# A COMPARISON OF VARIANCE ESTIMATORS USING THE TAYLOR SERIES APPROXIMATION

Cathryn S. Dippo, Bureau of Labor Statistics
Kirk M. Wolter, Bureau of the Census

## I. INTRODUCTION

The selection of a variance estimator for large complex sample surveys is not straightforward. Most of the methods of variance estimation for such surveys are based upon some form of repeated subsampling. The random group, jackknife and balanced repeated replication methods differ primarily in the procedures for forming the subsamples. Previous comparative studies have been primarily empirical. One of our goals is to compare analytically the accuracy of these different subsample variance estimators.

A first order Taylor series approximation is widely used in computing variances for complex surveys; however, the analytical properties of the random group, jackknife and balanced repeated replication variance estimators are indistinguishable in their first order term. Koop (1968) hypothesized the underestimate of variance he found was due to neglecting terms of order $1/n^2$ and $1/n^3$. Sukhatme and Sukhatme (1970) suggested the use of a second order approximation. Our method of comparing these variance estimators is to include all the terms of order $n^{-2}$ in the Taylor series expansion. A more complete description of these procedures, which are summarized in section II below, can be found in Dippo (1981). Concurrent with our work, Rao and Wu (1983) have made an asymptotic second order comparison of the linearization, jackknife and balanced repeated replication methods.

The second order Taylor series analytical expressions obtained for the bias of the variance estimators are complex and require the population moments and derivatives for evaluation. Since a second goal of this research is to investigate the properties of the variance estimators when the sample size is small and the underlying population is extremely skewed, the 1980-81 Consumer Expenditure Diary Survey (CES) is used to evaluate the second order Taylor series expressions in section III. The CES, which is a complex multistage sample of only 5000 housing units per year with a design similar to the Current Population Survey, produces national estimates of mean expenditures, which have a distribution that is closer to log-normal than normal.

Previous empirical studies such as Frankel (1971) and Bean (1975) have concentrated on the effects of the complex multistage cluster designs as represented by the Current Population Survey and the Health Interview Survey instead of the effects of the shape of the underlying finite population. Furthermore, although functionally equivalent, the ratio estimator of mean expenditures is conceptually different from the ratio estimator of a proportion investigated in Frankel and Bean. In section IV, the results of a Monte Carlo investigation of the confidence interval properties of the random group and balanced repeated replication variance estimators for the skewed CES data are presented.

## II. TAYLOR SERIES APPROXIMATION

The method used to compare the different variance estimators is to approximate each estimator using a Taylor series expansion, including terms of order $1/n^2$.

Consider a finite population of N units divided into L strata. A simple random sample of $n_h$ units are selected from the $N_h$ units in the h-th stratum with sampling independent between strata. Let $x_{rhi}$ be the observed value of the r-th variable for the i-th unit from the h-th stratum. Define the stratum mean for the r-th variable in the h-th stratum as

$\bar{X}_{rh} = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{rhi}$ and its corresponding sample mean as $\bar{x}_{rh} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{rhi}$.

The class of parameters to be considered is that which can be expressed as a function of stratum means, $\theta = F(\bar{X}_{11},\ldots,\bar{X}_{rh},\ldots,\bar{X}_{RL}) = F(\bar{\underset{\sim}{X}})$ where $r = 1, \ldots, R$ and $h = 1, \ldots, L$. The estimator of $\theta$ to be considered is the same function of the sample means, $\hat{\theta} = F(\bar{\underset{\sim}{x}})$.

Let us assume $F(\cdot)$ is a real-valued function on RL-dimensional Euclidian space with continuous partial derivatives of order five at $\bar{\underset{\sim}{X}}$ and define

$$F^{(1)}(\bar{X}_{rh}) = \left.\frac{\partial F(\bar{\underset{\sim}{x}})}{\partial \bar{x}_{rh}}\right|_{\bar{\underset{\sim}{X}}} \quad ,\ldots,$$

$$F^{(4)}(\bar{X}_{rh}\bar{X}_{sh}\bar{X}_{th}\bar{X}_{zh}) = \left.\frac{\partial F(\bar{\underset{\sim}{x}})}{\partial \bar{x}_{rh}\,\partial \bar{x}_{sh}\,\partial \bar{x}_{th}\,\partial \bar{x}_{zh}}\right|_{\bar{\underset{\sim}{X}}} ,$$

which are the first four partial derivatives of $\hat{\theta}$ evaluated at the point where each $\bar{x}_{rh}$ is equal to its expected values $\bar{X}_{rh}$. Furthermore, let $\bar{u}_{rh} = \bar{x}_{rh} - E(\bar{x}_{rh}) = \bar{x}_{rh} - \bar{X}_{rh}$. Then,

$$\begin{aligned}\hat{\theta} = {} & \theta + \sum_r \sum_h \bar{u}_{rh} F^{(1)}(\bar{X}_{rh}) \\ & + \frac{1}{2} \sum_{r,s} \sum_{h,h'} \bar{u}_{rh}\bar{u}_{sh'} F^{(2)}(\bar{X}_{rh}\bar{X}_{sh'}) \\ & + \frac{1}{6} \sum_{r,s,t} \sum_{h,h',h''} \bar{u}_{rh}\bar{u}_{sh'}\bar{u}_{th''} F^{(3)}(\bar{X}_{rh}\bar{X}_{sh'}\bar{X}_{th''}) \\ & + \frac{1}{24} \sum_{r,s,t,z} \sum_{h,h',h'',h'''} \bar{u}_{rh}\bar{u}_{sh'}\bar{u}_{th''}\bar{u}_{zh'''} \\ & \qquad F^{(4)}(\bar{X}_{rh}\bar{X}_{sh'}\bar{X}_{th''}\bar{X}_{zh'''}) + O_p(\max_h n_h^{-5/2}),\end{aligned} \tag{1}$$

and

$$\operatorname{var}\hat{\theta} = E(\hat{\theta} - E\hat{\theta})^2$$

$$= \sum_{r,s} \sum_h E(\bar{u}_{rh}\bar{u}_{sh})\, F^{(1)}(\bar{X}_{rh})F^{(1)}(\bar{X}_{sh})$$

$$+ \sum_{r,s,t} \sum_h E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})F^{(1)}(\bar{X}_{rh})F^{(2)}(\bar{X}_{sh}\bar{X}_{th})$$

$$+ \frac{1}{3} \sum_{r,s,t,z} \sum_h E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh})F^{(1)}(\bar{X}_{rh}) F^{(3)}(\bar{X}_{sh}\bar{X}_{th}\bar{X}_{zh})$$

$$+ \sum_{r,s,t,z} \sum_{h \neq h'} E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th'}\bar{u}_{zh'})F^{(1)}(\bar{X}_{rh}) F^{(3)}(\bar{X}_{sh}\bar{X}_{th'}\bar{X}_{zh'})$$

$$+ \frac{1}{4} \sum_{r,s,t,z} \sum_h [E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh}) -E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th}\bar{u}_{zh})]F^{(2)}(\bar{X}_{rh}\bar{X}_{sh})F^{(2)}(\bar{X}_{th}\bar{X}_{zh})$$

$$+\frac{1}{2} \sum_{r,s,t,z} \sum_{h \neq h'} E(\bar{u}_{rh}\bar{u}_{th})E(\bar{u}_{sh'}\bar{u}_{zh'}) F^{(2)}(\bar{X}_{rh}\bar{X}_{sh'})F^{(2)}(\bar{X}_{th}\bar{X}_{zh'})+0(\max n_h^{-3}), \quad (2)$$

where $O_p$ is the usual notation for "bounded in probability" and where $N_h$ and $n_h \xrightarrow{p} \infty$ in such a way that $n_h/N_h \to \lambda_h^{\circ}$. The expectation operator is defined with respect to the probability distribution generated by repeated sampling using the stratified design described above.

