ON THE VARIANCE OF A TOTAL ESTIMATE OVER A SUBPOPULATION

Shien S. Perng, Internal Revenue Service

1. Introduction

In sampling surveys it often occurs that a sample has to be selected from a larger population which contains units of interest and other unwanted units. Thus, the population of interest is a subpopulation of the population from which the sample is selected. Several methods have been introduced to estimate the total of a quantitative variable over such a subpopulation. The choice of methods depends on available information about the subpopulation. Cochran (1977), pages 35-38, presents three methods. The first one is applicable when M, the total number of units in the subpopulation, is known. In this case an estimate of the total is M times of the sample mean of units which fall in the subpopulation. The second one is applicable when the total of the quantitative variable over the entire population is known. In this case a ratio estimate may be employed. The third one is applicable when neither M nor the total of the quantitative variable is available. An estimate is the sample total over the subpopulation times the reciprocal of the sample fraction. This method is sometimes referred to as the simple expansion method. See Jones and Coopersmith (1976). Kish (1967), pages 204 and 438, mentions a ratio estimate using an auxiliary total over the subpopulation. Jones and Coopersmith (1976) studies a ratio estimate where the auxiliary total over the subpopulation is estimated. Neter and Loebbecke (1975 and 1977) and Cox and Snell (1979) study methods of estimating total error in accounting population, including an equivalence of the simple expansion method.

This paper concerns the simple expansion method. This method is especially useful in a large scale survey where a large volume of tables are produced. The estimation of each table cell may be considered as an estimation over a subpopulation where the portion of the entire population satisfying the table cell definition is the subpopulation of interest. In this case, a simple expansion method is the easiest one to use and is often the only feasible one due to various reasons.

Large scale surveys are usually multipurposed. Data are collected for many variables. There are rarely variables whose totals over all table cells are available and which are highly correlated to the survey variables and hence can be used as an auxiliary variable to reduce the variance of the estimate. The total numbers of units in the subpopulations (table cells) are often not available or too costly to obtain. Besides, these numbers are very often part of the items to be estimated. Sometimes the table cells are not well defined until a very late stage of the survey. In some circumstances, the census data may be used. But this may be rare due to different categorization, or data being out of date. Programming for the tabulation may further restrict the estimation methodology. Usage of simple expansion method could simplify the programming effort considerably. This is especially true when the auxiliary data are only partially available.

For the Statistics of Income and Taxpayer Compliance Measurement Program reports compiled by the Internal Revenue Service, the tabulation is done for line items on the tax forms according to filing status, income classes, return types, etc. Due to the complexity of the tabulation, it is hardly possible to use methodologies requiring auxiliary variables to improve precision of the estimate. The simple expansion method has been basically used for these reports.

The simple expansion method has been discussed in many sampling textbooks. Sukhatme and Sukhatme (1970), Cochran (1977), Hansen, Hurwitz and Madow (1953b), and Kish (1967), among others, present the estimation method and its variance estimation. Sukhatme and Sukhatme (1970), pages 36-37, gives an example of sample size computation using the coefficient of variation. Kish (1967), pages 434-436, and Cochran (1977), page 38, compare the variance of simple expansion with the variance of an estimate when the subpopulation count is known. Netter and Loebbecke (1975) compare several estimation methods in accounting context, including simple expansion (mean-per-unit) and stratified simple expansion methods.

The simple expansion estimate is obtained by expanding the sample total of units falling in the subpopulation by the reciprocal of the sample fraction. When the proportion of the subpopulation is close to 1, most of the sample units fall in the subpopulation. Hence the characteristics of the estimate are expected to be close to those of an estimate for the entire population. But when the proportion of the subpopulation is small, the number of sample units usable for the estimation may be few. In this case, the characteristics of the estimate may be completely different. Our uncertainty about the characteristics of the subpopulation estimate motivates the present investigation. We study how the variance of such estimate depends on the proportion and how the variance of its variance estimate behaves. We focus on the impact of the number of usable units on the reliability of the estimate when the proportion of the subpopulation is small.

Section 2 examines the behavior of the variance of the simple expansion estimate. The variance is found to be approximately proportional to the proportion of the subpopulation, when the coefficient of variation per unit in the subpopulation is not less than 1. Comparing to the estimate when the subpopulation count is available, the simple expansion estimate performs well when the coefficient of variation of the subpopulation is larger than or equal to 2.

