2.63	2.63	2/	2/	2/	1.80	1.75	2/	2/	2/
As private household workers			1.6	1.5	1.4	2/	2/	1.54	1.53	1.52	2/	2/	1.30	1.28	1.22
Unemployed															
Total	2.7	2.3	2.5	2.3	2.2	1.48	1.43	1.50	1.55	1.56	1.25	1.29	1.27	1.35	1.37
Wage and salary workers in durables or nondurables (mfg.)	.60	.60	.65	.56	.56	1.23	1.26	1.23	1.31	1.34	1.16	1.17	1.18	1.22	1.27
White Unemployed															
Males			1.0	.91	.88	2/	2/	1.35	1.27	1.23	2/	2/	1.25	1.16	1.15
Females			1.0	.91	.91	2/	2/	1.24	1.26	1.29	2/	2/	1.13	1.14	1.17
Males, 16-19			.28	.28	.27	2/	2/	1.17	1.23	1.15	2/	2/	1.08	1.13	1.08
Females, 16-19			.26	.26	.26	2/	2/	1.16	1.19	1.17	2/	2/	1.05	1.07	1.05

See footnotes at end of table.

TABLE 1 (Cont'd) Design Effects for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Characteristic	Percent of Population possessing characteristic					Design Effects[1]									
	Population 14+		Population 16+			Unbiased Estimate					1st & 2nd Stage Combined Est.				
	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969	1965	1966	1967	1968	1969
Negro and Other Races Unemployed															
Males			.24	.22	.20	2/	2/	1.40	1.53	1.72	2/	2/	1.07	1.26	1.32
Females			.28	.26	.25	2/	2/	1.49	1.53	1.52	2/	2/	1.19	1.21	1.28
Males, 16-19			.08	.08	.08	2/	2/	1.23	1.29	1.39	2/	2/	1.00	1.12	1.14
Females, 16-19			.08	.07	.08	2/	2/	1.23	1.24	1.39	2/	2/	1.10	1.14	1.13
Rural															
Total nonfarm	29	24	24	25	25	15.29	17.36	17.88	18.66	18.34	7.95	11.12	10.01	10.69	10.62
Total farm	6.4	6.1	5.6	5.4	5.1	12.35	12.66	7.77	7.66	7.48	5.37	5.19	5.65	5.63	5.48
Male farm	3.4	3.1	2.9	2.8	2.6	6.67	6.61	4.07	4.03	4.00	2.94	2.77	2.94	3.00	2.93
Miscellaneous															
Total Males	47	47	46	47	47	4.67	5.74	5.23	5.20	5.20	0	0	0	0	0
Total Females	53	53	54	53	53	5.00	6.14	5.92	6.02	5.78	0	0	0	0	0
Total Persons of Negro & Other Races	11	11	11	11	11	10.89	12.01	8.95	9.97	10.24	0	0	0	0	0
Household Heads	44	44	46	46	46	3.15	4.12	3.89	3.78	3.75	.61	.65	.58	.52	.54
At school	8.2	8.1	5.0	5.1	5.0	2.17	2.30	1.77	1.97	1.83	.71	.68	.90	1.04	1.05

TABLE 2 Design Effects for Characteristics which are Subsets of Age-Sex-Race Groups for the Unbiased Estimate and Ratio Estimate for 1965 through 1969

Civilian Labor Force															
Female	38	39	42	42	43	3.28	3.39	3.56	3.64	3.78	1.44	1.39	1.26	1.25	1.38
Under 20 Yr.	36	38	49	49	50	2.53	2.88	3.11	3.28	3.77	1.21	1.20	1.17	1.24	1.33
Employed															
Negro and Other Races	50	54	57	54	58	12.25	12.88	12.13	13.76	15.59	.92	1.10	1.04	1.09	1.33
Males	73	73	77	78	78	8.87	11.44	12.32	12.16	12.50	.78	.78	.77	.82	.90
Nonag. Employed															
Males	67	68	71	73	73	7.35	9.57	10.44	10.33	10.79	.98	1.00	1.04	1.05	1.08
Females	34	36	38	39	40	2.94	3.14	3.27	3.40	3.53	1.38	1.39	1.26	1.26	1.34
Ag. Employed															
Males	5.9	5.3	5.1	5.1	4.7	5.35	5.21	3.40	3.57	3.32	2.50	2.53	2.55	2.76	2.51
Females	1.2	1.1	1.0	1.0	.9	3.22	3.18	2.46	2.56	2.36	2.29	2.36	2.10	2.22	2.05
White Unemployed															
Males			2.3	2.2	2.1	2/	2/	1.37	1.28	1.25	2/	2/	1.26	1.17	1.16
Females			2.0	1.9	1.9	2/	2/	1.25	1.27	1.30	2/	2/	1.14	1.16	1.18
Males, 16-19			6.6	6.2	6.2	2/	2/	1.25	1.30	1.22	2/	2/	1.15	1.20	1.14
Females, 16-19			5.6	5.7	5.8	2/	2/	1.22	1.26	1.23	2/	2/	1.11	1.13	1.11
Negro & Other Races Unemployed															
Males			4.7	4.6	4.3	2/	2/	1.48	1.61	1.76	2/-	2/	1.12	1.32	1.43
Females			4.9	4.4	4.2	2/	2/	1.57	1.60	1.58	2/	2/	1.25	1.26	1.33
Males, 16-19			12	12	11	2/	2/	1.42	1.47	1.57	2/	2/	1.14	1.27	1.28
Females, 16-19			10	10	10	2/	2/	1.40	1.39	1.55	2/	2/	1.23	1.27	1.26
Rural															
Male farm	6.9	6.5	6.0	6.0	5.7	6.95	6.86	4.21	4.18	4.13	3.06	2.87	3.03	3.10	3.03