To make expressions such as (1) and (2) rigorous, we must work in terms of a conceptual sequence of finite populations of increasing size. Krewski and Rao (1981) and Isaki and Fuller (1982) give alternative formulations of the conceptual sequence. In this paper, the assumed sequence is such that L is fixed and the strata sample sizes $n_h$ increase without bound. To simplify the presentation, the formal definition of the sequence is omitted.

Each of the estimators considered is based upon dividing the sample of size $n = \sum_h n_h$ into different subgroups. For the random group (RG) method, the $n_h$ sample units in each stratum are divided into k equal groups of size $m_h$. The random group estimator of $\theta$ is

$\hat{\theta}_{RG} = \frac{1}{k} \sum_{\alpha=1}^{k} \hat{\theta}_\alpha$, where $\hat{\theta}_\alpha$ is the estimator of the same functional form as $\theta$ based upon the $\alpha$-th random group. The random group estimator of the variance of $\hat{\theta}$ to be considered here is

$$\hat{V}_{RG}(\hat{\theta}) = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}'_\alpha - \hat{\theta}'_{RG})^2,$$

where $\hat{\theta}'_\alpha = F(\sqrt{1-\lambda_h}\, \bar{x}_{rh\alpha})$, $\lambda_h = n_h/N_h$, and $\bar{x}_{rh\alpha}$ is the sample mean based upon the $\alpha$-th random group.

The jackknife (JKK) estimator to be considered is one proposed by Jones (1974)

$$\hat{\theta}_{JKK}= (1 + \sum_h W_h)\hat{\theta} - \sum_h \sum_i^{n_h} a_h \hat{\theta}_{(hi)}$$

where $W_h = (N_h - n_h)(n_h - 1)/N_h$, $a_h = W_h/n_h$, and $\hat{\theta}_{(hi)}$ is the estimator of the same functional form as $\theta$ but which omits the hi-th unit. Therefore, the jackknife estimator is based upon n groups of size n-1. Jones' jackknife estimator of the variance of $\hat{\theta}$ is

$\hat{V}_{JKK}(\hat{\theta})=\sum_h\sum_i a_h(\hat{\theta}_{(hi)}- \hat{\theta}_{(h)})^2$, where $\hat{\theta}_{(h)}=\frac{1}{n_h}\sum_i\hat{\theta}_{(hi)}$.

The balanced repeated replication (BRR) estimator to be considered is similar to the random group estimators in that the $n_h$ units in each stratum are first divided into (k=) two random groups of size $m_h=n_h/2$. However, instead of forming only two groups across strata as in the random group estimator, orthogonally balanced combinations of one random group from each stratum are created. In studying this estimator, we use k' balanced half samples, where k' is the smallest integral multiple of four greater than or equal to the number of strata L. The BRR estimator of $\theta$ is

$\hat{\theta}_{BRR} = \frac{1}{k'} \sum_{\alpha}^{k'} \hat{\theta}_\alpha$, where $\hat{\theta}_\alpha$ is the estimator of the same functional form as $\theta$ based upon the $\alpha$-th replicate. The BRR estimator of the variance of $\hat{\theta}$ is

$\hat{V}_{BRR}(\hat{\theta})=\frac{1}{k'}\sum_{\alpha}^{k'}(\hat{\theta}'_\alpha- \hat{\theta}'_{BRR})^2$, where $\hat{\theta}'_\alpha=F(\sqrt{1-\lambda_h}\,\bar{x}_{rh\alpha})$.

By expanding $\hat{\theta}_\alpha$, $\hat{\theta}'_\alpha$, $\hat{\theta}_{(hi)}$, etc, in a Taylor series, as in (1), approximations for the bias and variance of $\hat{\theta}_{RG}$, $\hat{\theta}_{JKK}$, $\hat{\theta}_{BRR}$, $\hat{V}_{RG}(\hat{\theta})$, $\hat{V}_{JKK}(\hat{\theta})$, and $\hat{V}_{BRR}(\hat{\theta})$ can be obtained. Table 1 presents the coefficients of the derivatives (columns) in terms of expected values needed to express the bias of the variance estimators in a Taylor series. To construct the exact expression for the bias of one of the estimators, substitute the coefficient in the table for the one corresponding to the same derivative in (2). For example,

$$\text{Bias } \hat{V}'_1(\hat{\theta}) = \sum_{r,s,t} \sum_h \left\{\frac{(1-\lambda)}{(k-1)}^{3/2}[E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})] - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})\right\} F^{(1)}(\bar{X}_{rh})F^{(2)}(\bar{X}_{sh}\bar{X}_{th}) + \ldots .$$

These expressions are complex and their interpretation requires knowledge of the population moments and derivatives.

The bias of the random group estimator with two random groups per stratum differs from that of the BRR estimator in the cross-stratum $F^{(2)}(\bar{X}_{rh}\bar{X}_{sh'})F^{(2)}(\bar{X}_{th}\bar{X}_{zh'})$ term only (see table 1). This larger between stratum component of the BRR estimator makes the BRR estimator more biased than the random group estimator with two random groups. However, when the number of random groups is increased, the random group estimator becomes more biased than the BRR estimator. This reflects the fact that as more random groups are formed, the sample size per random group used in computing each $\hat{\theta}_\alpha$ decreases. Again, the two parameters where the numerator is a subset of the denominator are exceptions. The reader will recall the variance of the variance estimator is generally a decreasing function of the number of random groups, k, whereas here we find the bias of variance estimator is an increasing function of k.

For sample sizes of $n_h$ = 6 and 12, the jackknife is the least biased with the exception of the parameter average weekly cost of gasoline per vehicle. However, when the sample size is $n_h$=24, the bias of the jackknife variance estimator is similar to that of the random group variance estimator with two random groups.