Section 3 considers variance analyses. Two decompositions of the variance are discussed. The two components in the first decomposition are due to the variation of unit measures in the subpopulation and to the variation of the number of sample units falling in the subpopulation, respectively. To describe the second decomposition, recall that the unit measures for units not in the subpopulation are set to 0. The second decomposition splits the variance into a component which is contributed by the deviation of unit measures from the overall average and a component which is contributed by the departure of the overall average from 0, the designated value for units not in the subpopulation. In both decompositions, the second

component contributes most of the variance when the coefficient of variation per unit in the subpopulation is very small.

Section 4 examines the variance of the variance estimate. The variance is graphed under several different configurations of the subpopulation. It is approximately linearly increasing with the magnitude of the proportion of the subpopulation when the coefficient of variation per unit in the subpopulation is not less than 1.

Some concluding comments are offered in Section 5.

2. Simple Expansion Estimate and its Variance

Let C be a subpopulation of the population from which a simple random sample has been taken. Let

- N = total number of units in the entire population,
- M = total number of units in the subpopulation C,
- n = total sample size,
- m = number of sample units falling in subpopulation C,
- y'_i = a unit measure in the entire population, where y'_i is set to 0 if the unit is not in subpopulation C,
- y_i = a unit measure in subpopulation C,
- Y = total of y-variable in subpopulation C,
- $\bar{Y}$ = Y/M.

Without loss of generality, we shall write y'_1, y'_2, ...y'_n, as units of the sample and y_1, y_2,..., y_m as units of the sample which fall in subpopulation C. Similar notations are used for units in the population. Write $\bar{y} = \sum_1^m y_i/m$ and $\bar{y}' = \sum_1^n y'_i/n$. The simple expansion estimate of Y is given by

$$\hat{Y} = \frac{N}{n}\sum_1^m y_i = \frac{N}{n} m\bar{y} = N\bar{y}'. \qquad (2.1)$$

Note that m is a random variable which has a hypergeometric distribution with mean and variance given by

$$E[m] = nP, \qquad (2.2)$$

$$V[m] = nP(1-P)\frac{N-n}{N-1}, \qquad (2.3)$$

where $P = M/N$ is the proportion of the subpopulation C.

From (2.2), $\hat{Y}$ is unbiased;

$$E(\hat{Y}) = E[E(\hat{Y}|m)] = \frac{N}{n}E[m]\bar{Y} = M\bar{Y} = Y.$$

The variance of $\hat{Y}$ is

$$V(\hat{Y}) = N^2V(\bar{y}') = \frac{N^2}{n}(1-n/N)S^2_{y'}, \qquad (2.4)$$

where

$$S^2_{y'} = \frac{1}{N-1}\sum_1^N (y'_i - \bar{Y}')^2 = \frac{1}{N-1}\left(\sum_1^M y_i^2 - \frac{1}{N}\left(\sum_1^M y_i\right)^2\right).$$

By using $(N-1)S^2_{y'} = \sum_1^M y_i^2 - \frac{1}{M}\left(\sum_1^M y_i\right)^2 + \left(\frac{1}{M} - \frac{1}{N}\right)\left(\sum_1^M y_i\right)^2$, we can rewrite (2.4) as

$$V(\hat{Y}) = \frac{N^2}{n}\left(1 - \frac{n}{N}\right)\left(\frac{M-1}{N-1}S^2_y + \frac{M}{N-1}\left(1 - \frac{M}{N}\right)\bar{Y}^2\right) \qquad (2.5)$$

$$\doteq \frac{N^2}{n}\left(1-\frac{n}{N}\right)P\left(S^2_y + (1-P)\bar{Y}^2\right), \qquad (2.5')$$

where $S^2_y = \frac{1}{M-1}\left(\sum_1^M y_i^2 - \frac{1}{M}\left(\sum_1^M y_i\right)^2\right)$. In the last approximation, 1/N is assumed to be negligible. This form of the variance has been given in many sampling textbooks. See, e.g. Sukhatme and Sukhatme (1970) page 35, or Cochran (1977), page 38. It expresses $V(\hat{Y})$ in terms of the mean and variance in the subpopulation C.

