PART I: SAMPLE SURVEY
CHAPTER 1

JACKKNIFING A MODIFIED RATIO ESTIMATOR UNDER A SUPER-POPULATION MODEL

SUMMARY: Quenouille's method of bias reduction, also called Quenouille-Tukey jackknifing, is applied to a modified ratio estimator due to Srivastava (1967). The performance of the jackknifed estimator with regard to bias and variance is examined and the condition of its superiority is derived. Finally, an optimum choice of the number of groups to be considered in connection with the jackknifing is obtained and the result is seen to agree with that of Rao (1965) and Rao and Webster (1966).

Rao (1991) has recently studied the different methods of improving, through modifications, the classical ratio and regression estimators, and concluded that there is no real gain in such modifications. We have, however, shown in this chapter that jackknifing of the modified estimators may result in a better precision in some cases.
1. INTRODUCTION

The classical ratio estimator for the population mean (total), which incorporates auxiliary information available on a variable correlated to the study variable has been extended in many ways. One such extension is due to Srivastava (1967) who defined a ratio estimator of the population mean \( \bar{Y} \) as

\[ \hat{Y}_\alpha = \bar{y} \left( \frac{\bar{x}}{\bar{X}} \right)^\alpha, \]

(1.1)

where \( \bar{y} \) and \( \bar{x} \) are the respective sample means of the study variable \( y \) and the auxiliary variable \( x \). Here, the population mean \( \bar{X} \) of \( x \) is assumed to be known, and \( \alpha \) is any real number. The estimator in (1.1) reduces, for \( \alpha = -1, 0 \) and 1, to ratio, mean-per-unit and product estimators, respectively. Recently, Rao (1991) has considered many such modified estimators available in the literature, including (1.1), which may be called the Srivastava's modified ratio estimator (SURE), and compared them with regard to bias and mean square error (MSE).

The classical ratio estimator is known to be biased under an SRS (Simple Random Sampling) design. With a view to reducing this bias, Durbin (1959) showed that if the auxiliary variable \( x \) is normally distributed and if the regression of \( y \) on \( x \) is linear, then the ratio estimator of Quenouille (1956), being obtained by jackknifing the classical ratio estimator, reduces not only the bias but also the variance, asymptotically. Durbin also considered a second model, where the regression is linear, but \( x \) now follows a gamma distribution.
One problem that arises in using the jackknifing technique is the determination of the optimal number of groups \( g \), say, into which the data are to be divided. Rao (1965) and Rao and Webster (1966) showed that the optimal choice is \( g = n \), \( n \) being the sample size. They, however, adopted the Durbin's models for their discussions.

The contents of this chapter can be divided into three parts. In the first part, we jackknife the SMRE to get an estimator which we call the JSMRE. The estimator JSMRE is shown to be unbiased upto the order \( 1/n \), while the SMRE is not. Also, the MSE's of the SMRE and JSMRE are compared for different values of \( \alpha \) and \( n \). It is observed that the JSMRE performs better.

In the second part, we study the estimators SMRE and JSMRE under the model of Durbin which is as follows:

\[
\begin{align*}
(i) & \quad y_i = \beta + \beta x_i + e_i, \\
(ii) & \quad E(y_i / x_i) = \beta + \beta x_i, i.e., E(e_i / x_i) = 0, \\
(iii) & \quad V(e_i / x_i) = \delta x_i, \quad \delta > 0,
\end{align*}
\]

and, finally, the \( x_i \) are assumed to be independent gamma variables with the common parameter \( \lambda \).

In the last chapter, the problem of optimal choice of \( g \) is considered. It is shown that the asymptotic relative efficiency (ARE) of the JSMRE compared to SMRE is an increasing function of \( g \), so that \( g = n \) is the optimum choice. This agrees with the earlier results in this context.
2. COMPARISON OF EFFICIENCIES OF SMRE AND JSMRE

