1. Introduction

This paper deals with the problem of efficient estimation of the mean and of the total when measurements from times $t_1$ and $t_2$ are available for the variable under study. We also consider estimation of change from $t_1$ and $t_2$, change being expressed either as a difference of means and totals or as a ratio of totals. These problems are often thoroughly examined in standard texts when the population is assumed to be infinite and composed of the same units at $t_1$ and $t_2$. See e.g. Cochran (1977) and Raj (1968).

Here, by contrast, we assume that
(i) The population is changeable, that is, units may have joined or left the population between $t_1$ and $t_2$.
(ii) The population is finite.

Linear minimum-variance unbiased (LMVU) estimators and their variances are presented. Results for the ratio estimator are presented also. Optimum fractions are given for the matched part of the sample taken at time $t_2$, with the corresponding minimal variances.

2. Basic concepts

We assume a situation where the mean and total of a variable under study are to be estimated at times $t_1$ and $t_2$ in a repetitive sample survey, and where the change between $t_1$ and $t_2$ is of special interest. The variable under study is called $x$ at time $t_1$ and $y$ at time $t_2$. The changeable population is called $U$ at $t_1$ and $U'$ at $t_2$, with size $N$ and $N'$ respectively.

A sample of $n$ units drawn at time $t$ is denoted $s_t(n)$.

At time $t_1$ the sample design is as follows (see figure 1): $n$ units are drawn with Simple Random Sampling (SRS) from the $N$ units of $U$. Measurements are made of the variable $x$. Naturally, at this time, the population $U'$ is unknown.

At time $t_2$ it is assumed that:
- Between $t_1$ and $t_2$, $N_1$ units have left $U$ and $N_2$ units have joined it. The corresponding subpopulations are called $U_1$ and $U_2$, respectively.
- The new population $U'$ contains $N' = N_1 + N_2$ units, where $N_2$ is the number of units in $U_2$, that is the intersection of $U$ and $U_2$.
- For the sample of $n$ drawn at $t_1$, $n_1$ units belong to $U_1$ and $n_12$ to $U_2$.

At time $t_2$ the sampling design is as follows:
- The sample consists of $n'$ units and is partitioned into three parts, namely
  - $n_{12m}$ ($m = matched$) units are drawn with SRS from $s_{t_1}(n_{12})$
  - $n_{12u}$ ($u = unmatched$) units are drawn with SRS from $U_{12} - s_{t_1}(n_{12})$
  - $n_2$ units are drawn with SRS from the $N_2$ units of $U_2$.

The sample is drawn in a way that the following relations hold true:

$$ n = n_1 + n_{12} $$
$$ n' = n_{12m} + n_{12u} + n_2 $$
$$ n_{12} = n_{12m} + n_{12u} $$

Figure 1: The sampling design

The relative size of the unmatched part of the sample of $U_{12}$ at $t_2$ is denoted

$$ \theta = \frac{n_{12u}}{n_{12}} $$

which implies that

$$ 1 - \theta = \frac{n_{12m}}{n_{12}} $$

The relative change of the population is denoted by $Q_1$ and $Q_2$, where

$$ Q_1 = \frac{N_2}{N_1} - 1 = \frac{N_1 - 1}{N_1} $$
$$ Q_2 = \frac{N_2'}{N_1'} - 1 = \frac{N_1'}{N_1'} - 1 $$

The population means, totals and variances for $x$ and $y$ in $U_1$, $U_{12}$ and $U_2$ are defined by

$$ \bar{X}_{1} = \frac{1}{N_1} \sum_{i=1}^{N_1} X_i ; \bar{X}_{12} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} X_{12}; \bar{Y}_{12} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} Y_{12} $$
$$ \bar{Y}_{1} = \frac{1}{N_1} \sum_{i=1}^{N_1} Y_i ; \bar{Y}_{12} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} Y_{12} $$
$$ \bar{Y}_{2} = \frac{1}{N_2} \sum_{i=1}^{N_2} Y_i $$

Furthermore, it is assumed that

$$ \sigma^2_{12x} = \sigma^2_{12y} = \sigma^2 $$

Then $\rho_{xy}$, the coefficient of correlation between $x$ and $y$ in $U_{12}$, is

$$ \rho_{xy} = \frac{\text{Cov}(x,y)}{\sigma^2} $$

where

$$ \text{Cov}(x,y) = \frac{\sum_{i=1}^{N_{12}} (x_{12} - \bar{x}_{12})(y_{12} - \bar{y}_{12})}{N_{12} - 1} $$

417
3. Two useful theorems

The following two theorems will be useful in the derivation of the estimators and their variances.

