Discussion Paper

Central Bureau of Statistics, P.B. 8131 Dep, 0033 Oslo 1, Norway

No. 3 20 March 1985

ON THE PREDICTION OF POPULATION TOTALS FROM SAMPLE SURVEYS BASED ON ROTATING PANELS

BY

ERIK BIØRN

ABSTRACT

The paper deals with the prediction (estimation) of the aggregate value of a variable on the basis of micro data from partly overlapping samples. This problem is of considerable interest for economic data, e.g. household budget data. We are particularly concerned with the interplay between the sampling design (degree of rotation) and the covariance structure of the data vector in a situation where the micro data are generated by a variance components mechanism with two components, one of which represents unobserved individual factors. The optimal choice of predictor is discussed, both with respect to the level of the variable under consideration and with respect to its change between two successive periods.

ON THE PREDICTION OF POPULATION TOTALS FROM SAMPLE SURVEYS BASED ON ROTATING PANELS *)

CONTENTS

*) This paper was written out during my leave of absence from the Central Bureau of Statistics 1983-1984, for teaching at the Department of Economics, University of Oslo. I am grateful to Tore Schweder for useful comments.

1. INTRODUCTION

The prediction of population totals on the basis of data from sample surveys is a problem of considerable practical interest in statistics and econometrics. Frequently the problem posed is that of predicting the aggregate value of a variable y in a period t from observations on y from a sample survey performed in this period. A more interesting problem may be to predict the aggregate *change* in y from period t_o to period t_1 on the basis of sample survey data collected in these two periods.

An econometrician facing such problems will often be in the situation that he has some a priori information on the mechanism generating the data. To him it may seem unrealistic to assume, as sampling statisticians often do, that all y's in a given period are generated by the same probability distribution. On the contrary, from economic theory he may have the notion of a *model* generating the different y values - both those observed and those unobserved - and he wants to utilize this information when making predictions of the population totals. Stated in sampling theoretic terms, he may want to combine *"design-based"* and *"model-based"* inference; confer e.g. Royall (1970), and Cassel, Särndal, and Wretman (1979).

In this paper, we shall be particularly concerned with a model in which y is determined by a *variance components mechanism*, i.e. we allow for unobserved, individual, random effects in the model specification. Within this framework, we shall consider two situations; that in which y is related to an observable exogenous variable x through a linear regression equation, and that in which no such relationship exists. Regression models with variance components specifications of the disturbance terms have received increasing interest in econometric research based on panel data in recent years, but as far as the author knows, little attention has been paid to their implications for prediction in sample survey contexts. The salient feature of this specification is that the covariance structure of the data vector will depend on the choice of sampling design. Hence, the sampling design becomes a crucial element in the construction of the optimal predictor of the aggregate variable y. Of course, this simple model has to be modified to be useful in practical situations, but it serves to illustrate the main points of interest.

The sampling design we shall consider is a design with *partly overlapping samples*, or rotating samples, between periods. (For a formal and fairly general treatment of such data structures and their relation to complete cross-section/time-series (panel) data, see Biørn (1981).) In particular, we shall focus on a situation where two periods are involved and in which some individuals are observed in the first period only, some are observed in the second period only, and some are observed in both periods. A main motivation for considering this particular data structure - but of course not the only one - is a desire to explore the possibilities for a more systematic utilization of the Norwegian household budget surveys for prediction purposes. From the year 1975, these surveys have been performed annually, using a sampling design of the format described above about 25 cent of the respondents in one year are asked to report their consumption expenditures again in the next year. The "predictions" we have in mind include (a) calculation of annual changes in the aggregate expenditures on the different consumption items for national accounting purposes, and (b) estimation of the annual changes in the vector of budget shares used as weights in the Consumer Price Index.

2. NOTATION, MODEL AND SAMPLING DESIGN

Consider a *population* of H individuals numbered consecutively from 1 to H. Let $P = \{1,2,...,H\}$. In each period, a sample of individuals, i.e. a subset of elements in the index set P, is drawn from this population. The samples are *partly overlapping* between periods, but no individual is observed more than twice. Let $Z_t \subset P$ be the sample selected in period t. These assumptions imply that

$$S_{t,t+1} = Z_t \cap Z_{t+1}$$

is non-empty, whereas $Z_t \cap Z_{t+\Theta}$ is empty for all $\Theta > 1$ or $\Theta < -1$. Let, moreover, S_t be the individuals among those selected in period t which are observed only once. It follows that Z_t can be expressed as the union of three disjoint sets as

(2.1) $$Z_t = S_{t-1,t} \cup S_t \cup S_{t,t+1},$$

where $S_{t-1,t}$ contains the individuals observed in periods t-1 and t, $S_{t,t+1}$ those observed in periods t and t+1, and S_t those observed in period t only. Finally, let Z_t^* represent the individuals not observed in period t, i.e. $Z_t \cup Z_t^* = P$, and S^* those not observed in any of the periods under consideration, 1,2,...,T, i.e.

(2.2) $$S^* = Z_1^* \cap Z_2^* \cap ... \cap Z_T^*.$$

We want to make inferences on the variable y. Its value for individual h in period t, y_{ht}, is assumed to be generated by the following process

(2.3) $$y_{ht} = a_{ht} + \mu_h + \nu_{ht},$$

where a_{ht} is a non-stochastic and (so far) unspecified parameter and μ_h and ν_{ht} are independent stochastic variables, with zero expectations and constant variances, equal to σ_μ^2 and σ_ν^2, respectively. Hence,

(2.4) $$E(y_{ht}) = a_{ht},$$

(2.5) $$E(\mu_h) = E(\nu_{ht}) = 0,$$

$$(2.6)\quad \begin{cases} E(\mu_h \mu_{h'}) = \delta_{hh'} \sigma_\mu^2, \\ E(\mu_h \nu_{h't}) = 0, \\ E(\nu_{ht} \nu_{h't'}) = \delta_{hh'} \delta_{tt'} \sigma_\nu^2, \end{cases}$$

where $\delta_{hh'} = 1$ for $h' = h$, 0 for $h' \neq h$; and $\delta_{tt'} = 1$ for $t' = t$, 0 for $t' \neq t$. The model is thus a variance components model with two components, the first, μ_h, representing unobservable factors which are specific to individual h, and ν_{ht} is a remainder.

We assume that the above specification applies to all the H individuals in the population in T successive periods, i.e. (2.3)-(2.6) are valid for

$$h, h' = 1, 2, \ldots, H,$$

$$t, t' = 1, 2, \ldots, T.$$

Letting ε_{ht} denote the composite disturbance,

$$(2.7)\quad \varepsilon_{ht} = \mu_h + \nu_{ht},$$

an equivalent way of writing the model is

$$(2.8)\quad E(y_{ht}) = a_{ht},$$

$$(2.9)\quad \operatorname{cov}(y_{ht}, y_{h't'}) = E(\varepsilon_{ht} \varepsilon_{h't'}) = \begin{cases} \sigma^2 & \text{for } h'=h,\ t'=t \\ \rho\sigma^2 & \text{for } h'=h,\ t' \neq t \\ 0 & \text{otherwise,} \end{cases}$$

where $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$, and $\rho = \sigma_\mu^2/\sigma^2$. The presence of the individual specific disturbance component implies that all observations on y from the same individual are positively correlated, with a coefficient of correlation equal to ρ.

Our main problem in the following will be to predict the total value of y in the population in period t, i.e.

$$(2.10)\quad Y_t = \sum_{h=1}^{H} y_{ht} \qquad t = 1, \ldots, T,$$

and its change

(2.11) $$\Delta Y_t = \sum_{h=1}^{H} \Delta y_{ht},$$

where $\Delta y_{ht} = y_{ht} - y_{h,t-1}$, on the basis of the values of y_{ht} observed in the different samples, i.e. from the observation sets

$$y_{ht}, \qquad h \in Z_t, t = 1,\ldots,T.$$

Let n_t denote the number of individuals in the sub-sample S_t and $n_{t,t+1}$ the number of elements in $S_{t,t+1}$. The total number of individuals included in the sample in period t is thus

(2.12) $$N_t = n_{t-1,t} + n_t + n_{t,t+1}.$$

We shall consider two specifications of the unknown parameters a_{ht}:

Model I : $a_{ht} = a_t$ *for* $h=1,\ldots,H;\ t=1,\ldots,T,$

where a_t *are unknown constants.*

Model II: a_{ht} *is linearly related to an observable variable* x_{ht}.