The variance $V(\hat{Y})$ varies with P. The manner of variation depends on the distribution of the y variable in the subpopulation C and, in particular, on S^2_y and $\bar{Y}^2$. To examine such variation, let G(P) be the right-handed side of (2.5') and $R_1(P) = G(P)/G(1)$. Then

$$R_1(P) = P(1 + (1-P)/C^2), \qquad (2.6)$$

where $C = S_y/\bar{Y}$, the coefficient of variation of y per unit in subpopulation C. (Note that C is used to denote both the coefficient of variation and the subpopulation. The meaning will be clear from the context.) For any fixed $C \neq 0$, $R_1(P)$ is a bounded function of P for $0 \leq P \leq 1$. Since the derivative

$$R'_1(P) = (C^2 + 1-2P)/C^2,$$

$R_1(P)$ has a maximum $1/4(1 + 1/C^2)$ at $P = (C^2 + 1)/2$ when $C^2 < 1$ and has a maximum 1 at P=1 when $C^2 \geq 1$. However, considering R_1 as a function of C, it is not bounded above for C close to 0. This will not present problems since in most social and economical surveys, the value of C is usually not less than 1/4. See Hansen, Hurwitz and Madow (1953a), pages 138-148, for examples of values for C in applications.

Figure 1 shows plots of $R_1(P)$ for C = 1/4, 1/2, 1, 2, 4. $R_1(P)$ is very much proportional to P when $C \geq 1$. For $C < 1/2$, $R_1(P)$ is less stable.

Since $V(\hat{Y}) \doteq \frac{N^2}{n}\left(1 - \frac{n}{N}\right)S^2_y R_1(P)$, for fixed N, n, and S^2_y the behavior of $V(\hat{Y})$ as a function of P is similar to that of $R_1(P)$. It is bounded for most practical values of C. For $C \geq 1$, $V(\hat{Y})$ is very much proportional to P. It is especially worthwhile to note that for small value of P, $V(\hat{Y})$ is relatively small compared to its value when P = 1. While this does not really guarantee that the estimate $\hat{Y}$ is reliabile, it is somewhat comforting to know that its variance does not go wild when P is small. We discuss the estimate of $V(\hat{Y})$ and the variance of such estimate in Section 4.

The square of the coefficient of variation of Y is obtained by dividing $Y^2 = (M\bar{Y})^2$ into (2.5) or (2.5'). Thus

$$C^2(\hat{Y}) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\frac{N^2}{M^2}\left(\frac{M-1}{N-1}C^2 + \frac{M}{N-1}\left(1 - \frac{M}{N}\right)\right) \qquad (2.7)$$

$$\doteq \frac{1}{n}\left(1 - \frac{n}{N}\right)(C^2 + 1-P)/P \qquad (2.7')$$

The value of $C^2(\hat{Y})$ is not bounded for any fixed C, n and N; it approaches infinite as P approaches 0. For small P, it takes a large sample size n to bring down the $C(\hat{Y})$ to desirable level. This is in contrast to the estimation of the total of a

quantitative variable in the entire population where $C^2(\hat{Y}) = \frac{1}{n}(1 - \frac{n}{N})C^2$. The situation is similar to the estimation of the proportion or total of a subpopulation where $C^2(\hat{M}) = \frac{1}{n}(1 - \frac{n}{N})\frac{1-P}{P}$.

Thus, one should be cautious when P is small if a small $C(\hat{Y})$ is desired.

It is noted that when P is small the coefficient of variation $C(\hat{Y})$ is large, but the standard error $s.e.(\hat{Y})$ is small. The precision is good in absolute sense, but not in relative sense. In this case, the requirement of a small coefficient of variation is debatable since its magnitude does not really reflect the reliability of the estimate. The standard error may be a better choice as measure of sampling variability.