## III. COMPARISON OF VARIANCE ESTIMATORS

### A. STUDY POPULATION

In order to obtain some insight into the properties of the variance estimators, data from the 1980-81 Consumer Expenditure Diary Survey has been treated as a finite population. The 14,360 consumer units (CU's) classified as complete income reporters have been divided into 20 approximately equal sized strata based upon region and city size. A consumer unit is a single financially independent consumer or a family of two or more persons living together, pooling incomes and drawing from a common fund for major expenditures. The following thirteen parameters have been considered:

$R_1$ = average cost per reporting CU for flour (FLOUR)

$R_2$ = average cost per reporting CU for ground beef (GRBEEF)

$R_3$ = average cost per reporting CU for eggs (EGGS)

$R_4$ = average cost per reporting CU for candy and chewing gum (CANDY)

$R_5$ = average cost per reporting CU for food away from home (FOODAWAY)

$R_6$ = average cost per reporting CU for food at home (FOODHOME)

$R_7$ = average cost per reporting CU for gasoline (GASCOST)

$R_8$ = average number of vehicles per CU owning at least one vehicle (VEHQ_FAM)

$R_9$ = average annual CU income before taxes (FINCBEFX)

$R_{10}$= average per capita wage and salary income (WAGE_CAP)

$R_{11}$= average weekly gasoline cost per vehicle (GAS_VEHQ)

$R_{12}$= proportion wage and salary income of total CU income before taxes (WAGE_INC)

$R_{13}$= proportion of civilian labor force that was unemployed during last 12 months (UNEM_CLF).

The variable names in the parenthesis above such as FLOUR and GRBEEF are used in the accompanying tables.

For $R_1$ to $R_{13}$ , the general form of the estimator is $\frac{\sum_h N_h \bar{x}_{1h}}{\sum_h N_h \bar{x}_{2h}}$ . For example, in $R_1$,

$x_{1hi}$ = the cost reported by the hi-th CU for flour in one week

and

$$x_{2hi} = \begin{cases} 1 \text{ if the hi-th CU purchased flour during the week} \\ 0 \text{ if the hi-th CU did not purchase flour during the week.} \end{cases}$$

In $R_{10}$, $x_{1hi}$ is the total annual wage and salary income reported by the hi-th CU and $x_{2hi}$ is the number of persons in the hi-th CU who reported wage and salary income. $R_9$ is a linear estimator since $x_{2hi}$ = 1 for all units. $R_{12}$ and $R_{13}$ differ from the others in that the numerator is a subset of the denominator, e.g., wage and salary income is a subset of total income.

Table 2 displays some basic distribution statistics for the expenditure and income variables. Figures 1 and 2 present the frequency distributions of weekly expenditures for ground beef and food away from home. All of the expenditure and income variables exhibit similarly skewed distributions. Since the second order term in the Taylor series is a function of the third order moments, one might expect the first order Taylor series approximations to the bias and variance of $\hat{\theta}$ to be biased when the finite population is highly skewed. Indeed, later results confirm this hypothesis.

### B. TAYLOR SERIES APPROXIMATION TO VAR $\hat{\theta}$

Table 3 indicates the general magnitude and sign of the population moments by presenting the average stratum population moments and derivatives, which are similar in magnitude. When the corresponding individual stratum moments and derivatives are substituted in the expression (2) for var $\hat{\theta}$, the second order Taylor series approximation to the variance is obtained.

Table 4 presents the second order Taylor series approximation to the variance along with the proportion of the variance associated with the first and second order terms for three sample sizes. For the ten nonlinear parameters where the numerator is not a subset of the denominator, the first order term accounts for about 98.9% of the variance and the second order terms 1.1% when the stratum sample size is $n_h$=6. When the sample size is doubled to $n_h$=12, the relationship is 99.4% to .6%. When doubled again to $n_h$=24, the relationship is 99.7% to .3%. On the other hand, when the numerator is a subset of the denominator (UNEM_CLF and WAGE_INC), the first order term accounts for 100.8% of the variance when the stratum sample size is 6, 100.4% when $n_h$ is 12, and 100.2% when $n_h$ is 24. That is, the total of the second order terms is negative and the first order approximation provides an overestimate of the variance. Overall, for the three sample sizes and the twelve nonlinear parameters examined, the percent of the Taylor series approximation to the variance associated with the second order terms is at most 5%.

In table 5, the percent of the Taylor series approximation to the variance associated with the second order terms when the sample size is six per stratum is given in rank order along with the derivatives which are not a function of the numerator. All the derivatives which are a function of $\bar{X}_{1h}$ or the numerator do not show a relationship with the relative importance of the second order portion of the variance and, therefore, are not shown. An examination of table 5, ignoring the four parameters with non-indicator function denominators, indicates the second order variance becomes more important as the proportion of the population purchasing an item in a given week decreases.

### C. COMPARISON OF THE EXPECTATION OF THE VARIANCE ESTIMATORS

The first order Taylor series approximations to the expectations of the random group, jackknife and BRR variance estimators are identical. Therefore, although the contribution to the expectation of the variance estimators from the second order terms may be small, an analysis of the second order terms should give some indication of the relative merits of the different estimators.

The expectations of the random group, jackknife and BRR variance estimators obtained by substituting the population moments and derivatives in the Taylor series approximation to $E[\hat{V}(\hat{\theta})]$ are compared to the Taylor series approximation to the variance of $\hat{\theta}$, var $\hat{\theta}$, in table 6. These expectations are computed by substituting the population moments and derivatives in the formulas given in table 1. The appropriate finite population sampling coefficients of the population moments are also needed. For example, $(N_h - n_h)/[n_h(N_h - 1)]$ is the coefficient of the full sample second order stratum population moment when sampling is without replacement. All of the variance estimators are positively biased for each sample size for all the ratio estimation parameters where the numerator is not a subset of the denominator. When the numerator is a subset of the denominator, the random group estimator is negatively biased for all three sample sizes and each choice of the number of random groups. The BRR estimator is negatively biased for the larger proportion. The jackknife estimator is negatively biased for only one of the proportions and for only the smallest sample size.

## IV. MONTE CARLO INVESTIGATION

For comparison purposes, 1000 without replacement samples of size 6, 12, and 24 units per stratum have been selected, resulting in samples with total size 120, 240 and 480. Two additional parameters have been considered for the Monte Carlo portion of this study:

$R_{14}$ = correlation between total food at home and family income (R_FH_INC)

$R_{15}$ = correlation between food away from home and family income (R_FA_INC).

Table 7 presents the population parameters and the average relbiases of the sample estimates of $\theta$. As in other similar empirical studies, e.g. Frankel (1971), the relbias is relatively small for the ratio estimates but not for the correlation coefficients. On average over the 12 nonlinear ratio type estimates, the relbias consistently decreases as the sample size increases.

The variation among the 1000 sample estimates of $\theta$, $\sum_{i=1}^{1000}(\hat{\theta}_i - \bar{\hat{\theta}})^2/999$, provides an empirical estimate of the var $\hat{\theta}$. When the Monte Carlo sampling variances are compared to the second order Taylor series variances discussed in section III, the two estimates are within 10% of each other in 8 of the 13 cases for the smallest sample size. For the largest sample size, the two estimates are within 10% of each other in all but one case.

For each of the 1000 samples, three sample sizes, and 15 parameters, random group and BRR variance estimates have been computed. Due to budget restrictions, jackknife variance estimates have been delayed until next fiscal year. Although a comparison of the variance estimators using these empirical estimates of variance does not show the same clear relationships as table 6 due to the noise in the data, they are useful in investigating the performance of the variance estimators with respect to confidence intervals for $\theta$.