Suppose we are given a random sample (without replacement) of $n$ pairs of observations $(x_i, y_i)$, $i = 1, 2, \ldots, n$, from a finite population of $N$ units. Then, assuming the model

\[
x_i = \bar{x} + \epsilon_i, \quad y_i = \bar{y} + \epsilon_i,
\]

with \( E(\epsilon_i) = E(\epsilon_i) = 0 \),

we can easily show, on taking \(|(\epsilon' / \bar{x})| < 1\), that

\[
\tilde{\gamma}_a = \bar{y} (1 + \epsilon / \bar{y}) (1 + \epsilon' / \bar{x})^\alpha
\]

\[
= \bar{y}(1 + \epsilon / \bar{y} + \alpha \epsilon' / \bar{x} + \alpha (\alpha - 1) \epsilon^2 / (2 \bar{x}^2) + \alpha \epsilon \epsilon' / (\bar{x} \bar{y}) + \ldots),
\]

where \( \bar{\epsilon} \) and \( \bar{\epsilon}' \) are the sample means of \( \epsilon_i \) and \( \epsilon_i' \), respectively.

Now, ignoring terms in \( \epsilon_i \) of order greater than two, and also taking expectation, we have from (2.2),

\[
E(\tilde{\gamma}_a) = \bar{y} + \bar{y} (\alpha f (\epsilon C_x C_y + (\alpha - 1) \sigma_x^2 / 2 ) / n),
\]

where \( f = (N-n)/N \) is the finite population correction (fpc), \( \rho \) is the correlation coefficient between \( x \) and \( y \) and \( C_x \) and \( C_y \) are the coefficients of variation (cv) of \( x \) and \( y \), respectively.

The bias of \( \tilde{\gamma}_a \), as obtained in (2.3), can be completely eliminated by jackknifing. Let us call the estimator \( \tilde{\gamma}_a \), defined above, the SMRE, while its jackknifed version \( \tilde{\gamma}_a^* \), say, will be called the JSMRE. The process of jackknifing \( \tilde{\gamma}_a \) is explained below.
For each \( i \), we define \( \hat{y}_\alpha^{(i)} \) to be the estimator \( \hat{y}_\alpha \) which is now calculated deleting the \( i \)-th pair, namely, \((x_i, y_i)\) from the given data set. The expression for \( \hat{y}_\alpha^{(i)} \) will then be

\[
\hat{y}_\alpha^{(i)} = \left( \sum_{\neq i} y_j/(n-1) \right) \left( \sum_{\neq i} x_j/(n-1) \right) \bar{x} \alpha^{-\alpha} \nonumber \\
= (1 - 1/n) \hat{y}_\alpha \left( 1 - y_i/(n\bar{y}) \right) \left( 1 - x_i/(n\bar{x}) \right), \tag{2.4}
\]

where ' \( \sum' \) ' denotes summation for 'i' missing.

We then obtain the arithmetic mean of \( \hat{y}_\alpha^{(i)} \) as

\[
\frac{\hat{y}_\alpha^{(o)}}{\sum_{i=1}^{n} \hat{y}_\alpha^{(i)}} \tag{2.5}
\]

Finally, the jackknifed estimator \( \tilde{y}_\alpha^* \) of the population mean \( \bar{Y} \) is obtained as

\[
\tilde{y}_\alpha^* = n \hat{y}_\alpha - (n-1) \hat{y}_\alpha^{(o)} \tag{2.6}
\]

which we call the JSMRE.

The unbiasedness of \( \tilde{y}_\alpha^* \) upto the order \( 1/n \) : To prove the unbiasedness of \( \tilde{y}_\alpha^* \) upto the order \( 1/n \), we assume \( f = 1 \). Again, \( E(\tilde{y}_\alpha^{(i)}) \) is the same as \( E(\hat{y}_\alpha) \) with \( n \) replaced by \( n-1 \). Under these considerations, we have

\[
E(\tilde{y}_\alpha^*) = n E(\hat{y}_\alpha) - \frac{(n-1)}{n} \sum_{i=1}^{n} E(\hat{y}_\alpha^{(i)}) \\
= \bar{Y},
\]

which proves the result.
We now obtain the MSE of the JSMRE for large sample size. With this in mind, we note that

\[
\text{MSE}(\tilde{y}_\alpha) = (\bar{y}^2 f/n)(\alpha^2 c_x^2 + c_y^2 + 2 \alpha \rho c_x c_y). \tag{2.7}
\]

Srivastava (1967) noted that the MSE attains its minimum at the value \( \alpha = -\rho c_y / c_x \), though at this value of \( \alpha \) the bias of the SMRE will not be zero.