Model I will be discussed in sections 3 and 4, and model II in sections 5 and 6.

Moreover, to simplify the exposition, we shall confine attention to the situation with only *two periods* involved, i.e. T = 2, and with the sets S_{01} and S_{23} empty, i.e. $n_{01} = n_{23} = 0$. Then $S^* = Z_1^* \cap Z_2^*$ is the index set of the individuals not observed and

(2.13) $$m = H - n_1 - n_{12} - n_2 = H - N_1 - N_2 + n_{12}$$

the number of these individuals. Our data set thus has the following structure:

n_1 individuals in subset S_1 are observed in period 1 only.

n_{12} individuals in subset S_{12} are observed in both periods 1 and 2.

n_2 individuals in subset S_2 are observed in period 2 only.

m individuals in subset S^* are unobserved.

3. ESTIMATION AND PREDICTION

MODEL I: CONSTANT EXPECTATIONS

3.1 The aggregate variables and their distribution

Let $\bar{Y}_t$ be the average value of y in the population in period t,

$$(3.1)\qquad \bar{Y}_t = \frac{1}{H}\sum_{h=1}^{H} y_{ht} = \frac{Y_t}{H} \qquad (t = 1,2),$$

and

$$(3.2)\qquad \bar{Y}_t(S_i) = \frac{1}{n_i}\sum_{h\in S_i} y_{ht} \qquad (t = 1,2,\ i = 1,2,12)$$

the corresponding averages in the samples S_1, S_2, and S_{12}. By assumption, $\bar{Y}_1(S_1)$, $\bar{Y}_1(S_{12})$, $\bar{Y}_2(S_{12})$, and $\bar{Y}_2(S_2)$ are observable, and $\bar{Y}_1(S_2)$, $\bar{Y}_2(S_1)$ are unobservable. Similarly,

$$(3.3)\qquad \bar{Y}_t(S^*) = \frac{1}{m}\sum_{h\in S^*} y_{ht} \qquad (h = 1,\ldots,H;\ t = 1,2)$$

is the average value in period t for the individuals which are not observed in either period. Obviously

$$(3.4)\qquad H\bar{Y}_t = n_1\bar{Y}_t(S_1)+n_{12}\bar{Y}_t(S_{12})+n_2\bar{Y}_t(S_2)+m\bar{Y}_t(S^*)\ (t=1,2).$$

When the expectation of y_{ht} is assumed to be the same for all individuals in period t, i.e.

$$(3.5)\qquad E(y_{ht}) = a_{ht} = a_t \qquad (h=1,\ldots,H;t=1,2),$$

it follows from (2.3) and (3.1)-(3.3) that

$$(3.6)\qquad \bar{Y}_t = a_t + \bar{\mu} + \bar{\nu}_t,$$

$$(3.7)\qquad \bar{Y}_t(S_i) = a_t + \bar{\mu}(S_i) + \bar{\nu}_t(S_i),$$

$$(3.8)\qquad \bar{Y}_t(S^*) = a_t + \bar{\mu}(S^*) + \bar{\nu}_t(S^*) \qquad (i = 1,2,12;\ t = 1,2),$$

where

$$(3.9) \qquad \bar{\mu} = \frac{1}{H} \sum_{h=1}^{H} \mu_h ,$$

$$(3.10) \qquad \bar{\nu}_t = \frac{1}{H} \sum_{h=1}^{H} \nu_{ht} ,$$

$$(3.11) \qquad \bar{\mu}(S_i) = \frac{1}{n_i} \sum_{h \in S_i} \mu_h ,$$

$$(3.12) \qquad \bar{\nu}_t(S_i) = \frac{1}{n_i} \sum_{h \in S_i} \nu_{ht} ,$$

$$(3.13) \qquad \mu(S^*) = \frac{1}{m} \sum_{h \in S^*} \mu_h ,$$

$$(3.14) \qquad \bar{\nu}_t(S^*) = \frac{1}{m} \sum_{h \in S^*} \nu_{ht} .$$

Using (2.4)-(2.6), we find that $\bar{Y}_t(S_i)$ and $\bar{Y}_t(S^*)$ have expectations

$$(3.15) \qquad E[\bar{Y}_t(S_i)] = E[Y_t(S^*)] = a_t \qquad (i = 1,2,12;\ t = 1,2),$$

and variances and covariances given by

$$(3.16) \qquad \mathrm{cov}[\bar{Y}_t(S_i), \bar{Y}_\tau(S_j)] = \begin{cases} \dfrac{\sigma_\mu^2 + \sigma_\nu^2}{n_i} = \dfrac{\sigma^2}{n_i} & \text{for } j = i,\ \tau = t \\ \dfrac{\sigma_\mu^2}{n_i} = \rho \dfrac{\sigma^2}{n_i} & \text{for } j = i,\ \tau \neq t \\ 0 & \text{otherwise,} \end{cases}$$

$$(3.17) \qquad \mathrm{cov}[\bar{Y}_t(S^*), \bar{Y}_\tau(S^*)] = \begin{cases} \dfrac{\sigma_\mu^2 + \sigma_\nu^2}{m} = \dfrac{\sigma^2}{m} & \text{for } \tau = t \\ \dfrac{\sigma_\mu^2}{m} = \rho \dfrac{\sigma^2}{m} & \text{for } \tau \neq t, \end{cases}$$

$$(3.18) \qquad \mathrm{cov}[\bar{Y}_t(S_i), \bar{Y}_\tau(S^*)] = 0 \qquad (i,j = 1,2,12;\ t,\tau = 1,2).$$

3.2 Estimation

In the case considered here, nothing is known *a priori* about a_1 and a_2 (or their possible relationship). Since, however, σ^2 and ρ (σ_μ^2 and σ_ν^2) are common parameters in the disturbance structure of all observations, it will be more efficient to estimate the four parameters simultaneously from the combined data set with $n_1 + 2n_{12} + n_2$ observations than estimating a_1 from the observations from period 1 and a_2 from the observations from period 2.

Assume that μ_h and ν_{ht} are *normally* distributed. Let $\varepsilon_{(1)}$ be the n_1x1 vector of disturbances from the n_1 individuals observed in period 1 only, $\varepsilon_{(2)}$ the n_2x1 vector of disturbances from the n_2 individuals observed in period 2 only, and $\varepsilon_{(12)}$ the $2n_{12}$x1 vector of disturbances from the n_{12} individuals observed in both periods, ordered first by individual, second by period. It follows from (2.9) that the covariance matrix of the stacked vector

$$(3.19) \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_{(1)} \\ \varepsilon_{(12)} \\ \varepsilon_{(2)} \end{bmatrix}$$

can be written as[1)]

$$(3.20) \qquad E(\varepsilon\varepsilon') = \Omega = \sigma^2\Omega_*,$$

where

$$(3.21) \qquad \Omega_* = \begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_{12}} \otimes F_2 & 0 \\ 0 & 0 & I_{n_2} \end{bmatrix},$$

I_{n_i} being the n_i x n_i identity matrix and $F_2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

Expressing (2.3) and (2.7) in vector notation as $\underset{\sim}{y} = \underset{\sim}{a} + \underset{\sim}{\varepsilon}$, we can write the log-likelihood function of $\underset{\sim}{y}$ as

$$L = -\frac{n_1 + 2n_{12} + n_2}{2} \log (2\pi) - \tfrac{1}{2} \log |\Omega| - \tfrac{1}{2} \underset{\sim}{\varepsilon}'\Omega^{-1}\underset{\sim}{\varepsilon} ,$$

where $\underset{\sim}{\varepsilon}$ is a shorthand notation for $\underset{\sim}{y} - \underset{\sim}{a}$.