If $\hat{\theta}$ is a normally distributed random variable and $\hat{V}(\hat{\theta})$ is a consistent estimator of var $\hat{\theta}$, then $(\hat{\theta} - \theta)/[\hat{V}(\hat{\theta})]^{1/2}$ has a standard normal distribution. Figures 3, 4 and 5 present the cumulative distribution functions of the t values computed as $(\hat{\theta} - \theta)/[\hat{V}(\hat{\theta})]^{1/2}$ for different choices of a variance estimator for each of the 1000 samples. The five lines on each graph correspond to the normal distribution and the empirical t-distributions for the smallest and largest sample size where the estimate of variance is either the BRR estimator or the random group estimator with the maximum number of random groups considered (k = 2 if n = 120, k = 8 if n = 480).

Figure 3 for ground beef is representative of the ratio estimation parameters when the numerator is a function of a variable from a skewed population and the denominator is a function of a Bernoulli variable. None of the sample t-distributions crosses the normal distribution for this type parameter. Theoretically, five percent of the observed values should be less than -1.645 and five percent should be greater than 1.645. For ground beef, an average of 35 percent of the 1000 values are less than -1.645 and less than 2 percent are greater than 1.645. While one-half of the observed values should be on either size of zero, the median is almost one standard error less than zero. The t-distribution of the random group estimator, when the sample size is small and the number of groups is therefore limited, has an especially long negative tail indicating the BRR variance estimator would be a better choice. When the sample size is 24 per stratum for a total of 480, there does not appear to be any significant differences between the BRR and random group estimator with a fairly large number of groups.

One hypotheses for explaining the greater than expected number of t values at the lower end of the distribution is that it is due to the high correlation between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$. In the following table, the seven mean expenditure per CU variables are listed by increasing skewness of the expenditure variable along with the correlation between $\hat{\theta}$ for $n_h = 6$ and $\hat{V}_{BRR}(\hat{\theta})$.

| Parameter | Skewness of Numerator Population | Correlation Between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ |
|---|---|---|
| FOODAWAY | 4.0 | .53 |
| GASCOST | 5.0 | .59 |
| FOODHOME | 5.7 | .53 |
| FLOUR | 6.7 | .67 |
| CANDY | 8.6 | .66 |
| GRBEEF | 22.6 | .86 |
| EGGS | 40.2 | .84 |

When the numerator population is very skewed, $\hat{\theta}$ is negatively biased if an extreme value (see figure 1) is not included in the sample. At the same time, the estimate of variance is a significant underestimate. Ground beef, which has the highest correlation, has the poorest coverage ratio. For the ratio parameters of this type, the correlation between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ appears to be related to the skewness of the population.

Figure 4 is for the ratio of wage and salary income to total income. The sample t-distribution has a median at approximately zero. Excluding the random group estimator for the smallest sample size, approximately 5% of the t-values are less than -1.645 and almost 10% of the t-values are greater than 1.645.

The t-distribution of the correlation between food at home and family income is presented in Figure 5. As in a previous study by Mulry and Wolter (1981), the lower end of the distribution appears close to the normal; but instead of only 5% of the values being greater than 1.645, more than 10% are greater. The median t-value is greater than zero.

WAGE_INC and R_FH_INC, which are examples of two different types of estimators, have negative correlations. Of the 15 parameters studied, only the correlation coefficients and WAGE_INC have negative correlations and only these three parameters with a negative correlation between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ have t-distributions with a heavier upper tail than lower tail.

Figures 6, 7 and 8 show the relationship between the number of random groups and the t-distribution for the largest sample size 480. RG3 refers to two random groups, each of size 12 per stratum. RG2 has four random groups of size 6 per stratum and RG1 has eight random groups of size 3 per stratum. For ground beef, the number of random groups has a significant effect on the lower tail and little effect on the median or upper tail. For the other variables, the effect is more symmetric. As the number of random groups increases, the distribution of the t-values approaches normality. Therefore, although the bias of the random group variance estimator is reduced as the size of the groups increases and the number decreases, a larger number of random groups is better with respect to coverage properties.

## V. CONCLUSION

The results of this study indicate the bias of the variance estimators studied is relatively small. If one assumes the variance estimators can be accurately approximated with a second order Taylor series, the random group, jackknife and BRR variance estimators are all positively biased when the estimator of interest is a ratio estimator where the numerator is not a subset of the denominator. If the ratio estimator is a proportion, the variance estimators could be negatively biased. The bias of the random group variance estimator decreases as the number of the random groups decreases or the size of the groups increases. But over all the parameters, sample sizes and variance estimators studied, the maximum relbias is only 9 percent. The Monte Carlo results support the conclusion that the relbias is small.

On the other hand, the variance of the variance estimators is not insignificant, and the normal-theory confidence intervals do not always have the desired coverage probabilities when the estimator is of a ratio type or a correlation coefficient. For the ratio estimator where the numerator is a function of a variable from a very skewed population, the sample may not include enough extreme values if the effective sample size is small. Consequently, not only may $\hat{\theta}$ be small and negatively biased, but $\hat{V}(\hat{\theta})$ may be significantly smaller yielding large negative t-values. These situations are indicated by a large correlation between $\hat{\theta}$ and $\hat{V}(\hat{\theta})$. When $\hat{\theta}$ and $\hat{V}(\hat{\theta})$ are negatively correlated, the t-distribution has a heavy upper tail. Users should be warned that the construction of confidence intervals and tests of hypothesis assuming normality may not be appropriate in these situations. As shown in the Mulry and Wolter paper, confidence intervals based upon transformations may be better.

When the effective sample size is small, the balanced repeated replication variance estimator is a better choice than the random group estimator with only two random groups. As the sample size increases allowing more random groups, the difference in these two variance estimators appears to be minimal.

**BIBLIOGRAPHY**

Bean, J.A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. Vital and Health Statistics, Series 2, No. 65.

Dippo, C.S. (1981). "Variance estimation for nonlinear estimators based upon stratified samples from finite populations." Ph.D. disseration, George Washington Unversity.

Frankel, M.R. (1971). Inference from survey samples: an empirical investigation. Ann Arbor. Institute for Social Research, University of Michigan.

Hajek, J. (1980). Limiting distributions in simple random sampling from a finite population. Periodica Mathematica. 5:361-374.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. Journal of the American Statistical Association. 77:89-96.

Jones, H.L. (1974). Jackknife estimation of functions of stratum means. Biometrika. 61:343-348.

Koop, J.C. (1968). An exercise in ratio estimation. The American Statistician. 22:29-30.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: properties of the linearization, jackknife, and balanced repeated replication methods. Annals of Statistics. 9:1010-1019.