3. Variance Analysis

We consider two different decompositions of the variance $V(\hat{Y})$. First, the variance can also be derived as follows.

$$V(\hat{Y}) = E[V(\hat{Y}|m)] + V[E(\hat{Y}|m)]$$

$$= \frac{N^2}{n^2}E[m(1 - \frac{m}{M})]S_y^2 + \frac{N^2}{n^2}V[m]\bar{Y}^2$$

$$\doteq \frac{N^2}{n}(1 - \frac{n}{N})P\,S_y^2 + \frac{N^2}{n}(1 - \frac{n}{N})P(1 - P)\bar{Y}^2,$$

using (2.2) and (2.3). This expression is the same as (2.5) or (2.5'). This derivation is more informative. The first term is due to the variation of the y- variable in the subpopulation C and the second term to the variation of m. It is interesting to know how these two terms vary with P for different values of C. Let $R_3(P)$ be the ratio of the second term over the first. Then

$$R_3(P) = (1-P)/C^2 \qquad (3.1)$$

For fixed C, the ratio ranges from 0 to $1/C^2$ and decreases linearly. The contribution of the second term never exceeds a $(1/C^2)$th of the first. However, for small C, the contribution can be extremely large. Thus, if C is known to be small, the variance $V(\hat{Y})$ may be largely due to the variation of m, especially when P is small, In this case, the simple expansion method should be avoided. That is, it may earn a good payoff to screen off unwanted units in the population before sampling if it is possible, or in mass tabulation situation, fine table cells should be avoided. Figure 3 shows the graph of $R_3(P)$ for C = 1/4,1/2, 1, 2, 4. If C is large (C > 2, for example), $V(\hat{Y})$ is mostly contributed by the first component. In this case, simple expansion estimate is very effective compared to $M\bar{y}$.

We now consider a second decomposition of $V(\hat{Y})$. Note that for a unit y' selected from the entire population, y' is in subpopulation C with probability P and is not in subpopulation C with probability 1-P. Recall y' = y if the unit is in C and y' = 0 if the unit is not in C. Thus, the distribution function of y can be written as

$$F_{y'}(\ ) = PF_c(\ ) + (1-P)F_d(\ ),$$

where $F_c(\ )$ represent the (continuous) distribution function of y variable in subpopulation C and $F_d(\ )$ is a degenerated distribution function with all its probability mass on the point 0. From this we get

$$E(y') = PE_c(y') + (1-P)E_d(y') = P\bar{Y},$$

$$V(y') = PE_c(y' - P\bar{Y})^2 + (1-P)E_d(y' - P\bar{Y})^2,$$

where $E_c(\ )$ and $E_d(\ )$ are the expectations with respect to $F_c(\ )$ and $F_d(\ )$, respectively. Since $E_c(y' - P\bar{Y})^2 = E_c(y-\bar{Y})^2 + (\bar{Y}-P\bar{Y})^2 = S_y^2 +$

$(1-P)^2\bar{Y}^2$ and $E_d(y' - P\bar{Y})^2 = P^2\bar{Y}^2$, we have

$$S_{y'}^2 = V(y') = P(S_y^2 + (1-P)^2\bar{Y}^2) + (1-P)P^2\bar{Y}^2 \qquad (3.2)$$

Thus

$$V(\hat{Y}) = V(N\bar{y}') = \frac{N^2}{n}(1- \frac{n}{N})\,P(S_y^2 + (1-P)^2\,\bar{Y}^2)$$

$$+ \frac{N^2}{n}(1- \frac{n}{N})(1-P)P^2\bar{Y}^2 \qquad (3.3)$$

The first component represents the contribution of the deviation of y variable from $E(y') = P\bar{Y}$ and the second the contribution of the deviation of 0 from E(y'). This decomposition enables one to assess the contribution of designating y' = 0 for units not in subpopulation C to the variance $V(\hat{Y})$ of the estimate $\hat{Y}$. The contribution reaches its maximum 4/27 of $\bar{Y}^2$, when P = 2/3.
The ratio of these two components is

$$R_4(P) = (1-P)P/(C^2 + (1-P)^2).$$

Figure 4 shows the graphs of $R_4(P)$ for C = 1/4, 1/2, 1, 2, 4. For small C and moderately large P, the contribution of the second term may be extremely large. In this case, the variance $V(\hat{Y})$ is mainly due to the deviation of y' = 0 from $P\bar{Y}$, the expected value of y'.

4. Variance of the Variance Estimate

The accuracy and precision of the estimate of the variance $V(\hat{Y})$ have been a great concern when the obtained m in the sample is small, or when P is small. There are not many y_i in the sample which can be used for the estimation. In this case how reliable is the estimate of the variance? In this section, we examine the variance of the variance estimate. We study how the precision of the variance estimate depends on P.