Since $|\Omega| = |\sigma^2\Omega_*| = \sigma^{2(n_1 + 2n_{12} + n_2)}(1-\rho^2)^{n_{12}}$ and $F_2^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$,

L can be written as

$$(3.22) \quad L = L(\underset{\sim}{y};\underset{\sim}{a},\rho,\sigma^2) = -\frac{n_1 + 2n_{12} + n_2}{2}\log(2\pi)$$

$$-\frac{n_1 + 2n_{12} + n_2}{2}\log\sigma^2 - \frac{n_{12}}{2}\log(1-\rho^2) - \frac{1}{2}\sigma^{-2}Q,$$

where

$$(3.23) \quad Q = \varepsilon'\Omega_*^{-1}\varepsilon$$

$$= \varepsilon'_{(1)}\varepsilon_{(1)} + \varepsilon'_{(12)}\{I_{n_{12}} \otimes F_2^{-1}\}\varepsilon_{(12)} + \varepsilon'_{(2)}\varepsilon_{(2)}$$

$$= \sum_{h\in S_1}\varepsilon_{h1}^2 + \frac{1}{1-\rho^2}\sum_{h\in S_{12}}\{\varepsilon_{h1}^2 - 2\rho\varepsilon_{h1}\varepsilon_{h2} + \varepsilon_{h2}^2\} + \sum_{h\in S_2}\varepsilon_{h2}^2 .$$

Maximum Likelihood (ML) estimates of a_1, a_2, ρ, and σ^2 can be obtained (provided that certain regularity constraints are satisfied) by an algorithm which switches between the following two subproblems:

(i) *Minimization of Q with respect to a_1 and a_2, conditionally on ρ and σ^2 (i.e. conditional Generalized Least Squares (GLS) estimation).*

(ii) *Minimization of $g = (n_1 + 2n_{12} + n_2)\log\sigma^2 + n_{12}\log(1-\rho^2) + \sigma^2 Q$ with respect to ρ and σ^2, conditionally on a_1 and a_2.*

It can be shown[2)] that *subproblem (i)* is solved by minimizing the following sum of squares

$$Q(1-\rho) = \sum_{h\in S_1}\{(1-\rho)^{\frac{1}{2}}\varepsilon_{h1}\}^2 + \sum_{h\in S_2}\{(1-\rho)^{\frac{1}{2}}\varepsilon_{h2}\}^2$$

$$+ \sum_{h\in S_{12}}\left[\{\varepsilon_{h1} - (1-(\tfrac{1-\rho}{1+\rho})^{\frac{1}{2}})\frac{\varepsilon_{h1}+\varepsilon_{h2}}{2}\}^2 + \{\varepsilon_{h2} - (1-(\tfrac{1-\rho}{1+\rho})^{\frac{1}{2}})\frac{\varepsilon_{h1}+\varepsilon_{h2}}{2}\}^2\right].$$

Subproblem (ii) involves solution of the following two nonlinear equations in σ^2 and ρ:

$$(1-\rho)\sigma^2(n_1 + 2n_{12} + n_2) = (1-\rho)\left[\sum_{h\in S_1}\varepsilon_{h1}^2 + \sum_{h\in S_2}\varepsilon_{h2}^2\right]$$

$$+ \sum_{h\in S_{12}}\left[\varepsilon_{h1}^2 + \varepsilon_{h2}^2 - \frac{1}{2}(1 - \frac{1-\rho}{1+\rho})(\varepsilon_{h1}+\varepsilon_{h2})^2\right],$$

$$\sigma^2\{n_1 + n_2 + (1+\rho)^{-1}2n_{12}\} = \sum_{h\in S_1}\varepsilon_{h1}^2 + \sum_{h\in S_2}\varepsilon_{h2}^2 + (1+\rho)^{-2}\sum_{h\in S_{12}}(\varepsilon_{h1}+\varepsilon_{h2})^2.$$

Let the estimates be denoted as $\hat{a}_1$, $\hat{a}_2$, $\hat{\rho}$, and $\hat{\sigma}^2$.

3.3 Prediction

Having obtained estimates of a_1, a_2 and ρ, we now proceed to the problem of predicting the population totals Y_1 and Y_2 and its increase from period 1 to period 2, $\Delta Y = Y_2 - Y_1$. We shall consider two different ways of attacking this problem:

(A) Direct prediction based on the observed values of y_{ht} and the estimate of ρ.

(B) Prediction utilizing not only the observed y_{ht} and the estimated value of ρ, but also the estimates of a_1 and a_2.

Both procedures emerge as special cases of the following linear prediction formulae:

$$(3.24) \quad \begin{cases} \hat{Y}_1 = v_{11}\bar{Y}_1(S_1) + v_{12}\bar{Y}_1(S_{12}) + v_{1*}\hat{a}_1, \\ \hat{Y}_2 = v_{21}\bar{Y}_2(S_{12}) + v_{22}\bar{Y}_2(S_2) + v_{2*}\hat{a}_2 , \end{cases}$$

where the v's are suitably defined weights. In case A, v_{1*} and v_{2*} are set equal to zero a priori; in case B, all weights are positive. The corresponding predictor of ΔY is

$$(3.25) \quad \hat{\Delta Y} = v_{22}\bar{Y}_2(S_2) - v_{11}\bar{Y}_1(S_1) + v_{21}\bar{Y}_2(S_{12}) - v_{12}\bar{Y}_1(S_{12}) + v_{2*}\hat{a}_2 - v_{1*}\hat{a}_1 .$$

Of course, the distinction between procedures (A) and (B) is of no interest if $\hat{a}_1$ is a linear function of the y's observed in period 1 and $\hat{a}_2$ is a linear function of the y's observed in period 2. This will for instance be the case if $\mu_h=0$ for all individuals, since then the ML estimates are simply the unweighted sample averages

$$\hat{a}_1 = (n_1\bar{Y}_1(S_1) + n_{12}\bar{Y}_1(S_{12}))/(n_1 + n_{12}),$$

$$\hat{a}_2 = (n_{12}\bar{Y}_2(S_{12}) + n_2\bar{Y}_2(S_2))/(n_{12} + n_2).$$

But if individual components are present, this distinction is highly relevant, as we shall see below.

Using (3.7), the three predictors can be reformulated as

$$(3.26)\quad \begin{cases} \hat{Y}_1 = (v_{11} + v_{12} + v_{1*})a_1 + v_{1*}(\hat{a}_1 - a_1) + U_1 , \\ \hat{Y}_2 = (v_{21} + v_{22} + v_{2*})a_2 + v_{2*}(\hat{a}_2 - a_2) + U_2 , \end{cases}$$

$$(3.27)\quad \widehat{\Delta Y} = (v_{21} + v_{22} + v_{2*})a_2 - (v_{11} + v_{12} + v_{1*})a_1 + v_{2*}(\hat{a}_2 - a_2) - v_{1*}(\hat{a}_1 - a_1) + U_2 - U_1,$$

where

$$(3.28)\quad \begin{cases} U_1 = v_{11}\{\bar{\mu}(S_1) + \bar{\nu}_1(S_1)\} + v_{12}\{\bar{\mu}(S_{12}) + \bar{\nu}_1(S_{12})\} , \\ U_2 = v_{21}\{\bar{\mu}(S_{12}) + \bar{\nu}_2(S_{12})\} + v_{22}\{\bar{\mu}(S_2) + \bar{\nu}_2(S_2)\} . \end{cases}$$

Since the ML estimates $\hat{a}_1$ and $\hat{a}_2$ are unbiased, it follows that the condition for the predictors to be unbiased is

$$(3.29)\quad v_{11} + v_{12} + v_{1*} = v_{21} + v_{22} + v_{2*} = H.$$

We shall discuss case A and B in turn.