**Theorem 1**: Consider a population \( U \) which is fixed over times \( t_1 \) and \( t_2 \) and contains \( M \) units. The following sampling design is assumed:

At \( t_1 \), a SRS of \( u \) units is drawn from the \( M \) units of \( U \).

At \( t_2 \), a SRS of \( g \) units is drawn from \( \mathcal{S}(u) \) and a SRS of \( v-g \) units is drawn from \( \mathcal{U}_{t_2}(v) \) (Note that \( \mathcal{U}_{t_2}(v) = \mathcal{S}(v) \) is a SRS from \( U \)).

Assume also that \( x \) and \( y \) are variables under study at \( t_1 \) and \( t_2 \) respectively and that \( X_U \) and \( Y_{t_2} \) are simple estimators of means and that \( \text{Cov}(x, y) = \rho \).

Then

\[
- \frac{1}{s^2} \text{Cov}(aX_U, bY_{t_2}) = ab(u-v-M) \rho,
\]

where \( a \) and \( b \) are arbitrary constants. This can be proved easily.

In our case, \( u, v \), and \( M \) correspond to \( n_{12}, n_{12}', n_{12} \) and \( N_{12} \). With theorem 1, covariances are easily calculated.

The next theorem is due to Rao, C.R. (1952). See also Raj (1968).

**Theorem 2**: A necessary and sufficient condition for \( \hat{T}_0 \) to be the minimum-variance-unbiased estimator (MVU) of a parameter is that \( \hat{T}_0 \) is unbiased and that \( \text{Cov}(T_0, z) = 0 \) for all \( z \) where \( z \) is a zero function, i.e., \( E(z) = 0 \).

**Corollary 2.1**: If \( T_0 \) is a MVU-estimator of a parameter and \( \hat{T}_1 \) is an unbiased estimate of that parameter then \( \text{V}(T_0) = \text{V}(T_0 - T_1) \). This follows from Theorem 2 since

\[
\begin{align*}
\text{Cov}(T_0, z) = 0 & \Rightarrow \text{Cov}(T_0, T_0 - T_1) = 0 \\
& \Rightarrow \text{Cov}(D, z) = \text{Cov}(T_0, z) - \beta \text{Cov}(T_0, z) = 0
\end{align*}
\]

This follows from theorem 2 since

\[
\text{Cov}(T_0, z) = \text{Cov}(T_0, z) = 0
\]

Theorem 2.2 with corollaries will be used to derive linear MVU-estimators (LMVU) and their variances.

**Linear MVU- and MV-estimators**

Linear MVU-estimators (LMVU) and their variances have been derived for the following parameters:

\[\begin{align*}
\tilde{X} &= \frac{\tilde{X}_{t_1}}{N} & X &= N \tilde{X} & \text{in } U & \text{at } t_1 \\
\tilde{Y} &= \frac{\tilde{Y}_{t_2}}{N'} & Y &= N' \tilde{Y} & \text{in } U' & \text{at } t_2 \\
\end{align*}\]

Minimum variance-estimator (MV) and its variance have been derived for the parameter

d. \( Y/X \).

Optimum variances have been calculated by minimizing the variances of the LMVU- and MV-estimators with respect to \( \theta = \frac{n_{12}u}{n_{12}'} \).

The method of calculating LMVU-estimators and their variances is, in principal, the same for the above parameters a-c. The techniques to derive the LMVU-, the MV-estimators and their variances are shown in Forsman and Garås (1982).

### 4. Results for the LMVU- and MV-estimators

**LMVU**- and **MV-estimators** have their variances and optimal variances given in this section for the parameters \( X, X, Y, Y, Y_X, Y_X \) and \( Y/Y \). Let us introduce the following concepts:

An LMVU-estimator, here considered, is denoted by \( \hat{e}_{LMVU} \). Then

\[
\text{V}_{\text{opt}}(\hat{e}_{LMVU}) = \min_{\theta} \text{V}(\hat{e}_{LMVU}) \quad \text{where } \theta = \frac{n_{12}u}{n_{12}'} \quad 0 < \theta < 1
\]

The \( \theta \)-value minimizing \( \text{V}(\hat{e}_{LMVU}) \) is denoted \( \theta_{\text{opt}} \).