1/ Twelve months of Keyfitz data were averaged for the numerators of the factors.
2/ No Keyfitz variance estimate available.

TABLE 3 Total Between[1/] and Between Stratum Variance as Proportions of Total Variance, 1967-1969 Averages 1st and 2nd Stage Combined Ratio Estimate

Characteristic	Total Between	Between Stratum
Civilian Labor Force		
Total	.099	.019
Females	.082	.013
Ages, 16-19	.037	.006
Employed (Average)	.076	.010
Negro and Other Races	.062	.002
Males	.065	.019
Working 1-14 hrs.	.069	.002
Part-time (1-34 hrs) usually full-time	.117	.016
Part-time for economic reasons, usually full-time	.090	.010
Self-employed	.053	.013
Nonagriculture Employed (Average)	.048	.012
Total	.036	.014
Males	.027	.013
Females	.071	.011
Working 35+ hours	.050	.014
With a job, not at work	.054	.007
Agriculture Employed (Average)	.114	.017
Total	.123	.016
Males	.111	.021
Females	.255	.003
Working 35+ hours	.105	.026
Unpaid family and self-employed	-.026	.018
Wage & Salary Workers (Average)	.121	.020
In nonagriculture	.043	.016
In agriculture	.288	.016
In durables or nondurables (mfg.)	.224	.059
In durables	.225	.039
In construction	.035	.002
In retail trade	.011	.008
As private household workers	.024	.000
Unemployed		
Total	.023	.008
Wage and salary workers in durables or nondurables (mfg.)	.008	.010
White Unemployed		
Males	.006	.008
Females	.034	.006
Males, 16-19	-.003	.008
Females, 16-19	.039	.002
Negro & Other Races Unemployed		
Males	.005	.003
Females	.012	-.002
Males, 16-19	.014	-.003
Females, 16-19	-.026	-.009
AVERAGE OF ALL UNEMPLOYMENT ITEMS	.011	.003
Rural		
Total nonfarm	-.136	-.007
Total farm	-.130	.019
Male farm	-.133	.017
Miscellaneous		
At school	.024	.008
Household heads	.008	.004
AVERAGE OF ALL ITEMS	.051	.011

TABLE 4 Ratio of the Variance of the Composite Estimate to the Variance of the Noncomposite Estimate, 1969 Data

	Ratio of the Annual Averages of Monthly Level	Monthly Change
Civilian Labor Force		
Total	.8560	.6895
Females	.8384	.6619
Ages, 16-19	.9788	.8537
Employed (Average)		
Negro and Other Races	.8746	.7799
Males	.8976	.8043
Working 1-14 hrs.	1.0620	.9583
Part-time (1-34 hrs) usually full-time	1.2226	1.1478
Part-time for economic reasons, usually full-time	1.1834	1.0866
Self-employed	.8663	.6992
Nonagriculture Employed (Average)		
Total	.8628	.6957
Males	.9061	.8045
Females	.8385	.6398
Working 35+ hours	.9725	.9035
With a job, not at work	1.1485	1.0633
Agriculture Employed (Average)		
Total	.9033	.7162
Males	.8899	.7003
Females	.9274	.8223
Working 35+ hours	.8861	.7655
Unpaid family and self-employed	.9095	.6889
Wage & Salary Workers (Average)		
In nonagriculture	.8560	.6881
In agriculture	.9353	.8791
In durables or nondurables (mfg.)	.8373	.5940
In durables	.8280	.6245
In construction	.8319	.7390
In retail trade	.8305	.6893
As private household workers	.8796	.8271
Unemployed		
Total	1.1096	.9939
Wage and salary workers in durables or nondurables (mfg.)	1.0514	.9714
White Unemployed		
Males	1.0663	.9972
Females	1.0824	1.0044
Males, 16-19	1.1371	1.0453
Females, 16-19	1.0926	1.0391
Negro & Other Races Unemployed		
Males	.9706	.9584
Females	1.1172	.9973
Males, 16-19	1.0223	1.0004
Females, 16-19	1.1594	.9959
AVERAGE OF ALL UNEMPLOYMENT ITEMS	1.0809	1.0003
Rural		
Total nonfarm	.9081	.4382
Total farm	.8785	.5345
Male farm	.8736	.5359
Miscellaneous		
At school	.9245	.7996
Household heads	.7693	.5358
AVERAGE OF ALL ITEMS	.9558	.8139

1/ Total between variance is the sum of between PSU variance and "between stratum" variance.