Case A: $v_{1*} = v_{2*} = 0$

Let $v_{1*} = v_{2*} = 0$ and define

$$(3.30)\quad k_1 = v_{11}/H, \qquad k_2 = v_{22}/H;$$

i.e. k_1 and $1-k_1$ are the relative weights assigned to observations from individuals observed once and twice, respectively, when making predictions for period 1; and k_2 and $1-k_2$ are the corresponding weights for period 2. Using (3.1), (3.4), and (3.29), we find that the *prediction errors* of Y_1 and Y_2 can be written as

$$(3.31)\quad \begin{cases} \delta_1 = \hat{Y}_1 - Y_1 = \{k_1H - n_1\}\bar{Y}_1(S_1) + \{(1-k_1)H - n_{12}\}\bar{Y}_1(S_{12}) \\ \qquad - n_2\bar{Y}_1(S_2) - m\bar{Y}_1(S^*) , \\ \delta_2 = \hat{Y}_2 - Y_2 = \{k_2H - n_2\}\bar{Y}_2(S_2) + \{(1-k_2)H - n_{12}\}\bar{Y}_2(S_{12}) \\ \qquad - n_1\bar{Y}_2(S_1) - m\bar{Y}_2(S^*) . \end{cases}$$

From (3.16)-(3.18) and (2.13) it follows that their variances are

$$(3.32) \qquad \text{var } \delta_1 = \sigma^2 H \left[\frac{Hk_1^2}{n_1} + \frac{H(1-k_1)^2}{n_{12}} - 1 \right] = V_1,$$

$$(3.33) \qquad \text{var } \delta_2 = \sigma^2 H \left[\frac{Hk_2^2}{n_2} + \frac{H(1-k_2)^2}{n_{12}} - 1 \right] = V_2,$$

and that they have a covariance equal to

$$(3.34) \qquad \text{cov } (\delta_1, \delta_2) = \sigma^2 \rho H \left[\frac{H(1-k_1)(1-k_2)}{n_{12}} - 1 \right].$$

If ρ is positive, the prediction errors will have positive, zero, and negative correlation according as $H(1-k_1)(1-k_2) \gtreqless n_{12}$.

We are also interested in the prediction error of $\hat{\Delta Y}$,

$$(3.35) \qquad \delta_\Delta = \hat{\Delta Y} - \Delta Y = (\hat{Y}_2 - Y_2) - (\hat{Y}_1 - Y_1) = \delta_2 - \delta_1.$$

Its variance is

$$(3.36) \qquad \text{var } \delta_\Delta = \text{var } \delta_1 + \text{var } \delta_2 - 2 \text{ cov } (\delta_1, \delta_2)$$

$$= \sigma^2 H \left[\frac{Hk_1^2}{n_1} + \frac{Hk_2^2}{n_2} - 2(1-\rho) + \frac{H}{n_{12}} \{(1-k_1)^2 - 2\rho(1-k_1)(1-k_2) + (1-k_2)^2\} \right]$$

$$= V_\Delta.$$

We see that the variances of the prediction errors δ_1 and δ_2 are functions of the population size H, the sample sizes n_1, n_2, and n_{12}, and the relative weights k_1 and k_2. The variance of δ_Δ also depends on ρ, the share of the disturbance variance which is due to individual variations. This has notable implications for the optimal choice of predictor, as we shall see in section 4.

Case B: $v_{1*}, v_{2*} > 0$

When we also utilize the estimated values of a_1 and a_2 in constructing the predictors, we find from (2.13), (3.4), (3.7), (3.8), (3.26), (3.28), and (3.29) that the prediction errors become

$$(3.37)\quad d_1 = \hat{Y}_1 - Y_1 = v_{1*}(\hat{a}_1 - a_1) + U_1 - H(\bar{\mu} + \bar{\nu}_1)$$

$$= v_{1*}(\hat{a}_1 - a_1) + (v_{11} - n_1)\{\bar{\mu}(S_1) + \bar{\nu}_1(S_1)\}$$

$$+ (v_{12} - n_{12})\{\bar{\mu}(S_{12}) + \bar{\nu}_1(S_{12})\}$$

$$- n_2\{\bar{\mu}(S_2) + \bar{\nu}_1(S_2)\} - m\{\bar{\mu}(S^*) + \bar{\nu}_1(S^*)\},$$

$$(3.38)\quad d_2 = \hat{Y}_2 - Y_2 = v_{2*}(\hat{a}_2 - a_2) + U_2 - H(\bar{\mu} + \bar{\nu}_2)$$

$$= v_{2*}(\hat{a}_2 - a_2) + (v_{21} - n_{12})\{\bar{\mu}(S_{12}) + \bar{\nu}_2(S_{12})\}$$

$$+ (v_{22} - n_2)\{\bar{\mu}(S_2) + \bar{\nu}_2(S_2)\}$$

$$- n_1\{\bar{\mu}(S_1) + \bar{\nu}_2(S_1)\} - m\{\bar{\mu}(S^*) + \bar{\nu}_2(S^*)\}.$$

Three sources of prediction errors can be discerned in this case. The first is errors in the estimates $\hat{a}_1$ and $\hat{a}_2$; its contribution to the total error depends on the weights v_{1*} and v_{2*}. The second source is the disturbances of the $n_1 + n_{12}$, resp. $n_2 + n_{12}$, individuals included in the samples. This component can be controlled by changing either the weights or the sampling design. Thirdly we have the disturbances of the individuals which are not observed in the period under consideration. This component cannot be controlled by changing the weighting system, it can only be affected by the sampling design.

Since the estimates $\hat{a}_1$ and $\hat{a}_2$ are based on the y_{ht} values in the samples S_1, S_2, and S_{12}, they will be correlated with the error components in (3.37) and (3.38). [3] The derivation of general expressions for the variances of d_1 and d_2 would thus involve rather messy algebra. In the following, we shall, for simplicity, neglect the first source of prediction error by letting $\hat{a}_t = a_t$ $(t=1,2)$. This, of course, implies that we *proceed as if the common non-stochastic part of y_{ht} were known with certainty for all individuals.* The variances of the prediction errors then become

(3.39) $$\operatorname{var} d_1 = \sigma^2[\frac{1}{n_1}(n_1-v_{11})^2 + \frac{1}{n_{12}}(n_{12}-v_{12})^2 + n_2 + m]$$

$$= \sigma^2[\frac{v_{11}^2}{n_1} + \frac{v_{12}^2}{n_{12}} + H-2v_{11}-2v_{12}] = W_1,$$

(3.40) $$\operatorname{var} d_2 = \sigma^2[\frac{1}{n_2}(n_2-v_{22})^2 + \frac{1}{n_{12}}(n_{12}-v_{21})^2 + n_1 + m]$$

$$= \sigma^2[\frac{v_{22}^2}{n_2} + \frac{v_{21}^2}{n_{12}} + H-2v_{22}-2v_{21}] = W_2,$$

and their covariance is

(3.41) $$\operatorname{cov}(d_1,d_2) = \sigma^2\rho\,[\frac{(n_{12}-v_{12})(n_{12}-v_{21})}{n_{12}} + (n_1-v_{11})+(n_2-v_{22})+ m]$$

$$= \sigma^2\rho\,[\frac{v_{12}v_{21}}{n_{12}} + H-v_{11}-v_{12}-v_{21}-v_{22}].$$

If $\rho > 0$, this covariance is positive, zero, and negative according as

$$\frac{v_{12}v_{21}}{n_{12}} \gtreqless v_{11}+v_{12}+v_{21}+v_{22}-H = H-v_{1*}-v_{2*}.$$

The variance of the error of the predicted change, $d_\Delta = d_2-d_1$, is in this case

(3.42) $$\operatorname{var} d_\Delta = \operatorname{var} d_1 + \operatorname{var} d_2-2\operatorname{cov}(d_1,d_2)$$

$$= \sigma^2\,[\frac{v_{11}^2}{n_1} + \frac{v_{22}^2}{n_2} -2(1-\rho)(v_{11}+v_{12}+v_{21}+v_{22}-H)$$

$$+ \frac{1}{n_{12}}\{v_{12}^2-2\rho v_{12}v_{21}+v_{21}^2\}] = W_\Delta .$$

Like the corresponding variance in case A, given in (3.36), it depends in a crucial way on the individual share of the total disturbance variance.[4)]

4. OPTIMAL CHOICE OF PREDICTORS

MODEL I: CONSTANT EXPECTATIONS

Since the variances of the prediction errors depend on the weighting system as well as on the composition of the samples, an interesting problem is to find the optimal choice of these parameters, i.e. the ones that *minimize the variances*. Three problems may be defined:

(a) Determination of optimal choice of weights, given the sampling design.
(b) Determination of optimal sampling design, given the weighting system.
(c) Joint determination of optimal weighting system and sampling design.