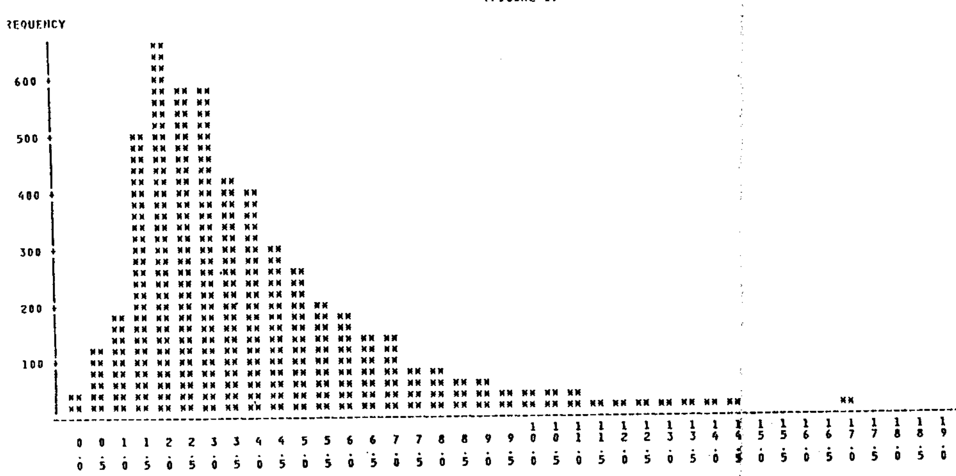
Mulry, M., and Wolter, K.M. (1981). The effect of Fisher's t-transformation in confidence intervals for the correlation coefficient. 1981 Proceedings of the Section on Survey Research Methods.

Sukhatme, P.V., and Sukhatme, B.V. (1970). Sampling theory of surveys with applications. Ames: Iowa State University Press.

Rao, J.N.K., and Wu, C.F.J. (1983). Inferences from stratified samples: second order analysis of three methods for nonlinear statistics. Technical Report Series of the Laboratory for Research in Statistics and Probability, 7.

Wolter, K. (1980). "Introduction to variance estimators." (Typewritten)

POPULATION FREQUENCY DISTRIBUTION
(FIGURE 1)




POPULATION FREQUENCY DISTRIBUTION
(FIGURE 2)

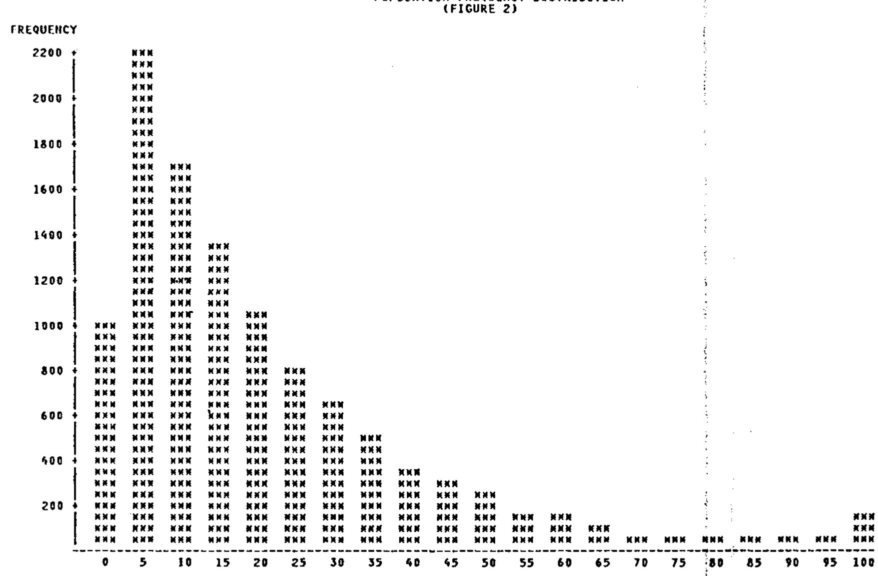


Figure 3

GROUNDBEEF

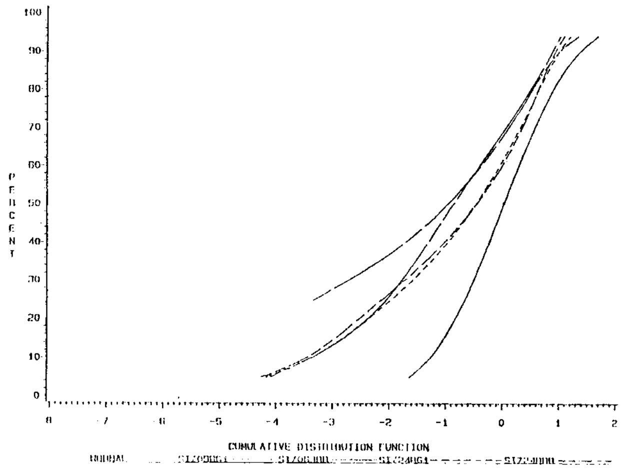
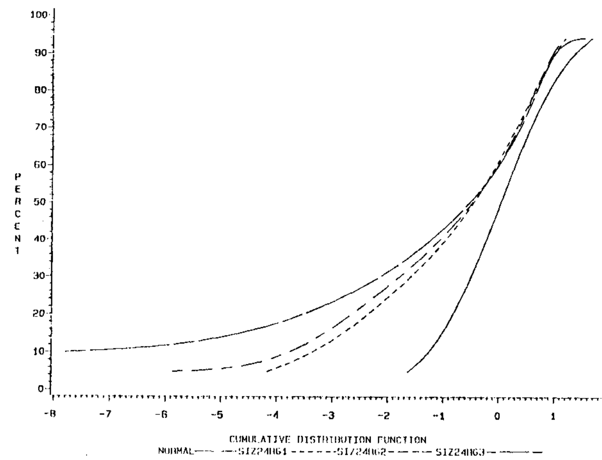