From (2.4), $V(\hat{Y})$ may be estimated by

$$v(\hat{Y}) = \frac{N^2}{n}(1-\frac{n}{N})\,s_{y'}^2, \qquad (4.1)$$

where

$$s_{y'}^2 = \frac{1}{n-1}(\sum_1^m y_i^2 - \frac{1}{n}(\sum_1^m y_i)^2) \qquad (4.2)$$

Our main concern will be the variance of $s_{y'}^2$. For simplicity, we shall assume sampling with replace-

ment for the remainder of this section. Thus, the result will be only an approximation of the result when the sampling is without replacement. The approximation will be close when N is large and n/N is small (e.g. less than 0.05). Under the new assumption, we may rewrite (4.1)

$$v(\hat{Y}) = \frac{N^2}{n} s^2_{y'}. \tag{4.1}$$

The question is how reliable is $s^2_{y'}$ as an estimator of $S^2_{y'}$, or $s_{y'}$ as an estimator of $S_{y'}$?

It is clear that $s^2_{y'}$ is unbiased for $S^2_{y'}$. From (5.9) of Hansen, Hurwitz and Madow (1953b), page 101, we have the variance of $s^2_{y'}$,

$$V(s^2_{y'}) = 1/n(\mu_{y'4} - \frac{n-3}{n-1}\sigma^4_{y'}) \tag{4.3}$$

where

$$\mu_{y'4} = \frac{1}{N}\sum_1^N (y'_i - \bar{Y}')^4 \tag{4.4}$$

$$\sigma^2_{y'} = \frac{1}{N}\sum_1^N (y'_i - \bar{Y}')^2 \tag{4.5}$$

and $\bar{Y}' = \sum_1^N y'_i/N$. These formulas are expressed in terms of y'_i. It is more informative to express them in terms of moments in the subpopulation C. Let

$$\mu_j = \frac{1}{M}\sum_1^M (y_i - \bar{Y})^j, \quad j = 2, 3, 4. \tag{4.6}$$

Write $\sigma^2_y = \mu_2$. Then, by a straight forward algebraic manipulation, we obtain

$$\mu_{y'4} = P\{\mu_4 + 4(1-P)\bar{Y}\mu_3 + 6(1-P)^2\bar{Y}^2\sigma^2_y \tag{4.7}$$

$$+ (1-P)(1-3P+3P^2)\bar{Y}^4\}$$

$$\sigma^2_{y'} = P\sigma^2_y + P(1-P)\bar{Y}^2 \tag{4.8}$$

Substituting (4.7) and (4.8) into (4.3), we have

$$V(s^2_{y'}) = \frac{P}{n}\{\mu_4 + 4(1-P)\bar{Y}\mu_3 + 6(1-P)^2\sigma^2_y + (1-P)(1-3P+2P^2)\bar{Y}^4 - \frac{n-3}{n-1}P(\sigma^2_y + (1-P)\bar{Y}^2)^2\} \tag{4.9}$$

When P = 1, (4.9) is in the same form as (4.3). For large n, $(n-3)/(n-1) \doteq 1$. We shall employ this approximation.

We are interested in how $V(s^2_{y'})$ depends on P. For this purpose, we write

$$F(P) = nV(s^2_{y'})/\sigma^4_y$$

$$= P\{\gamma_2 + 4(1-P)\gamma_1/C + 2(1-P)(3-4P)/C^2 + (3-P) + (1-P)(1-2P)^2/C^4\} \tag{4.10}$$

where $\gamma_1 = \mu_3/\sigma^3_y$, $\gamma_2 = \mu_4/\sigma^4_y - 3$ and $C = \sigma_y/\bar{Y}$.