If \( \rho = \pm 1 \) and \( \theta \to 1 \Rightarrow \text{V}(\hat{e}_{LMVU}) \to \text{V}_{\text{opt}}(\hat{e}_{LMVU}) / \rho \).

\( \rho = 0 \Rightarrow \text{V}_{\text{opt}}(\hat{e}_{LMVU}) \) is independent of \( \theta \).

The formulas below are valid for \( -1 < \rho < 1 \). The above mentioned about \( \hat{e}_{LMVU} \) is also valid for the MV-estimator, although negative \( \rho \) is not considered here.

a. **Estimator of total at \( t_1 **

\[
\hat{X}_{LMVU} = \hat{X}_{t_1} + \hat{X}_{t_2} \frac{1}{1-\rho^2} \left[ (1-\theta) \hat{X}_{t_2} - \frac{\rho}{1-\rho^2} \hat{X}_{t_1} \right]
\]

\[
\text{V}(\hat{X}_{LMVU}) = N_{12}^2 \left( \frac{1}{n_{12}'} + \frac{1}{n_{12}} \right) \frac{1}{1-\rho^2} \frac{\rho}{1-\rho^2} \left[ 1 - \frac{n_{12}}{n_{12}'} \left( \frac{1}{n_{12}'} - \frac{1}{n_{12}} \right) \right]
\]

b. **Estimator of mean at \( t_1 **

\[
\hat{X}_{LMVU} = \frac{\hat{X}_{LMVU}}{N}
\]

\[
\text{V}(\hat{X}_{LMVU}) = \text{V}(\hat{X}_{LMVU}) / N^2
\]

c. **Estimator of total at \( t_2 **

\[
\hat{Y}_{LMVU} = \frac{\hat{Y}_{LMVU}}{N'}
\]

\[
\text{V}(\hat{Y}_{LMVU}) = \text{V}(\hat{Y}_{LMVU}) / N'^2
\]

d. **Estimator of mean at \( t_2 **

\[
\hat{Y}_{LMVU} = \frac{\hat{Y}_{LMVU}}{N'}
\]

\[
\text{V}(\hat{Y}_{LMVU}) = \text{V}(\hat{Y}_{LMVU}) / N'^2
\]
e. Estimator of difference between totals

\[(\hat{Y} - \hat{X})_{LMVU} = \hat{Y}_{LMVU} - \hat{X}_{LMVU} = \hat{Y}_2 - \hat{X}_1 + \hat{Y}_{12u} - \hat{X}_{12} + \]

\[+ \frac{(1-\theta)}{1-\rho^2} \left[ \hat{Y}_{12m} - \hat{Y}_{12u} + \hat{X}_{12m} - \hat{X}_{12u} \right] \]

\[V(\hat{Y} - \hat{X})_{LMVU} = \frac{2n_2^2s_2^2(1-\rho)}{n_{12}} \left[ \frac{1}{1-\rho^2} \cdot \frac{n_{12}}{N_{12}} \right] \]

\[
\theta_{min} = 0 \Rightarrow V_{opt}(\hat{Y} - \hat{X})_{LMVU} = N_2^2s_2^2 \left( \frac{1}{n_2} - \frac{1}{N_2} \right) + \]

\[+ N_1^2s_1^2 \left( \frac{1}{n_1} - \frac{1}{N_1} \right) + 2n_2^2s_2^2(1-\rho) \left( \frac{1}{n_{12}} - \frac{1}{N_{12}} \right) \]

\[
\theta_{min} = 1 \Rightarrow V_{opt}(\hat{Y} - \hat{X})_{LMVU} = N_2^2s_2^2 \left( \frac{1}{n_2} - \frac{1}{N_2} \right) + \]

\[+ N_1^2s_1^2 \left( \frac{1}{n_1} - \frac{1}{N_1} \right) + 2n_2^2s_2^2(1-\rho) \left( \frac{1}{n_{12}} - \frac{1}{N_{12}} \right) \]

f. Estimator of difference between means

\[(\hat{Y} - \hat{X})_{LMVU} = \hat{Y}_{LMVU} - \hat{X}_{LMVU} = (1-\rho) \hat{Y}_2 - (1-\theta) \hat{X}_1 + \rho \hat{Y}_{12u} - \]