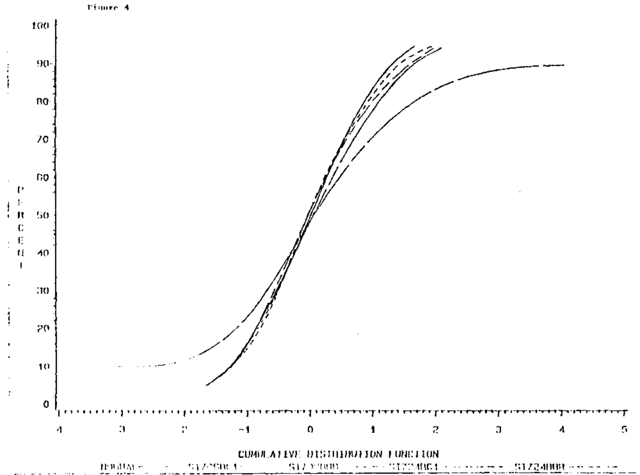

Figure 6

GROUNDBEEF



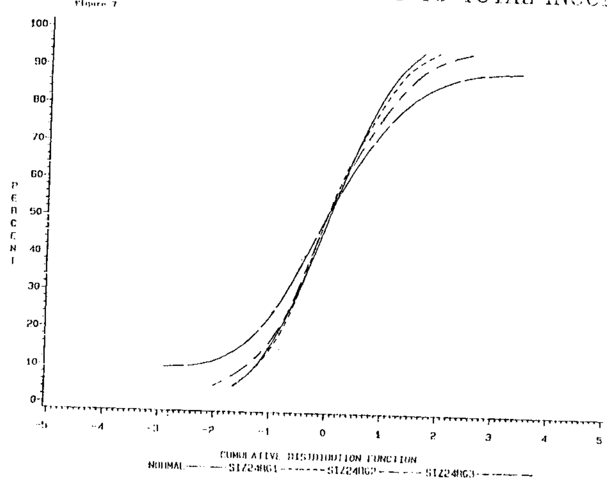

RATIO OF WAGES AND SALARIES TO TOTAL INCOME

Figure 4



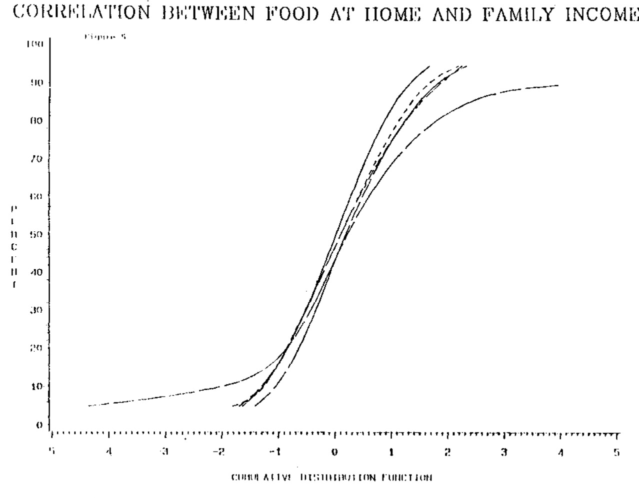

RATIO OF WAGES AND SALARIES TO TOTAL INCOME

Figure 7




CORRELATION BETWEEN FOOD AT HOME AND FAMILY INCOME

Figure 5




CORRELATION BETWEEN FOOD AT HOME AND FAMILY INCOME

Figure 8

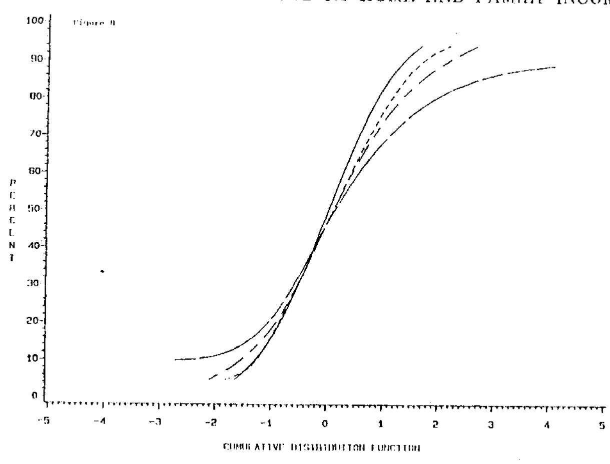


Table 1. The Bias of the Variance Estimators Assuming a Taylor Series Expansion

| | $F^{(1)}(\bar{X}_{rh})F^{(1)}(\bar{X}_{sh})$ | $F^{(1)}(\bar{X}_{rh})F^{(2)}(\bar{X}_{sh}\bar{X}_{th})$ |
|---|---|---|
| Bias $\hat{V}_1'$ | 0 | $\frac{(1-\lambda)^{3/2}}{k-1}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}) - E(\bar{u}_{rh}\bar{u}_{sh\alpha}\bar{u}_{th\alpha})\big) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})$ |
| Bias $\hat{V}_3$ | 0 | $c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hi)}\bar{u}_{t(hi)}) - c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hj)}\bar{u}_{t(hj)}) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})$ |
| Bias $\hat{V}_5'$ | 0 | $(1-\lambda)^{3/2}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th})-E(\bar{u}_{rh\alpha}\bar{u}_{sh\beta}\bar{u}_{th})\big) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th})$ |

$$c_h = \frac{(N_h-n_h)(n_h-1)^2}{n_h N_h}$$

Table 1. The Bias of the Variance Estimators Assuming a Taylor Series Expansion (Continued)