Moreover, each problem may be discussed from the point of view of predicting Y and of predicting ΔY. We shall not be concerned with problem (b) in the following, but concentrate on (a) and touch (c) briefly.

Case A: $v_{1*}=v_{2*}=0$

From (3.32) and (3.33) it follows that V_1 and V_2 are minimized for

$$k_1=k_1^*=\frac{n_1}{n_1+n_{12}}$$

and

$$k_2=k_2^*=\frac{n_2}{n_2+n_{12}},$$

respectively. This implies, cf. (3.24) and (3.29), that each observation in period t is given the same weight, $H/(n_t+n_{12})(t=1,2)$, regardless of whether it comes from an individual which is observed once or twice.

These weights will not, however, minimize the variance of the error of the predicted change, V_Δ. From (3.36) we find that this variance is minimized for

$$k_1=k_1^{\Delta} = \frac{n_1(1-\rho)[n_{12}+n_2(1+\rho)]}{(n_1+n_{12})(n_2+n_{12})-\rho^2 n_1 n_2} ,$$

$$k_2=k_2^{\Delta} = \frac{n_2(1-\rho)[n_{12}+n_1(1+\rho)]}{(n_1+n_{12})(n_2+n_{12})-\rho^2 n_1 n_2} .$$

We see that k_t^{Δ} $(t=1,2)$ attains its maximal value, k_t^*, for $\rho = 0$ and decreases monotonically towards zero as ρ goes to 1: The larger the individual part of the disturbance variance, the larger weight should be given to observations from individuals observed twice and the smaller weight to those observed once when predicting aggregate changes.

To simplify, we now assume that the same number of individuals is observed in both periods, i.e. $n_1=n_2=n$. Let $N=n+n_{12}$ be the sample size in each period and $c=n_{12}/N$ the share of the samples which is overlapping. Then,

(4.1) $$k_1^*=k_2^*=k^* = \frac{n}{n+n_{12}} = 1-c,$$

(4.2) $$k_1^{\Delta}=k_2^{\Delta}=k^{\Delta} = \frac{n(1-\rho)}{n(1-\rho)+n_{12}} = \frac{(1-c)(1-\rho)}{(1-c)(1-\rho)+c} .$$

Values of k^* and k^{Δ} for selected combinations of c and ρ are given in table 1.

Let $V_t(k,c,N)$ and $V_{\Delta}(k,c,N)$ denote the variances V_t and V_{Δ} considered as functions of k,c, and N, i.e., from (3.32), (3.33) and 3.36),

(4.3) $$V_t(k,c,N) = \sigma^2 H\,[\,\frac{H}{N}\{\frac{k^2}{1-c} + \frac{(1-k)^2}{c}\}-1] \qquad (t = 1,2)$$

(4.4) $$V_{\Delta}(k,c,N) = 2\sigma^2(1-\rho)H[\frac{H}{N}\{\frac{k^2}{(1-c)(1-\rho)} + \frac{(1-k)^2}{c}\}-1].$$

Their minimum values are, respectively,

(4.5) $$V_t(k^*,c,N) = \sigma^2 H\,[\,\frac{H}{N} - 1\,] \qquad (t = 1,2)$$

(4.6) $$V_{\Delta}(k^{\Delta},c,N) = 2\sigma^2(1-\rho)H\,[\,\frac{H}{N}\,\frac{1}{1-\rho+\rho c} -1\,].$$

We note that the minimum value of V_t is independent of c, i.e. it is impossible, by changing the composition of the sample, to get a better prediction of the *level* of Y. The prediction of the *change* in Y, however, can be improved upon by changing the sample design; $V_\Delta(k^\Delta,c,N)$ is a decreasing function of c when ρ is positive. Thus, given the total sample size, we will obtain the best predictor of ΔY by letting c = 1, i.e. by using identical samples in the two periods. Or stated differently: Since $N(1-\rho+\rho c) = n(1-\rho) + n_{12}$, a change in the sampling design such that n is decreased by $-\Delta n$ units and n_{12} is increased by $(1-\rho)\Delta n$ units will leave V_Δ unaffected. One observation from an individual observed once has the same "value" as $(1-\rho)$ observation from an individual observed twice when predicting ΔY. The minimum variance is $V_\Delta(k^\Delta,1,N) = 2\sigma^2(1-\rho)H(H/N-1)$, which is $2(1-\rho)$ times the error variance of the optimal predictor of Y.

In the following, we shall refer to the predictors based on $k=k^*$ as the *unweighted* and those based on $k=k^\Delta$ as the *weighted* predictors, since the former gives all observations the same weight, whereas the latter does not. The *relative prediction loss* incurred by using the unweighted instead of the weighted predictor of ΔY can be expressed as

$$(4.7) \qquad \lambda=\lambda(c,\rho,\tfrac{H}{N}) = \frac{V_\Delta(k^*,c,N)}{V_\Delta(k^\Delta,c,N)} = \frac{\frac{H}{N}\cdot\frac{1-\rho c}{1-\rho} - 1}{\frac{H}{N}\cdot\frac{1}{1-\rho+\rho c} - 1} .$$

Function values of λ for H/N = 100[5] are given in table 2. We see that the loss of efficiency may be substantial. If c = 0.5 and $\rho = 0.9$, λ is larger than 3. The optimal choice of k in this case is $k^\Delta = 0.09$, whereas $k^* = 0.5$, cf. table 1. When H/N is sufficiently large, we have approximately

$$\lambda\approx\lambda'(c,\rho) = \frac{(1-\rho c)(1-\rho+\rho c)}{1-\rho},$$

where obviously $\lambda'(1-c,\rho) = \lambda'(c,\rho)$. This function attains its maximal value, $(1-\rho/2)^2/(1-\rho)$, for c = 1/2, i.e. it is when (approximately) one half of the sample is observed once and the other half is observed twice that we will obtain the largest gain by using the weighted predictor instead of the unweighted one.

We can derive a similar expression for the prediction loss of Y. The relative prediction loss obtained by using the weighted instead of the unweighted predictor of this variable is

$$(4.8) \qquad \mu=\mu(c,\rho,\frac{H}{N}) = \frac{V_t(k^{\Delta},c,N)}{V_{\Delta}(k^*,c,N)} = \frac{\frac{H}{N}\frac{(1-c)(1-\rho)^2+c}{(1-\rho+\rho c)^2} - 1}{\frac{H}{N} - 1}$$

Values of this function for H/N = 100 are given in table 3. We see that the loss of efficiency may be substantial in this case as well - in particular when ρ is large and c is small. There may thus be a conflict between the optimal choice of predictor for the level of Y and for its change, ΔY. The conflict is more likely to arise the larger is the individual share of the total error variance, ρ, and the smaller the fraction of the samples which is overlapping. The only way in which it can be resolved is by letting all individuals be observed twice (c = 1), in which case $k^*=k^{\Delta}=0$ and $\lambda=\mu=1$.