TABLE 5 Ratio of the Variances of Monthly Averages for Three, Six, Nine and Twelve Months' to One Month's Estimate and Ratio of the Variance of Differences Between Two Months to the Variance of Monthly Level 1st and 2nd Stage Combined Ratio Estimate, 1969 and 1970 Data[1/]

Characteristic	Averages of Three Months	Six Months	Nine Months	Twelve Months	Differences of Two Adjacent Months	Two Months A Year Apart
Civilian Labor Force						
Total	.75	.58	.49	.46	.66	1.00
Females	.76	.58	.48	.43	.63	1.11
Ages, 16-19	.60	.41	.33	.29	1.12	1.56
Employed						
Negro and Other Races	.72	.53	.44	.40	.70	1.31
Males	.68	.49	.41	.37	.82	1.26
Working 1-14 hrs.	.53	.34	.25	.21	1.31	1.65
Part-time (1-34 Hrs) usually full-time	.43	.26	.20	.18	1.67	1.82
Part-time for economic reasons, usually full-time	.43	.25	.18	.14	1.67	1.88
Self-employed	.77	.60	.50	.47	.61	1.16
Nonagriculture Employed						
Total	.76	.60	.51	.49	.61	1.03
Males	.74	.58	.50	.47	.66	1.12
Females	.76	.58	.49	.44	.62	1.21
Working 35+ hours	.65	.48	.41	.38	.97	1.38
With a job, not at work	.43	.25	.16	.13	1.66	1.62
Agriculture Employed						
Total	.83	.69	.61	.58	.42	.73
Males	.84	.70	.63	.60	.39	.71
Females	.71	.51	.41	.34	.74	1.12
Working 35+ hours	.77	.62	.55	.51	.56	.81
Unpaid family and self-employed	.83	.68	.59	.55	.42	.79
Wage and Salary Workers						
In nonagriculture	.77	.61	.53	.51	.58	1.03
In agriculture	.77	.61	.53	.48	.57	1.06
In durables or nondurables (mfg)	.80	.65	.58	.54	.45	.79
In durables	.78	.63	.55	.50	.49	.79
In construction	.72	.50	.38	.31	.68	1.39
In retail trade	.75	.55	.44	.37	.64	1.44
As private household workers	.65	.45	.35	.30	.94	1.43
Unemployed						
Total	.51	.31	.23	.20	1.33	1.71
Wage and salary workers in durables or nondurables (mfg.)	.49	.30	.22	.19	1.38	1.75
White Unemployed						
Males	.48	.30	.23	.19	1.42	1.66
Females	.49	.30	.21	.16	1.40	1.88
Males, 16-19	.45	.26	.19	.16	1.59	2.00
Females, 16-19	.45	.25	.17	.12	1.60	1.78
Negro & Other Races Unemployed						
Males	.49	.29	.21	.18	1.45	1.84
Females	.47	.28	.20	.17	1.52	1.80
Males, 16-19	.46	.29	.22	.20	1.55	1.86
Females, 16-19	.47	.27	.19	.16	1.44	1.86
AVERAGE OF ALL UNEMPLOYMENT ITEMS	.48	.29	.21	.17	1.47	1.81
Rural						
Total nonfarm	.90	.79	.76	.73	.22	.64
Total farm	.87	.73	.66	.63	.31	.78
Male farm	.86	.71	.64	.62	.34	.78
Miscellaneous						
At school	.60	.42	.34	.30	1.06	1.45
Household heads	.74	.53	.43	.40	.57	1.10
AVERAGE OF ALL ITEMS	.66	.48	.40	.36	.92	1.33

1/ January 1969 through June 1970 data were used.