| | $F^{(1)}(\bar{X}_{rh})F^{(3)}(\bar{X}_{sh}\bar{X}_{th}\bar{X}_{zh})$ | $F^{(1)}(\bar{X}_{rh})F^{(3)}(\bar{X}_{sh}\bar{X}_{th'}\bar{X}_{zh'})$ | $F^{(2)}(\bar{X}_{rh}\bar{X}_{sh})F^{(2)}(\bar{X}_{th}\bar{X}_{zh})$ | $F^{(2)}(\bar{X}_{rh}\bar{X}_{sh'})F^{(2)}(\bar{X}_{th}\bar{X}_{zh'})$ |
|---|---|---|---|---|
| Bias $\hat{V}_1$ | $\frac{(1-\lambda)^2}{3(k-1)}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}\bar{u}_{zh\alpha}) - E(\bar{u}_{rh}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}\bar{u}_{zh\alpha})\big) - \frac{1}{3}E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh})$ | $\frac{(1-\lambda)^2}{k-1}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha})E(\bar{u}_{th'\alpha}\bar{u}_{zh'\alpha}) - E(\bar{u}_{rh}\bar{u}_{sh\alpha})E(\bar{u}_{th'\alpha}\bar{u}_{zh'\alpha})\big) - E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th'}\bar{u}_{zh'})$ | $\frac{(1-\lambda)^2}{4k}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}\bar{u}_{zh\alpha}) - E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\beta}\bar{u}_{zh\beta})\big) - \frac{1}{4}E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh}) + \frac{1}{4}E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th}\bar{u}_{zh})$ | $\frac{(1-\lambda)^2}{2k}\big(E(\bar{u}_{rh\alpha}\bar{u}_{th\alpha})E(\bar{u}_{sh'\alpha}\bar{u}_{zh'\alpha}) - E(\bar{u}_{rh\alpha}\bar{u}_{th\beta})E(\bar{u}_{sh'\alpha}\bar{u}_{zh'\beta})\big) - \frac{1}{2}E(\bar{u}_{rh}\bar{u}_{th})E(\bar{u}_{sh'}\bar{u}_{zh'})$ |
| Bias $\hat{V}_3$ | $\frac{1}{3}\big(c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hi)}\bar{u}_{t(hi)}\bar{u}_{z(hi)}) - c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hj)}\bar{u}_{t(hj)}\bar{u}_{z(hj)}) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh})\big)$ | $c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hi)})E(\bar{u}_{th'}\bar{u}_{zh'}) - c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hj)})E(\bar{u}_{th'}\bar{u}_{zh'}) - E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th'}\bar{u}_{zh'})$ | $\frac{1}{4}\big(c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hi)}\bar{u}_{t(hi)}\bar{u}_{z(hi)}) - c_h E(\bar{u}_{r(hi)}\bar{u}_{s(hi)}\bar{u}_{t(hj)}\bar{u}_{z(hj)}) - E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh}) + E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th}\bar{u}_{zh})\big)$ | $c_h E(\bar{u}_{r(hi)}\bar{u}_{t(hi)})E(\bar{u}_{sh'}\bar{u}_{zh'}) - c_h E(\bar{u}_{r(hi)}\bar{u}_{t(hj)})E(\bar{u}_{sh'}\bar{u}_{zh'}) - \frac{1}{2}E(\bar{u}_{rh}\bar{u}_{th})E(\bar{u}_{sh'}\bar{u}_{zh'})$ |
| Bias $\hat{V}_5$ | $\frac{(1-\lambda)^2}{24}\big[E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}\bar{u}_{zh\alpha}) + 3E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\beta}\bar{u}_{zh\beta}) - 4E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th\alpha}\bar{u}_{zh\beta})\big] + \frac{(1-\lambda)^2}{2}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha}\bar{u}_{th}\bar{u}_{zh}) - E(\bar{u}_{rh\alpha}\bar{u}_{sh\beta}\bar{u}_{th}\bar{u}_{zh})\big) - \frac{1}{3}E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh})$ | $\frac{(1-\lambda)^2}{4}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha})E(\bar{u}_{th'\alpha}\bar{u}_{zh'\alpha}) + E(\bar{u}_{rh\alpha}\bar{u}_{sh\beta})E(\bar{u}_{th'\alpha}\bar{u}_{zh'\beta}) - 2E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha})E(\bar{u}_{th'\alpha}\bar{u}_{zh'\beta})\big) + \frac{(1-\lambda)^2}{2}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh\alpha})E(\bar{u}_{th'}\bar{u}_{zh'}) - E(\bar{u}_{rh\alpha}\bar{u}_{sh\beta})E(\bar{u}_{th'}\bar{u}_{zh'})\big) - E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th'}\bar{u}_{zh'})$ | $\frac{(1-\lambda)^2}{2}\big(E(\bar{u}_{rh\alpha}\bar{u}_{sh}\bar{u}_{th\alpha}\bar{u}_{zh}) - E(\bar{u}_{rh\alpha}\bar{u}_{sh}\bar{u}_{th\beta}\bar{u}_{zh})\big) - \frac{1}{4}\big(E(\bar{u}_{rh}\bar{u}_{sh}\bar{u}_{th}\bar{u}_{zh}) - E(\bar{u}_{rh}\bar{u}_{sh})E(\bar{u}_{th}\bar{u}_{zh})\big)$ | $\frac{(1-\lambda)^2}{8}\big(E(\bar{u}_{rh\alpha}\bar{u}_{th\alpha})E(\bar{u}_{sh'\alpha}\bar{u}_{zh'\alpha}) + E(\bar{u}_{rh\alpha}\bar{u}_{th\beta})E(\bar{u}_{sh'\alpha}\bar{u}_{zh'\beta}) - 2E(\bar{u}_{rh\alpha}\bar{u}_{th\alpha})E(\bar{u}_{sh'\alpha}\bar{u}_{zh'\beta}) + 4E(\bar{u}_{rh\alpha}\bar{u}_{th\alpha})E(\bar{u}_{sh'}\bar{u}_{zh'}) - 4E(\bar{u}_{rh\alpha}\bar{u}_{th\beta})E(\bar{u}_{sh'}\bar{u}_{zh'})\big) - \frac{1}{2}E(\bar{u}_{rh}\bar{u}_{th})E(\bar{u}_{sh'}\bar{u}_{zh'})$ |

TABLE 2
DISTRIBUTION STATISTICS FOR POPULATION VALUES

| PARAMETER | N | MEAN | STD | SKEWNESS | KURTOSIS |
|---|---|---|---|---|---|
| FLOUR | 1990 | 1.07 | 1.3294 | 6.6883 | 97.7728 |
| CANDY | 3800 | 2.10 | 3.4714 | 8.5827 | 121.8822 |
| GRBEEF | 5415 | 4.91 | 11.6441 | 22.6241 | 680.2998 |
| EGGS | 6668 | 1.47 | 1.7851 | 40.1976 | 2508.2701 |
| GASOLINE | 10071 | 25.33 | 21.4416 | 4.9581 | 53.5673 |
| FOODAWAY | 10934 | 21.36 | 23.3353 | 3.9731 | 37.1022 |
| WAGEX | 11268 | 18790.88 | 14656.3233 | 2.0601 | 15.2658 |
| FOODHOME | 13113 | 39.17 | 38.3054 | 5.6814 | 91.8096 |
| FINCBEFX | 14360 | 18503.50 | 15674.2341 | 2.6230 | 24.1462 |

TABLE 3
AVERAGE STRATUM POPULATION MOMENTS

| AMETER | M2 $(X_1X_1)$* | M2 $(X_2X_2)$ | M2 $(X_1X_2)$ | M3 $(X_1X_1X_1)$ | M3 $(X_1X_2X_2)$ | M3 $(X_1X_1X_2)$ | M3 $(X_2X_2X$ |
|---|---|---|---|---|---|---|---|
| DY | 4.0402 | 0.1935 | 0.4062 | 109.7989 | 0.1866 | 2.6708 | 0.0899 |
| S | 1.9692 | 0.2474 | 0.3618 | 105.3453 | 0.0256 | 0.8196 | 0.0177 |
| CBEFX | 2.3781 E08 | 0.0000 | 0.0000 | 9.6022 E12 | 0.0000 | 0.0000 | 0.0000 |
| UR | 0.3750 | 0.1162 | 0.1251 | 2.9231 | 0.0879 | 0.2940 | 0.0827 |
| DAWAY | 486.0245 | 0.1791 | 3.7475 | 42011.7899 | -1.9642 | 48.7938 | -0.0917 |
| DHOME | 1450.4602 | 0.0788 | 3.0627 | 299670.4069 | -2.4921 | 17.8264 | -0.0640 |
| COST | 450.6975 | 0.2078 | 5.2381 | 39407.0712 | -2.0720 | 40.3502 | -0.0818 |
| _VEHQ | 450.6975 | 1.1476 | 9.0153 | 39407.0712 | 5.5500 | 278.0115 | 2.2732 |
| EEF | 56.1506 | 0.2332 | 1.1455 | 13673.7427 | 0.2810 | 32.4989 | 0.0569 |
| M_CLF | 0.0246 | 0.9576 | 0.0225 | 0.0308 | 0.0330 | 0.0276 | 0.7975 |
| Q_FAM | 1.1476 | 0.1406 | 0.2435 | 2.2732 | -0.1557 | -0.1531 | -0.0894 |
| E_CAP | 2.2152 E08 | 0.8322 | 8644.1466 | 6.0509 E12 | 3088.1557 | 6.3185 E07 | 0.5935 |
| E_INC | 2.2152 E08 | 2.3782 E08 | 1.9651 E08 | 6.0509 E12 | 6.2841 E12 | 6.0931 E12 | 9.6022 E12 |