Table 1. Optimal choice of k for predicting levels (k^*) and changes (k^Δ).

Overlapping share of each sample, c	Individual share of error variance, ρ					
	0.1		0.5		0.9	
	k^*	k^Δ	k^*	k^Δ	k^*	k^Δ
0.1	0.90	0.89	0.90	0.82	0.90	0.47
0.5	0.50	0.47	0.50	0.33	0.50	0.09
0.9	0.10	0.09	0.10	0.05	0.10	0.01

Table 2. Relative prediction loss by using the unweighted instead of the weighted predictor of ΔY, $\lambda=\lambda(c,\rho,H/N)$. $H/N = 100$.

c	ρ		
	0.1	0.5	0.9
0.1	1.001	1.05	1.73
0.3	1.003	1.11	2.71
0.5	1.003	1.13	3.04
0.7	1.002	1.11	2.71
0.9	1.001	1.05	1.74

Table 3. Relative prediction loss by using the weighted instead of the unweighted predictor of Y, $\mu=\mu(c,\rho,H/N)$. $H/N = 100$.

c	ρ		
	0.1	0.5	0.9
0.1	1.001	1.08	3.04
0.3	1.002	1.13	2.26
0.5	1.003	1.20	1.68
0.7	1.002	1.07	1.32
0.9	1.001	1.03	1.09

Case B: $v_{1*}, v_{2*} > 0$

We now relax the zero restrictions on v_{1*} and v_{2*}. From (3.39) and (3.40) it follows that W_1 and W_2 are minimized for

$$(4.9)\quad \begin{cases} v_{11} = n_1, v_{12} = n_{12},\ v_{1*} = H-n_1-n_{12} = n_2 + m, \\ v_{22} = n_2, v_{21} = n_{12},\ v_{2*} = H-n_2-n_{12} = n_1 + m, \end{cases}$$

respectively. From (3.24) we see that this implies that all the individuals actually observed are represented by the observed values in the prediction formulae, whereas those not observed are represented by the (estimated) value of their common expectation.

This simple predictor will not, however, minimize the variance of the error of the predicted change. From (3.42) we find that W_Δ is minimized for

$$(4.10)\quad \begin{cases} v_{11} = n_1(1-\rho),\ v_{12} = n_{12},\ v_{1*} = H-n_1(1-\rho)-n_{12} = n_2+m+\rho n_1, \\ v_{22} = n_2(1-\rho),\ v_{21} = n_{12},\ v_{2*} = H-n_2(1-\rho)-n_{12} = n_1+m+\rho n_2. \end{cases}$$

Inserting these values in (3.25), while using (3.2) and (3.3), we find that the optimal predictor of ΔY can be written as

$$\hat{\Delta Y} = \sum_{h=1}^{H} \hat{\Delta y}_h,$$

where

$$\hat{\Delta y}_h = y_{h2} - y_{h1} \qquad h \in S_{12}$$

$$\hat{\Delta y}_h = a_2 - (\rho a_1 + (1-\rho) y_{h1}) \qquad h \in S_1$$

$$\hat{\Delta y}_h = (\rho a_2 + (1-\rho) y_{h2}) - a_1 \qquad h \in S_2$$

$$\hat{\Delta y}_h = a_2 - a_1 \qquad h \in S^*.$$

The interpretation of this is that the individuals observed twice should be represented by their observed values, whereas each observation from those observed once should be replaced by a weighted average of the observed value and its estimated expectation, with weights equal to $(1-\rho)$ and ρ, respectively. All missing observations should be represented by their estimated expectation. Thus, the larger is ρ, the less useful are the observations from individuals observed once when predicting aggregate changes.

Assume, as before, that $n_1=n_2=n$ and let $N=n+n_{12}$ and $c=n_{12}/N$. The minimum values of W_t (t=1,2) and W_Δ are then, respectively

$$W_t^{min} = \sigma^2(H-N) \qquad (t=1,2), \tag{4.11}$$

$$W_\Delta^{min} = 2\sigma^2(1-\rho)[H-N(1-\rho+\rho c)]. \tag{4.12}$$

Again, we note that the variance of the prediction error of ΔY is a decreasing function of c, and attains its minimum, $2\sigma^2(1-\rho)(H-N)$, for c=1. The minimum values (4.11) and (4.12) are less than the corresponding minima in case A, (4.5) and (4.6); their ratios are N/H and $N(1-\rho+\rho c)/H$, respectively. This is not surprising since the predictors in case B utilizes knowledge of the expectations a_1 and a_2, which the predictors in case A neglect.[6)]

Let W_t^Δ denote the value of W_t when using the weights (4.10) and, correspondingly, W_Δ^* the value of W_Δ based on the weights (4.9). From (3.39), (3.40), and (3.42) we find

$$W_t^\Delta = W_t^{min} + \sigma^2\rho^2(1-c)N, \tag{4.13}$$

$$W_\Delta^* = W_\Delta^{min} + 2\sigma^2\rho^2(1-c)N. \tag{4.14}$$

In this case, as in case A, the loss incurred by using the "wrong" prediction formula is larger the larger is ρ and the smaller is c. Only when c=1, there is no conflict between the optimal choice of predictors for Y and ΔY.

5. ESTIMATION AND PREDICTION
MODEL II: LINEAR REGRESSION

5.1 The aggregate variables

We then consider the case where the systematic part of y_{ht} in (2.3), a_{ht}, is related to an observable variable x_{ht}.[7] The relationship is assumed to be linear, $a_{ht} = \alpha+\beta x_{ht}$, i.e.

$$(5.1)\qquad y_{ht} = \alpha+\beta x_{ht}+\mu_h+\nu_{ht} \qquad (h=1,\dots,H;\ t=1,2),$$

where α and β are unknown constants and x_{ht} is *stochastic* and uncorrelated with the disturbance components μ_h and ν_{ht}.[8] Eqs. (3.6) - (3.8) should then be replaced by

$$(5.2)\qquad \bar{Y}_t = \alpha+\beta\bar{X}_t+\bar{\mu}+\bar{\nu}_t,$$

$$(5.3)\qquad \bar{Y}_t(S_i) = \alpha+\beta\bar{X}_t(S_i)+\bar{\mu}(S_i)+\bar{\nu}_t(S_i),$$

$$(5.4)\qquad Y_t(S^*) = \alpha+\beta\bar{X}_t(S^*)+\bar{\mu}(S^*)+\bar{\nu}_t(S^*) \qquad (i=1,2,12;\ t=1,2),$$

where the $\bar{\mu}$'s and $\bar{\nu}$'s are defined as in (3.9)-(3.14) and

$$(5.5)\qquad \bar{X}_t = \frac{1}{H}\sum_{h=1}^{H} x_{ht} ,$$

$$(5.6)\qquad \bar{X}_t(S_i) = \frac{1}{n_i}\sum_{h\in S_i} x_{ht} ,$$

$$(5.7)\qquad \bar{X}_t(S^*) = \frac{1}{m}\sum_{h\in S^*} x_{ht} .$$

We have joint observations on y_{ht} and x_{ht} from all individuals in the samples.

5.2 Estimation

The parameters $\alpha,\beta,\rho,$ and σ^2 can be estimated by means of the Maximum Likelihood principle in a similar way as the estimation of $a_1, a_2, \rho,$ and σ^2 in model I; see section 3.2.[9] The iterative algorithm consists in switching between GLS estimation of α and β, conditional on ρ and σ^2, and estimation of ρ and σ^2, conditional on α and β. Let the estimates be denoted as $\hat{\alpha},\hat{\beta}$, $\hat{\rho}$, and $\hat{\sigma}^2$.