Note that γ_1 and γ_2 are the coefficients of skewness and kurtosis of the y variable in subpopulation C, respectively, and that C is the coefficient of variation. F(P) is a polynomial of degree 4 in P with F(0) = 0 and $F(1) = \gamma_2 + 2$. It has double peaks and a single valley for moderate values of γ_1, and γ_2. Figures 5(a) - 5(d) display F(P) for various combination of γ_1 and γ_2 and for C = 1/4, 1/2, 1, 2, 4. It is noted that for $C \geq 1$ and moderate value of γ_2, F(P) is almost linear in P. For $C < 1$, F(P) varies wildly. These properties carry over to $V(s^2_{y'})$. Thus, if $C < 1$, the variance $V(s^2_{y'})$ is large for P in the interval (0,1) and has maximums near P = 0.20 and 0.80. Therefore, the simple expansion method is not desirable when $C < 1$.

5. Concluding Remarks

This paper concerns a simple expansion method of estimating the total of a quantitative random variable over a subpopulation. This method is used when the units of interest are in a subpopulation of the population from which the sample is selected. In a simple survey, the sampling is designed in such a way that the subpopulation is as close to the sampled population as possible and the proportion of the subpopulation is very close to one. In this case, the simple expansion estimate should have similar characteristics as an estimate using the entire sample. The loss in precision of the simple expansion estimate is not expected to be large.

In some surveys, the result is post-stratified and tabulated. The estimation in each table cell may be considered as a subpopulation estimation problem. In this case, the proportion of such subpopulation is usually small. For instance, if there were 10 table cells, some of these cells would have proportion of no more than 10 percent. Due to such low proportion of the subpopulation, the characteristics of the estimate may be quite different from those when the proportion is close to 1. Does the estimate have normal distribution approximately? How reliable are the estimate and the variance estimate? This paper investigates some of these problems.

We consider the problem from several angles. We first study the variance and relvariance of the simple expansion estimate to see how they depend on the proportion of subpopulation under various circumstances. We then decompose the variance into components to see some insight of the variance. We also study the variance of the variance estimate. A simulation is done to study the empirical distributions of the estimate and its variance estimate. Due to the limitation of the number of pages, the detail of simulation study is not reported.

We found that the simple expansion estimate performs well when the coefficient of variation per unit in the subpopulation is no less than 1. The variance is basically proportional to the proportion under this circumstance; the proportion of the subpopulation does not have very much effect. But when the coefficient of variation is small, the variance is very unstable and is extremely large for most value of the proportion. In this case a large portion of the variance is contributed by factors other than the variation

of unit measures in the subpopulation. The simple expansion method should be avoided under this circumstance.

From the simulation study, the distribution of the estimate seems to be approximately normal when the coefficient of variation in the subpopulation is small. When the coefficient of variation is large (C=4) and when the proportion is small, the empirical distribution is skew. Overall, the distribution of the estimate when the proportion is small seems to behave well; it is not much worse than the case when the proportion is close to 1.

This paper studies the (unconditional) variance of the estimate. It assesses the performance of the estimate disregarding the number of units in the subpopulation realized in the sample. The conditional variance of the estimate will be the topic of a separate paper.

References

Cochran, W.G. (1977), Sampling Techniques, (3rd edition) John Wiley & Sons, New York.

Cox, D.R. & Snell, E.J. (1979), On Sampling and Estimation of Rare Errors, Biometrica, (66) 125-32.

Hansen, M.H.; Hurwitz, W.N. & Madow, W.G. (1953a), Sample Survey Methods and Theory, Vol. 1, John Wiley,& Sons, New York.

Hansen, M.H.; Hurwitz, W.N. & Madow, W.G. (1953b), Sample Survey Methods and Theory, Vol. 2, John Wiley & Sons, New York.

Holt, D. & Smith, T.M.F. (1979), Post Stratification, J.R. Statist. Soc. A 142, Part 1, 33-46.

Jones, D.H. & Coppersmith, L. (1976), A Ratio Estimator of the Total of A Subpopulation, Commun. Statist. Theor. Meth., A5(3), 251-60.

Kish, L. (1967), Survey Sampling, John Wiley & Sons, New York.

Neter, John & Loebbecke, James K. (1975), Behavior of Major Statistical Estimators in Sampling Accounting Populations, An Empirical Study AICPA, New York.

Neter, John & Loebbecke, James K. (1977), On the Behavior of Statistical Estimators When Sampling Accounting Population, J. of Amer. Statist. Assoc. Vol. 72 No. 359 501-07.

Sukhatme, P.V. & Sukhatme, B.V. (1970), Sampling Theory of Surveys with Applications (2nd edition), Iowa State University Press, Ames, Iowa.