From (2.4) and (2.6), it is easy to show that the jackknifed estimator admits the representation

\[
\tilde{y}_\alpha^* = n \tilde{y}_\alpha - \left( 1 - l/n \right) \tilde{y}_\alpha \sum_{i=1}^{n} (1 - y_i / (n \bar{y}))(1 - x_i / (n \bar{x}))^\alpha. \tag{2.8}
\]

Assuming that \( n \) is large and \( |y_i / (n \bar{y})| < 1 \), \( |x_i / (n \bar{x})| < 1 \), and also approximating the summand on the right-hand side of (2.8) by \( n - \alpha - 1 \), we get

\[
\tilde{y}_\alpha^* = (n - (1 - l/n) \tilde{y}_\alpha) \left( n - \alpha - 1 \right) \tilde{y}_\alpha. \tag{2.9}
\]

Thus, up to the order \( 1/n \), we have

\[
\tilde{y}_\alpha^* = (1 + \alpha (\alpha + 1)/(2n)) \tilde{y}_\alpha. \tag{2.10}
\]

The mean square error of \( \tilde{y}_\alpha^* \) is finally obtained as

\[
\text{MSE}(\tilde{y}_\alpha^*) = (1 + \alpha (\alpha + 1)/(2n))^2 (\bar{y}^2 f/n)(\alpha^2 c_x^2 + c_y^2 + 2 \alpha \rho c_x c_y). \tag{2.11}
\]

Clearly, for \( \alpha = 0 \) or \(-1\), the SMRE and JSMRE are equally efficient.
Rao (1991) studied the SMRE, along with other modified estimators, for four natural populations and concluded that $\alpha = -1$ is as good as the choice of $\alpha_{opt}$. He thus found no merit in considering the modified estimators, e.g., the estimator SMRE. We, however, suggest here an improvement over the SMRE, namely, the jackknifed version JSMRE. The MSE's of these estimators are compared for different values of $\alpha$ and $n$. This is shown in Table 1 below.

Here, we define the asymptotic relative efficiency (ARE) of $\hat{y}_\alpha^{*}$ compared to $\hat{y}_\alpha$ as the ratio $\text{MSE}(\hat{y}_\alpha^{*}) / \text{MSE}(\hat{y}_\alpha)$, expressed as a percentage. It is observed that this ARE of $\hat{y}_\alpha^{*}$ is greater than 100 for values of $\alpha$ lying in the interval (-1,0), and smaller otherwise.

Also, as the sample size $n$ increases, the estimators SMRE and JSMRE become equally efficient. In other words, if the sample size $n$ is large, then there is no real gain from jackknifing the modified estimator.

However, for values of $\alpha$ near -0.5 and also for small values of $n$, the jackknifed estimator JSMRE gives a better precision.
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.05</td>
<td>100.96</td>
<td>100.23</td>
<td>100.24</td>
<td>100.16</td>
</tr>
<tr>
<td>-0.10</td>
<td>101.82</td>
<td>100.91</td>
<td>100.45</td>
<td>100.30</td>
</tr>
<tr>
<td>-0.25</td>
<td>103.86</td>
<td>101.90</td>
<td>100.94</td>
<td>100.63</td>
</tr>
<tr>
<td>-0.40</td>
<td>104.98</td>
<td>102.44</td>
<td>101.21</td>
<td>100.80</td>
</tr>
<tr>
<td>-0.50</td>
<td>105.19</td>
<td>102.55</td>
<td>101.26</td>
<td>100.84</td>
</tr>
<tr>
<td>-0.75</td>
<td>104.98</td>
<td>102.44</td>
<td>101.21</td>
<td>100.80</td>
</tr>
<tr>
<td>-0.90</td>
<td>101.82</td>
<td>100.91</td>
<td>100.45</td>
<td>100.30</td>
</tr>
<tr>
<td>-0.95</td>
<td>100.96</td>
<td>100.23</td>
<td>100.24</td>
<td>100.16</td>
</tr>
</tbody>
</table>
3. JACKKNIFING THE SMRE UNDER DURBIN'S MODEL

We now jackknife the estimator SMRE under Durbin's model given in (1.2). This is shown below.