\[\hat{X}_{12} + \frac{(1-\theta)(\rho + \theta \rho)}{1-\rho^2} \left[ \hat{Y}_{12m} - \hat{Y}_{12u} + \hat{X}_{12m} - \hat{X}_{12u} \right] \]

\[V(\hat{Y} - \hat{X})_{LMVU} = (1-\rho)^2s_2^2 \left( \frac{1}{n_2} - \frac{1}{N_2} \right) + (1-\theta)^2s_1^2 \left( \frac{1}{n_1} - \frac{1}{N_1} \right) + \]

\[\frac{s^2}{n_{12}} \left[ \frac{(Q_1^2 + Q_2^2)(1-\theta^2) - 2\rho Q_1 Q_2}{1-\rho^2} + \frac{n_{12}}{N_{12}} \left( \frac{n_1}{n_1} - \frac{1}{N_1} \right) \right] \]

For the calculation of \( \theta_{min} \) we consider two cases. 

Case 1: \( 0 < \rho < 1 \)

\[V(\theta) = V(\hat{Y} - \hat{X})_{LMVU}; \quad \frac{dV(\theta)}{d\theta} = 0 \text{ gives} \]

\[\theta = k^2 \sqrt{1 - \xi}; \quad k = \frac{Q_1^2 + Q_2^2 - 2Q_1 Q_2 \rho}{(Q_1^2 + Q_2^2) \rho - 2Q_1 Q_2 \rho} \]

And \( \xi = \frac{1}{\rho^2} \) (4.1.1)

It can be proved that \( V(\theta) < V(1) \) when \( 0 < \theta < 1 \) and \( 0 < \rho < 1 \).

Then it follows that

a. If there exist a solution \( \theta \in [0,1] \text{ then } \theta = \theta_{min} \), which gives \( V_{opt}(\hat{Y} - \hat{X})_{LMVU} \).

b. If there does not exist such a solution, then \( \theta_{min} = 0 \).

Case 2: \( -1 < \rho < 0 \)

According to (4.1.1) \( \theta = k^2 \sqrt{1 - \xi} \)

It can be proved that \( V(\theta) > V(0) \) when \( 0 < \theta < 1 \) and \( -1 < \rho < 0 \).

Then it follows that

a. If there exist a solution \( \theta \in [0,1] \text{ then } \theta = \theta_{min} \), which gives \( V_{opt}(\hat{Y} - \hat{X})_{LMVU} \).

b. If there does not exist such a solution, then \( \theta_{min} = 1 \).

g. Estimator of ratio between totals

\[\hat{R}_{MV} = \frac{\hat{Y}_{LMVU}}{\hat{X}_{LMVU}} = \]

\[\frac{Y_{2} + \hat{Y}_{12u} + \frac{(1-\theta)}{(1-\rho^2)} \left[ \hat{Y}_{12m} - \hat{Y}_{12u} + \hat{X}_{12m} - \hat{X}_{12u} \right]}{X_{1} + \hat{X}_{12} + \frac{(1-\theta)}{(1-\rho^2)} \left[ \hat{Y}_{12m} - \hat{Y}_{12u} + \hat{X}_{12m} - \hat{X}_{12u} \right]} \]

\[V(\hat{R}_{MV}) = \frac{1}{\hat{X}^2} \left[ N_2^2s_2^2 \left( \frac{1}{n_2} - \frac{1}{N_2} \right) + N_1^2s_1^2 \left( \frac{1}{n_1} - \frac{1}{N_1} \right) - \frac{n_2^2s_2^2}{N_{12}} \left( \frac{n_1}{n_1} - \frac{1}{N_1} \right) \right] + \]

\[+ \frac{N_2^2s_2^2}{n_{12}} \left[ \frac{(1+R^2) - 2R \rho}{(1-\rho^2)} \right] \]

For the calculation of \( \theta_{min} \) we here only consider \( 0 < \rho < 1 \).

The minimum variance of \( \hat{R}_{MV} \) is obtained below by minimizing \( V(\hat{R}_{MV}) \) with respect to \( \theta \). This leads to a second degree equation of \( \theta \).

Let \( R > 0, 0 < \rho < 1 \) and \( V(\theta) = V(\hat{R}_{MV}) \).