FIGURE 1 -- THE DEPENDENCE OF THE VARIANCE OF THE SIMPLE EXPANSION ESTIMATE ON P

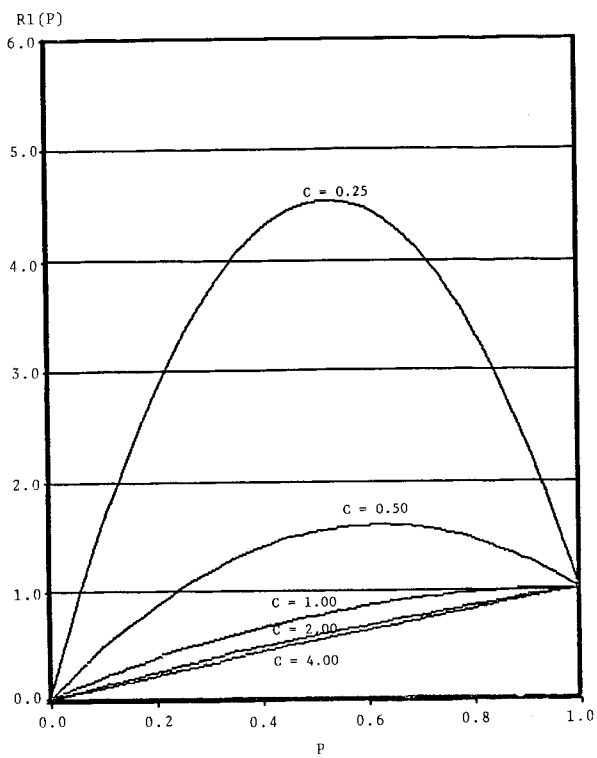

FIGURE 3 -- COMPONENT RATIO OF THE FIRST DECOMPOSITION OF THE VARIANCE OF THE SIMPLE EXPANSION ESTIMATE

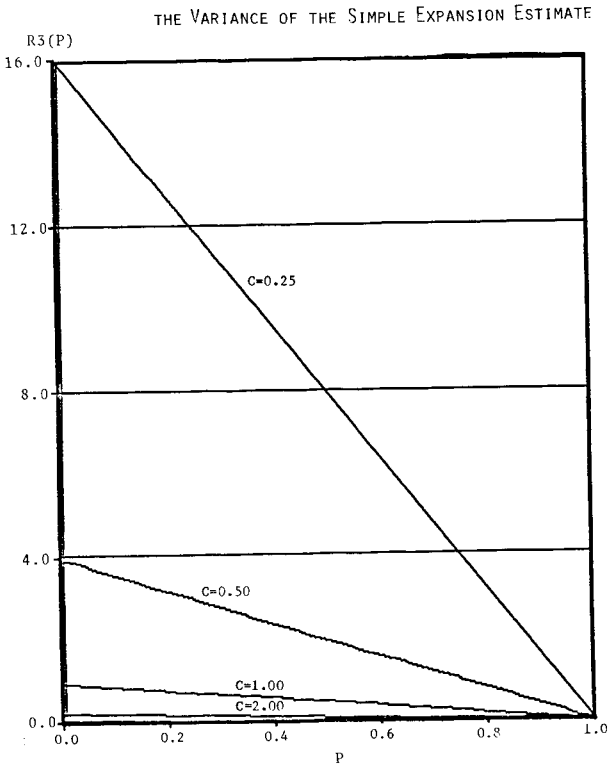

FIGURE 4 -- COMPONENT RATIO OF THE SECOND DECOMPOSITION OF THE VARIANCE OF THE SIMPLE EXPANSION ESTIMATE