It is easy to show that

\[ E(\bar{x}^\alpha) = \begin{cases} \frac{\Gamma(n\lambda + \alpha)}{\Gamma(n\lambda)}, & \text{for } n > -\alpha/\lambda \\ \text{undefined, otherwise.} \end{cases} \]

Hence, on using \( E(\bar{e}/x) = 0 \) and assuming the existence of all the relevant expectations, we have

\[ E(\tilde{y}_\alpha) = E(\bar{y} \bar{x}^\alpha \bar{x}^{-\alpha}) \]
\[ = \bar{x}^{-\alpha} E(\bar{y}^\alpha E(\bar{y}/x)) \]
\[ = \bar{x}^{-\alpha} \left( \sum_{i} i n^{-\alpha} \Gamma(n\lambda + \alpha) + \sum_{k} \Gamma(n\lambda + \alpha + 1) \right). \]

(3.1)

A particular case: Suppose \( \beta = 0 \) and \( \alpha = -1 \). Then, \( E(\tilde{y}_\alpha) = \beta \lambda = \bar{y} \), so that the estimator SMRE becomes unbiased for the population mean. Moreover, if \( \beta = 0 \), but \( \alpha \neq -1 \), then

\[ E(\tilde{y}_\alpha - \beta \lambda) = \beta \lambda (\Gamma(n\lambda^{-(\alpha + 1)})/\Gamma(n\lambda)) \]
\[ = \beta \lambda (\Gamma(n\lambda + \alpha + 1)/\Gamma(n\lambda) - 1), \]

(3.2)

which gives the bias of SMRE in this case. An approximate expression for this bias can be obtained by using the asymptotic expansion of the gamma function, which is

\[ \Gamma(n\lambda + \alpha + 1)/\Gamma(n\lambda) \approx (n\lambda)^{\alpha+1}. \]

(3.3)
Using (3.3) in (3.2), we have $E(\tilde{y}_\alpha - \beta \lambda) = 0$, so that the estimator SMRE is asymptotically unbiased under Durbin's model.

Large sample variance of the SMRE when both $\beta$ and $t$ are zero:

Using the relation

$$V(\tilde{y}_\alpha) = V E(\tilde{y}_\alpha / x) + E V(\tilde{y}_\alpha / x),$$

we can easily show that $V(\tilde{y}_\alpha)$ is approximately

$$V(\tilde{y}_\alpha) \approx \beta^2 \bar{X}^{-2\alpha} \left[ (E(\bar{X}^{2\alpha+1}) - (E(\bar{X}^{\alpha+1}))^2 \right] + (\delta/n) \bar{X}^{-2\alpha} E(\bar{X}^{2\alpha}).$$

(3.4)

Then, using the same asymptotic expansion as in (3.3), we have for large $n$,

$$V(\tilde{y}_\alpha) \approx \delta/n.$$  

(3.5)

We have thus proved, under Durbin's model with $\beta=0, t=0$ and also for large sample size, that

(a) the SMRE is unbiased for the population mean,

(b) the standard error (s.e.) of the SMRE is $\sqrt{(\delta/n)}$.

Finally, returning to the jackknifed estimator JSMRE, we have

under the same model and assumptions,

$$E(\tilde{y}^{\text{J}}_\alpha) = (1 + \alpha(\alpha + 1)/(2n)) \beta \lambda,$$

(3.6)

and

$$V(\tilde{y}^{\text{J}}_\alpha) = (1+\alpha(\alpha+1)/(2n))^2 (\delta/n).$$

(3.7)
4. OPTIMUM CHOICE OF THE NUMBER OF GROUPS FOR JACKKNIFING

In performing the process of jackknifing, the data has to be divided into a number of groups and then one group is deleted at a time. In deriving the jackknifed estimator, as explained in the preceding Sections, we have assumed that the number of groups is equal to \( n \), so that only one observation is deleted at a time. We shall now show that this choice of \( g \) is, in fact, the optimum choice.

Let, for a given sample of size \( n \), \( g \) be the number of groups with \( m \) observations in each group, so that we have \( n = mg \). Then, in jackknifing the SMRE we delete one group of \( m \) observations at a time. This is explained below.