5.3 Prediction

We now turn to the problem of predicting the population totals $Y_1 = H\bar{Y}_1$ and $Y_2 = H\bar{Y}_2$, and their difference $\Delta Y = Y_2 - Y_1$. The information available for prediction in this case is the values observed of y_{ht} and x_{ht} and the estimates $\hat{\alpha},\hat{\beta}$ and $\hat{\rho}$. We define the following predictors:

$$(5.8)\quad \begin{cases} \tilde{Y}_1 = v_{11}\bar{Y}_1(S_1) + v_{12}\bar{Y}_1(S_{12}) + w_{11}\{\hat{\alpha}+\hat{\beta}\bar{X}_1(S_1)\} \\ \qquad + w_{12}\{\hat{\alpha}+\hat{\beta}\bar{X}_1(S_{12})\}, \\ \tilde{Y}_2 = v_{21}\bar{Y}_2(S_{12}) + v_{22}\bar{Y}_2(S_2) + w_{21}\{\hat{\alpha}+\hat{\beta}\bar{X}_2(S_{12})\} \\ \qquad + w_{22}\{\hat{\alpha}+\hat{\beta}\bar{X}_2(S_2)\}, \end{cases}$$

where the v's and w's are suitable weights. These predictors are linear combinations of the y values observed and estimates of their (unconditional) expectations, with different weights assigned to individuals observed once and twice. When the w's are allowed to be different from zero, $\tilde{Y}_1$ and $\tilde{Y}_2$ define *combined model and design based predictors* since they utilize information on the sampling design along with information on the mechanism which connects the y's and the x's. The corresponding predictor of ΔY is

$$(5.9)\quad \Delta\tilde{Y} = v_{22}\bar{Y}_2(S_2) - v_{11}\bar{Y}_1(S_1) + v_{21}\bar{Y}_2(S_{12}) - v_{12}\bar{Y}_1(S_{12})$$
$$+ (w_{21} + w_{22} - w_{12} - w_{11})\hat{\alpha}$$
$$+ \{w_{21}\bar{X}_2(S_{12}) + w_{22}\bar{X}_2(S_2) - w_{12}\bar{X}_1(S_{12}) - w_{11}\bar{X}_1(S_1)\}\hat{\beta}.$$

Using (5.2) and (5.3), we find that the *prediction errors* of Y_1 and Y_2 can be written as

$$(5.10) \qquad d_1 = \tilde{Y}_1 - Y_1 = (v_{11}+v_{12}+w_{11}+w_{12}-H)\alpha + (Q_1 - H\bar{X}_1)\beta + (w_{11} + w_{12})(\hat{\alpha}-\alpha) + \{w_{11}\bar{X}_1(S_1) + w_{12}\bar{X}_1(S_{12})\}(\hat{\beta}-\beta) + U_1 - H(\bar{\mu}+\bar{\nu}_1),$$

$$(5.11) \qquad d_2 = \tilde{Y}_2 - Y_2 = (v_{21}+v_{22}+w_{21}+w_{22}-H)\alpha + (Q_2 - H\bar{X}_2)\beta + (w_{21} + w_{22})(\hat{\alpha}-\alpha) + \{w_{21}\bar{X}_2(S_{12}) + w_{22}\bar{X}_2(S_2)\}(\hat{\beta}-\beta) + U_2 - H(\bar{\mu}+\bar{\nu}_2),$$

where

$$(5.12) \qquad \begin{cases} Q_1 = (v_{11}+w_{11})\bar{X}_1(S_1) + (v_{12}+w_{12})\bar{X}_1(S_{12}), \\ Q_2 = (v_{21}+w_{21})\bar{X}_2(S_{12}) + (v_{22}+w_{22})\bar{X}_2(S_2), \end{cases}$$

and U_1 and U_2 are defined as in (3.28).

We impose a similar restriction of unbiasedness on the weighting system of these predictors as in model I (cf. 3.29)), namely

$$(5.13) \qquad v_{11}+v_{12}+w_{11}+w_{12} = v_{21}+v_{22}+w_{21}+w_{22} = H,$$

which implies that the first term in (5.10) - (5.11) vanishes. The second term represents the errors in the exogenous variables; $Q_t - H\bar{X}_t$ is the difference between the predicted and actual value of its population total in period t (t=1,2). These errors can be controlled by changing either the sampling design or the weighting system, since Q_1 and Q_2 depend on these parameters. Thirdly, the effect of the errors in the estimates $\hat{\alpha}$ and $\hat{\beta}$, can be controlled by changing the weights w_{ij}. (The estimates, of course, are affected by the sampling design.) Finally, the disturbance components in the regression equation give the same contribution to the prediction error,

$U_t - H(\bar{\mu}+\bar{\nu}_t)$ (t=1,2), as in model I; cf. (3.37)-(3.38). As noted in section 3.3, this error will be affected partly by the sampling design and partly by our choice of weighting system.

The sampling design thus affects the total prediction error through several "channels". For simplicity, we assume in the following that the samples are so large that the errors in the estimated regression coefficients can be neglected; i.e. we let $\hat{\alpha}=\alpha$ and $\hat{\beta}=\beta$. The prediction errors for the level of Y_1 and Y_2 then become

$$(5.14) \qquad d_t = R_t\beta + u_t \qquad (t=1,2),$$

with a corresponding error for the change ΔY equal to

$$(5.15) \qquad d_\Delta = d_2 - d_1 = (R_2 - R_1)\beta + u_2 - u_1,$$

where

$$(5.16) \qquad R_t = Q_t - H\bar{X}_t \qquad (t=1,2)$$

and

$$(5.17) \qquad u_t = U_t - H(\bar{\mu}+\bar{\nu}_t) \qquad (t=1,2).$$

5.4 Distribution of the exogenous variables and the prediction errors

From the assumptions made so far, we can only draw conclusions on the prediction errors d_1, d_2, and d_Δ which are *conditional* on the values of the exogenous variable x_{ht}, i.e. conditional on R_1 and R_2. This discussion would proceed exactly as in case B in section 3.3, and we shall not repeat it here.

In order to focus more specifically on the effect of variations in the exogenous variable, we now make the following assumption about its distribution (or the "super-population" model which generates x_{ht}): All x's in period t have the same expectation, ξ_t, and satisfy the following variance components specification:

$$(5.18)\qquad x_{ht} = \xi_t + \eta_h + \kappa_{ht} \qquad (h=1,\dots,H;\ t=1,2),$$

where η_h and κ_{ht} are uncorrelated with μ_h and ν_{ht}, and

$$(5.19)\qquad E(\eta_h) = E(\kappa_{ht}) = 0,$$

$$(5.20)\qquad \begin{cases} E(\eta_h\eta_{h'}) = \delta_{hh'}\tau_\eta^2, \\ E(\eta_h\kappa_{h't}) = 0, \\ E(\kappa_{ht}\kappa_{h't'}) = \delta_{hh'}\delta_{tt'}\tau_\kappa^2, \end{cases}$$

$\delta_{hh'}$ and $\delta_{tt'}$ denoting, as before, Kronecker deltas.[10] This implies

$$(5.21)\qquad \mathrm{cov}\,(x_{ht}, x_{h't'}) = \begin{cases} \tau^2 & \text{for } h'=h,\ t'=t \\ \rho_x\tau^2 & \text{for } h'=h,\ t'\neq t \\ 0 & \text{otherwise,} \end{cases}$$

where $\tau^2 = \tau_\eta^2 + \tau_\kappa^2$, and $\rho_x = \tau_\eta^2/\tau^2$. The latter ratio obviously has the alternative interpretation as the coefficient of correlation between x_{h1} and x_{h2}. Furthermore, we assume that *the sampling design is independent of the values of the individual components* η_h.

In the following, we shall let "$|S$" symbolize conditioning on the sample $S=S_1\cup S_{12}\cup S_2$. We shall interpret this not as conditioning on the values of x_{ht} from the individuals in this sample, but as *conditioning with respect to the individual components* of x_{ht} and of the regression disturbances of all individuals in S, i.e. "$|S$" is a shorthand notation for "$|\eta_h,\mu_h; h\in S$". What we do is thus to condition on the part of the regressors and disturbances which are particular to the individuals actually observed, and hence can be "controlled" by means of the sampling design.