Then \( \frac{dV(\theta)}{d\theta} = 0 \) leads to

\[\theta = k^2 \sqrt{1 - \xi}; \quad k = \frac{1+R^2 - 2R \rho}{(1+R^2 - 2R \rho)} \text{ and } \xi = \frac{1}{\rho^2} \]

It can be proved that \( V(\theta) < V(1) \) when \( 0 < \theta < 1 \) and \( 0 < \rho < 1 \). Then we can calculate optimum values of \( \theta \).

a. If there exist a solution \( \theta \in [0,1] \text{ then } \theta = \theta_{min} \), which gives \( V_{opt}(\hat{R}_{MV}) \).

b. If there does not exist such a solution, then \( \theta_{min} = 1 \).
4.2 Design comparisons

The variances of the LMVU-estimators of the design described above have been compared to the variances of the estimators of the same parameters under a simple design. This design is at \( t_1 \) a SRS of \( n \) units drawn from the \( N \) units of \( U \) and, at \( t_2 \), a SRS of \( n \) units from the \( N' \) units of \( U' \), independent of the sample at \( t_1 \). The following simplifying assumptions have been done: \( Q_1 = Q_2 = Q, 0 < p < 1, n_1 = n_2, n_2 = N_2 / N \) and \( n_1 / n = N_1 / N \). The population variances of \( U, U', U_1, U_12 \) and \( U_2 \) are assumed equal and the sampling fractions are assumed ignorable.

The estimators for the simple design and their variances are shown in Forsman and Garås (1982). In table 1 (levels) and table 2 (differences) the reductions of variances by using the matched design described in section 2 instead of the simple design, are expressed as the ratio

\[
Z = \frac{V_{opt}(LMVU-estimator, matched sample design)}{V( estimator, simple design)}
\]

Table 1: \( Z = 1 - Q_0 \) (levels)

<table>
<thead>
<tr>
<th>( p )</th>
<th>0.99</th>
<th>0.80</th>
<th>0.60</th>
<th>0.40</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.57</td>
<td>0.80</td>
<td>0.90</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>0.99</td>
<td>0.57</td>
<td>0.80</td>
<td>0.90</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>0.90</td>
<td>0.61</td>
<td>0.82</td>
<td>0.91</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>0.80</td>
<td>0.66</td>
<td>0.84</td>
<td>0.92</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>0.70</td>
<td>0.70</td>
<td>0.86</td>
<td>0.93</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>0.60</td>
<td>0.74</td>
<td>0.88</td>
<td>0.94</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>0.50</td>
<td>0.79</td>
<td>0.90</td>
<td>0.95</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 2: \( Z = (1 - Q) + Q \frac{1 + \sqrt{1 - p^2}}{2} \) (differences)

<table>
<thead>
<tr>
<th>( p )</th>
<th>0.99</th>
<th>0.80</th>
<th>0.60</th>
<th>0.40</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.20</td>
<td>0.40</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td>0.99</td>
<td>0.02</td>
<td>0.21</td>
<td>0.41</td>
<td>0.60</td>
<td>0.80</td>
</tr>
<tr>
<td>0.90</td>
<td>0.11</td>
<td>0.28</td>
<td>0.46</td>
<td>0.64</td>
<td>0.82</td>
</tr>
<tr>
<td>0.80</td>
<td>0.21</td>
<td>0.36</td>
<td>0.52</td>
<td>0.68</td>
<td>0.84</td>
</tr>
<tr>
<td>0.70</td>
<td>0.31</td>
<td>0.44</td>
<td>0.58</td>
<td>0.72</td>
<td>0.86</td>
</tr>
<tr>
<td>0.60</td>
<td>0.41</td>
<td>0.52</td>
<td>0.64</td>
<td>0.76</td>
<td>0.88</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0.60</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
</tr>
</tbody>
</table>

The results are following:

Large variance reductions can be achieved with the matched sample strategy for the estimators of change, even when the population is changeable. For the estimators of levels there are also variance reductions when using the matched sample strategy, but they are smaller.

5. Concluding remarks

The variances of the LMVU-estimators for levels and differences and of the MV-estimators for ratios are minimized with respect to the relative size of the overlapping part of the samples at \( t_1 \) and \( t_2 \). The optimum size of this part differs depending on whether the estimation concerns levels or changes. In practice the size chosen is a matter of judgment. As for the comparisons made between the sampling design and a simpler one with independent samples a numerical example shows that considerable variance reductions are possible. This is true even if the change between \( t_1 \) and \( t_2 \) is as large as 10 to 20 percent provided that the correlation coefficient in the overlapping part of the population is large. The variance reduction is especially large for estimation of change.

6. References