TABLE 6

Comparison of Relvariances Computed by the Keyfitz and Replication[1/] Methods, 1964 Data

Item[3/]	1st & 2nd Stage Combined Ratio Estimate					Composite Estimate				
	Keyfitz	Replication	Column (2) / Column (1)	Keyfitz	Replication	Keyfitz[2/]	Replication	Column (7) / Column (6)	Keyfitz[2/]	Replication
	Average Relvariance ($x10^{-8}$)			1/6 of range of monthly figures ($x10^{-8}$)		Average Relvariance ($x10^{-8}$)			1/6 of range of monthly figures ($x10^{-8}$)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Total labor force	1,049	781	.74	60	163	1,068	654	.61	60	119
Females in labor force	5,604	4,355	.78	300	365	5,301	4,359	.82	360	681
Employed males	855	987	1.15	70	86	647	915	1.41	60	142
Total nonwhite employed	12,575	15,032	1.20	1,300	3,276	11,787	14,191	1.20	1,300	3,869
Total ag. employed	126,176	137,657	1.09	7,000	24,923	113,742	127,708	1.12	6,700	29,242
Male ag. employed	109,862	125,492	1.14	10,000	19,855	96,723	113,007	1.17	7,200	22,501
Female ag. employed	559,273	545,447	.98	54,000	110,163	538,806	522,559	.97	60,000	93,949
Total nonag. employed	1,669	1,275	.76	70	169	1,411	1,154	.82	60	94
Employed males in nonag.	1,375	1,378	1.00	55	271	1,130	1,344	1.19	75	267
Employed females in nonag.	6,702	5,563	.83	550	797	5,743	5,279	.92	320	1,026
Full-time nonag. employed	2,741	2,180	.80	650	483	2,339	2,526	1.08	900	604
Employed in nonag. - With job, not at work	63,577	81,612	1.28	16,700	26,032	73,953	95,287	1.29	19,000	22,645
Unemployed	53,913	55,909	1.04	6,150	8,842	46,612	63,127	1.35	5,300	8,747
Average			.98					1.07		

1/ Twenty replications were used. 2/ These have been adjusted by applying correction factors. See text.

3/ All items used were for population 14 and over.

APPENDIX A.-Percent Reduction in Variance Due to Use of Three Months' Data Rather than One

For the estimate of a total (x) based on a single rotation group,

(1) $VAR(x) = N^2 \frac{\sigma^2}{n}$, where n is the number of sample cases in 1 rotation group and N is total population for the country.

For the estimate of a total based on 1 full month's data,

(2) $VAR(x) = \sum_{i=1}^{8} \frac{N^2}{(8)^2} \frac{\sigma^2}{n} = \frac{N^2\sigma^2}{8n}$

For the estimate of a total based on 3 consecutive months' data,

(3) $VAR(x) = \sum_{i=1}^{24} \frac{N^2\sigma^2}{(24)^2 n} + 2\sum_{i=1}^{16} \frac{N^2\rho_s\sigma^2}{(24)^2 n} + 2\sum_{i=1}^{8} \frac{N^2\rho_d\sigma^2}{(24)^2 n}$

where ρ_s is the correlation between identical rotation groups, and ρ_d is the correlation between non-identical, but matching, rotation groups.

In other words,

(4) $VAR(x) = \frac{N^2\sigma^2}{8n} \left(\frac{3}{9} + \frac{4}{9}\rho_s + \frac{2}{9}\rho_d\right)$

(5) $R_x = \frac{1}{3} + \frac{4\rho_s}{9} + \frac{2\rho_d}{9}$,

where R_x = the ratio of the variance of the monthly average for three months to the variance for one month's estimate.

If we assume $\rho_s = 1$ and $\rho_d = 0$, we get $R_x = \frac{7}{9}$,

i.e., there is a $\frac{2}{9}$, or 22% reduction in the variance compared to 1 month's data.

REFERENCES

1. Gurney, Margaret, "Sampling for Time Series", Internal Census Bureau memo. to Messrs. Wm. Hurwitz & Max Bershad, April 26, 1963.
2. Keyfitz, Nathan, "Estimates of Sampling Variance When Two Units are Selected From Each Stratum," Journal of the ASA 52 (Dec. 1957) pp. 503-510.
3. McCarthy, P. J., "Pseudo-Replication: Half-Samples," Review of the Internat'l. Statistical Inst. 37, No. 3 (1969), pp. 239-264.
4. Shapiro, Gary M., "Keyfitz Method of Estimating Variance & Its Application to the Current Pop. Survey" Internal Census Bureau Document, Sept. 1, 1966.
5. Tepping, Benj. J., "Variance Estimation in Complex Surveys" Proceedings of the Social Statistics Section 1968 ASA, pp. 11-18.
6. Woltman, Henry & Turner, Anthony G., "Two Additional Methods for Estimating Total NSR Variance by the Keyfitz Method for the CPS-MLS Merged Surveys", Internal Census Bureau memo. to Robert Hanson, Jan. 13, 1967.
7. Woltman, Henry, "Variance Estimation Formulas to be Used in Computing Total NSR PSU Variances for the 449-Area Design", Internal Census Bureau document, March 12, 1968.
8.__________U.S. Bureau of the Census, the Current Population Survey-A Report on Methodology, Technical Paper No. 7, 1963.