AVERAGE STRATUM DERIVATIVES FOR THE POPULATION

| PARAMETER | $F^{(1)}(\bar X_{1h})$ | $F^{(1)}(\bar X_{2h})$ | $F^{(2)}(\bar X_{1h}\bar X_{2h})$ | $F^{(2)}(\bar X_{2h}\bar X_{2h})$ | $F^{(3)}(\bar X_{1h}\bar X_{2h}\bar X_{2h})$ | $F^{(3)}(\bar X_{2h}\bar X_{2h}\bar X_{2h})$ |
|---|---|---|---|---|---|---|
| CANDY | 0.1889 | -0.3972 | -0.0357 | 0.1503 | 0.0135 | -0.0854 |
| EGGS | 0.1077 | -0.1578 | -0.0116 | 0.0340 | 0.0025 | -0.0110 |
| FINCBEFX | 0.0500 | -925.1747 | -0.0025 | 92.6449 | 0.0002 | -13.9348 |
| FLOUR | 0.3652 | -0.3959 | -0.1335 | 0.2896 | 0.0978 | -0.3182 |
| FOODAWAY | 0.0656 | -1.4026 | -0.0043 | 0.1844 | 0.0005 | -0.0364 |
| FOODHOME | 0.0547 | -2.1446 | -0.0030 | 0.2351 | 0.0003 | -0.0387 |
| GASCOST | 0.0712 | -1.8057 | -0.0050 | 0.2578 | 0.0007 | -0.0552 |
| GAS_VEHQ | 0.0344 | -0.4206 | -0.0011 | 0.0289 | 0.0000 | -0.0030 |
| GRBEEF | 0.1326 | -0.6509 | -0.0176 | 0.1728 | 0.0046 | -0.0689 |
| UNEM_CLF | 0.0358 | -0.0005 | -0.0012 | 0.0000 | 0.0000 | -0.0000 |
| VEHQ_FAM | 0.0603 | -0.1059 | -0.0036 | 0.0128 | 0.0004 | -0.0023 |
| WAGE_CAP | 0.0410 | -497.6624 | -0.0016 | 40.9444 | 0.0001 | -5.0598 |
| WAGE_INC | 0.0000 | -0.0000 | -0.0000 | 0.0000 | 0.0000 | -0.0000 |

$M_2(X_1X_1) = \frac{1}{n_h}\Sigma(X_{1hi}-\bar X_{1h})^2$, $M_3(X_1X_2X_2) = \frac{1}{n_h}\Sigma(X_{1hi}-\bar X_{1h})(X_{2hi}-\bar X_{2h})^2$, etc.

TABLE 4
COMPONENT PERCENTAGES FOR VARIANCE OF THETA HAT

| PARAMETER | TOTAL | PROP'N FIRST ORDER | PROP'N SECOND ORDER |
|---|---|---|---|
| | SIZE6 | | |
| CANDY | 0.384395 | 0.97804 | 0.0220 |
| EGGS | 0.058610 | 0.99194 | 0.0081 |
| FINCBEFX | 1965738.083960 | 1.00000 | 0.0000 |
| FLOUR | 0.112399 | 0.95125 | 0.0487 |
| FOODAWAY | 5.853407 | 0.99733 | 0.0027 |
| FOODHOME | 13.223444 | 0.99925 | 0.0008 |
| GAS_VEHQ | 1.590538 | 0.99862 | 0.0014 |
| GASCOST | 5.454668 | 0.99660 | 0.0034 |
| GRBEEF | 3.034593 | 0.98736 | 0.0126 |
| UNEM_CLF | 0.000104 | 1.00152 | -0.0015 |
| VEHQ_FAM | 0.008743 | 0.99824 | 0.0018 |
| WAGE_CAP | 754106.541680 | 0.99351 | 0.0065 |
| WAGE_INC | 0.001416 | 1.01517 | -0.0152 |
| | SIZE12 | | |
| CANDY | 0.188521 | 0.98871 | 0.0113 |
| EGGS | 0.028949 | 0.99586 | 0.0041 |
| FINCBEFX | 974587.111325 | 1.00000 | 0.0000 |
| FLOUR | 0.054396 | 0.97455 | 0.0255 |
| FOODAWAY | 2.898276 | 0.99864 | 0.0014 |
| FOODHOME | 6.553612 | 0.99962 | 0.0004 |
| GAS_VEHQ | 0.788117 | 0.99924 | 0.0008 |
| GASCOST | 2.700002 | 0.99827 | 0.0017 |
| GRBEEF | 1.495301 | 0.99352 | 0.0065 |
| UNEM_CLF | 0.000052 | 1.00068 | -0.0007 |
| VEHQ_FAM | 0.004331 | 0.99910 | 0.0009 |
| WAGE_CAP | 372674.196766 | 0.99672 | 0.0033 |
| WAGE_INC | 0.000708 | 1.00729 | -0.0073 |
| | SIZE24 | | |
| CANDY | 0.092145 | 0.99420 | 0.0058 |
| EGGS | 0.014206 | 0.99787 | 0.0021 |
| FINCBEFX | 479011.625007 | 1.00000 | 0.0000 |
| FLOUR | 0.026405 | 0.98684 | 0.0132 |
| FOODAWAY | 1.423615 | 0.99931 | 0.0007 |
| FOODHOME | 3.220542 | 0.99980 | 0.0002 |
| GAS_VEHQ | 0.387269 | 0.99956 | 0.0004 |
| GASCOST | 1.326104 | 0.99911 | 0.0009 |
| GRBEEF | 0.732740 | 0.99668 | 0.0033 |
| UNEM_CLF | 0.000025 | 1.00027 | -0.0003 |
| VEHQ_FAM | 0.002128 | 0.99954 | 0.0005 |
| WAGE_CAP | 182872.231948 | 0.99835 | 0.0017 |
| WAGE_INC | 0.000349 | 1.00342 | -0.0034 |

Table 5. Comparison of the Second Order Component of the Var $\hat\theta$ with Some Derivatives ($n_h$=6)

| | Percent Second Order | $X_2$ | $F^{(1)}(\bar X_1)$ $=N^2X_2^{-1}$ | $F^{(2)}(\bar X_1\bar X_2)$ $=-N^2X_2^{-2}$ | $F^{(3)}(\bar X_1\bar X_2^2)$ $=2N^3X_2^{-3}$ |
|---|---|---|---|---|---|
| WAGE_INC | -1.52 | | 0 | 0 | 0 |
| UNEM_CLF | - .15 | | .04 | -.001 | .00009 |
| FOODHOME | .08 | 13113 | .05 | -.003 | .00033 |
| GAS_VEHQ | .14 | | .03 | -.001 | .00008 |
| VEHQ_FAM | .18 | | .06 | -.004 | .00044 |
| FOODAWAY | .27 | 10934 | .07 | -.004 | .00057 |
| GASCOST | .34 | 10071 | .07 | -.005 | .00073 |
| WAGE_CAP | .65 | | .04 | -.002 | .00014 |
| EGGS | .81 | 6668 | .11 | -.012 | .00251 |
| GRBEEF | 1.26 | 5415 | .13 | -.018 | .00468 |
| CANDY | 2.20 | 3800 | .19 | -.036 | .01356 |
| FLOUR | 4.87 | 1990 | .37 | -.134 | .09782 |