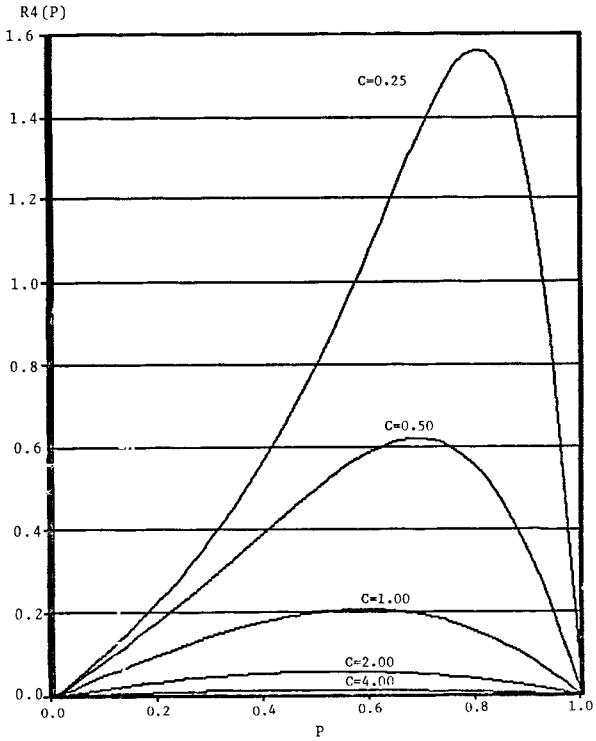

FIGURE 5(A) -- THE DEPENDENCE OF THE VARIANCE OF A VARIANCE ESTIMATE ON P

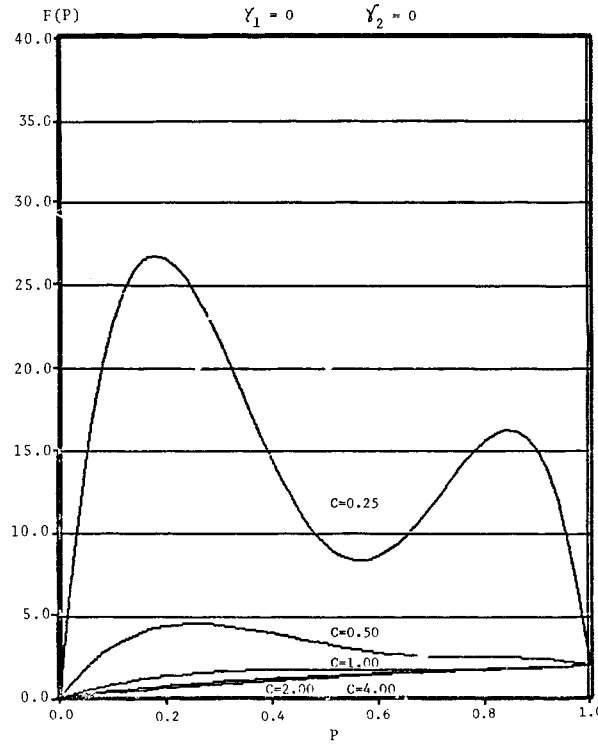

FIGURE 5(B) -- THE DEPENDENCE OF THE VARIANCE OF A VARIANCE ESTIMATE ON P

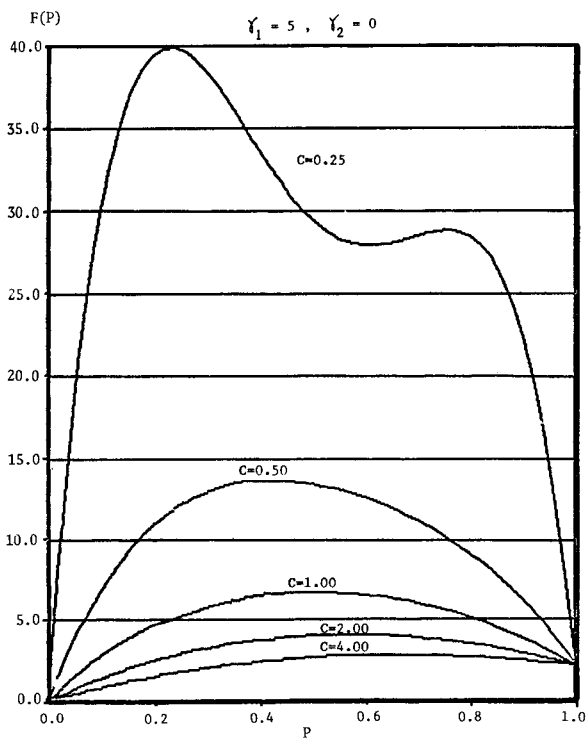

FIGURE 5(C) -- THE DEPENDENCE OF THE VARIANCE OF A VARIANCE ESTIMATE ON P