Let \( \hat{y}_d^{(i)} \) denote the estimator \( \hat{y}_d \) obtained from the sample after omitting the \( i \)-th group, \( i = 1, 2, \ldots, g \). The jackknifed estimator JSMRE is, therefore, given by

\[
\hat{y}_d^* = g \hat{y}_d - (g - 1) \hat{y}_d^{(o)}, \tag{4.1}
\]

where

\[
\hat{y}_d^{(o)} = \frac{1}{g} \sum_{i=1}^{g} \hat{y}_d^{(i)},
\]

wherein

\[
\hat{y}_d^{(i)} = \bar{y}_i^{(i)} \left( \bar{x}_i^{(i)} / \bar{x} \right)^d, \tag{4.2}
\]

where \( \bar{y}_i \) and \( \bar{x}_i \) denote, respectively, the sample means of \( y \) and \( x \) after omitting the \( i \)-th group.

We also note that

\[
n \bar{y} = \text{sample total of } y
\]

\[
= (n - m) \bar{y}_i^{(i)} + m \bar{y}_i, \tag{4.3}
\]
where $\bar{y}_i$ is the sample mean of $y$ in the $i$-th group.

Similarly,

$$n \bar{x} = (n - m) \bar{x}_i + m \bar{x}_i, \quad (4.4)$$

where $\bar{x}_i$ is the sample mean of $x$ in the $i$-th group.

Now, following the approach of Section 2, (4.2) can be re-written as

$$\tilde{y}_d^{(i)} = \frac{1}{(1-m/n)^{(\alpha+1)}} \tilde{y}_d \left( \frac{1-m\bar{y}_i}{(n\bar{y})} \right) \left( \frac{1-m\bar{x}_i}{(nx)} \right). \quad (4.5)$$

We also assume that $|m \bar{y}_i/(n \bar{y})| < 1$, $|m \bar{x}_i/(nx)| < 1$, which might be ensured if $n$ is large compared to $m$. Under these considerations, we have

$$\sum_{i=1}^{g} \tilde{y}_d^{(i)} = \frac{1}{(1-m/n)^{(\alpha+1)}} \tilde{y}_d \sum_{i=1}^{g} \left( \frac{1-m\bar{y}_i}{(n\bar{y})} \right) \left( \frac{1-m\bar{x}_i}{(nx)} \right).$$

The jackknifed estimator JSMRE is then obtained as

$$\tilde{y}_d^* = (g - \left( 1 - 1/g \right) \left( 1 - m/n \right)^{(\alpha+1)} \left( g - \alpha - 1 \right)) \tilde{y}_d, \quad (4.6)$$

which, for $g = n$, reduces to (2.9).

In case $g$ is large, keeping $m$ finite, we have as in (2.10),

$$\tilde{y}_d^* = \left( 1 + \alpha (\alpha + 1)/(2g) \right) \tilde{y}_d. \quad (4.7)$$
Now, it is known that, asymptotically, $E(\tilde{\gamma}_\alpha) = \bar{Y}$. The relative bias of $\gamma^*$ is, therefore, obtained from (4.7) as

$$\text{Relative bias of } \tilde{\gamma}_\alpha^* = \frac{E(\tilde{\gamma}_\alpha^*) - \bar{Y}}{\bar{Y}}$$

$$= \alpha(\alpha + 1)/(2g), \quad (4.8)$$

which, for fixed $\alpha$, decreases as $g$ increases.

Also, the asymptotic relative efficiency (ARE) of $\tilde{\gamma}_\alpha^*$ compared to $\tilde{\gamma}_\alpha$ is given by

$$\text{ARE of } \tilde{\gamma}_\alpha^* = \frac{\text{MSE}(\tilde{\gamma}_\alpha^*)}{\text{MSE}(\tilde{\gamma}_\alpha^*)}.100$$

$$= (1 + \alpha(\alpha + 1)/(2g))^2.100, \quad (4.9)$$

which, for fixed $\alpha$, increases as $g$ increases.

Since the maximum value that the number of groups $g$ can take is $n$, we have, from (4.8) and (4.9), the optimum value of $g$ as $g_{\text{opt}} = n$. Rao (1965) and Rao and Webster (1966) obtained similar results in this context.