From (5.5)-(5.7), (5.12), (5.13), (5.16), and (5.18)-(5.20) we then obtain

$$(5.22)\qquad \begin{cases} E(R_1|S) = (v_{11}+w_{11}-n_1)\bar{\eta}(S_1)+(v_{12}+w_{12}-n_{12})\bar{\eta}(S_{12})-n_2\bar{\eta}(S_2) = A_1, \\ E(R_2|S) = (v_{21}+w_{21}-n_{12})\bar{\eta}(S_{12})+(v_{22}+w_{22}-n_2)\bar{\eta}(S_2)-n_1\bar{\eta}(S_1) = A_2, \end{cases}$$

and

$$(5.23)\quad \begin{cases} \text{var}\ (R_1|S) = \tau^2[(1-\rho_x)\{\frac{(v_{11}+w_{11})^2}{n_1} + \frac{(v_{12}+w_{12})^2}{n_{12}} - H\} + \rho_x m] = C_{11}, \\ \text{var}\ (R_2|S) = \tau^2[(1-\rho_x)\{\frac{(v_{21}+w_{21})^2}{n_{12}} + \frac{(v_{22}+w_{22})^2}{n_2} - H\} + \rho_x m] = C_{22}, \\ \text{cov}\ (R_1,R_2|S) = \tau^2\rho_x m = C_{12}, \end{cases}$$

where $\bar{\eta}(S_i) = \frac{1}{n_i}\sum_{h\in S_i}\eta_h$ (i=1,2,12), and A_t and C_{ts} are defined by the last equalities. In a similar way, (2.5), (2.6), (3.9)-(3.14),(3.28),and (5.17) imply

$$(5.24)\quad \begin{cases} E(u_1|S) = (v_{11}-n_1)\bar{\mu}(S_1) + (v_{12}-n_{12})\bar{\mu}(S_{12})-n_2\bar{\mu}(S_2) = B_1, \\ E(u_2|S) = (v_{21}-n_{12})\bar{\mu}(S_{12}) + (v_{22}-n_2)\bar{\mu}(S_2)-n_1\bar{\mu}(S_1) = B_2, \end{cases}$$

and

$$(5.25)\quad \begin{cases} \text{var}\ (u_1|S) = \sigma^2[(1-\rho)\{\frac{v_{11}^2}{n_1} + \frac{v_{12}^2}{n_{12}} + H-2(v_{11}+v_{12})\} + \rho m] = D_{11}, \\ \text{var}\ (u_2|S) = \sigma^2[(1-\rho)\{\frac{v_{21}^2}{n_{12}} + \frac{v_{22}^2}{n_2} + H-2(v_{21}+v_{22})\} + \rho m] = D_{22}, \\ \text{cov}\ (u_1,u_2|S) = \sigma^2\rho m = D_{12}, \end{cases}$$

where B_t and D_{ts} are defined by the last equalities.

We can now write the expectations and variances of the prediction errors, *conditional on the sample*, as follows

$$(5.26)\quad \begin{cases} E(d_1|S) = \beta A_1 + B_1, \\ E(d_2|S) = \beta A_2 + B_2, \\ E(d_\Delta|S) = \beta(A_2-A_1) + B_2 - B_1, \end{cases}$$

and

$$(5.27)\quad \begin{cases} \text{var}\,(d_1|S) = \beta^2 C_{11} + D_{11}, \\ \text{var}\,(d_2|S) = \beta^2 C_{22} + D_{22}, \\ \text{var}\,(d_\Delta|S) = \beta^2(C_{11} + C_{22} - 2C_{12}) + (D_{11} + D_{22} - 2D_{12}). \end{cases}$$

Since A_t and B_t are different from zero, the same will, in general, be the case for the conditional expectations of the prediction errors, (5.26). The values of these expectations reflect the values of η_h and μ_h of the individuals in the sample.

Since, however, $E(A_t) = E(E(R_t|S)) = 0$ and $E(B_t) = E(E(u_t|S)) = 0$ in view of (5.19), (2.5), and our assumptions about the sampling design, we have

$$(5.28)\quad E(d_1) = E(d_2) = E(d_\Delta) = 0,$$

i.e. unconditionally, the predictors $\tilde{Y}_1$, $\tilde{Y}_2$ and $\widetilde{\Delta Y}$ are unbiased. The unconditional variances of the prediction errors are

$$(5.29)\quad \begin{cases} \text{var}\,(d_1) = E[\text{var}\,(d_1|S)] + \text{var}[E(d_1|S)] \\ \qquad = \beta^2\{C_{11} + \text{var}\,(A_1)\} + D_{11} + \text{var}\,(B_1), \\ \text{var}\,(d_2) = E[\text{var}\,(d_2|S)] + \text{var}[E(d_2|S)] \\ \qquad = \beta^2\{C_{22} + \text{var}\,(A_2)\} + D_{22} + \text{var}\,(B_2), \\ \text{var}\,(d_\Delta) = \beta^2\{C_{11} + C_{22} - 2C_{12} + \text{var}\,(A_1) + \text{var}\,(A_2) - 2\,\text{cov}\,(A_1,A_2)\} \\ \qquad + D_{11} + D_{22} - 2D_{12} + \text{var}\,(B_1) + \text{var}\,(B_2) - 2\,\text{cov}\,(B_1,B_2). \end{cases}$$

There is an important difference between conditional and unconditional inference in this case. All the conditional variances (5.27) depend on ρ_x and ρ, since C_{ts} and D_{ts} are functions of these parameters. The same is true for the unconditional variance of d_Δ. The unconditional variances of d_1 and d_2 in (5.29), however, will be independent of ρ and ρ_x, since it is easy to verify that the terms including ρ_x in C_{tt} cancel against the the corresponding terms in var (A_t) and that the terms including ρ in D_{tt} cancel against those in var (B_t) (t=1,2) ;cf. (6.2) below.

6. OPTIMAL CHOICE OF PREDICTORS
MODEL II: LINEAR REGRESSION

The variances of the prediction errors, given in (5.27) and (5.29), represent the joint effect of the random disturbances in the regression equation and the stochastic elements of the exogenous variable x_{ht}. Let us now examine the optimal choice of predictors on the basis of these formulae.

6.1 Conditional prediction

Consider first the problem from the point of view of conditional prediction, in the sense defined in section 5.3. Since D_{tt} in (5.25) is independent of w_{ts} and since $\partial C_{tt}/\partial w_{ts} = \partial C_{tt}/\partial v_{ts}$ (t=1,2; s=1,2), we find, by using simple calculus, that the values of v_{ts} and w_{ts} that minimize var $(d_1|S)$ and var $(d_2|S)$, subject to (5.13), are, respectively

$$(6.1)\quad \begin{cases} v_{11} = n_1, w_{11} = n_1\left[\frac{H}{n_1+n_{12}} - 1\right], v_{12} = n_{12}, w_{12} = n_{12}\left[\frac{H}{n_1+n_{12}} - 1\right], \\ v_{21} = n_{12}, w_{21} = n_{12}\left[\frac{H}{n_2+n_{12}} - 1\right], v_{22} = n_2, w_{22} = n_2\left[\frac{H}{n_2+n_{12}} - 1\right]. \end{cases}$$

Moreover, exactly the same choice of weights will minimize var $(d_\Delta|S)$. This follows from the fact that neither of the covariances C_{12} or D_{12} in (5.27) depends on v_{ts} or w_{ts}, and so they can be disregarded in the process of minimization.

Our conclusion, then, is that although the conditional variances of the prediction errors depend on ρ_x and ρ, the optimal choice of weights for conditional prediction will not be affected by these parameters. The intuitive explanation of this is, of course, that in the conditional distribution, where η_h and μ_h are treated as fixed, all x_{ht} and ε_{ht} will be uncorrelated, and so the composition of the sample between individuals observed once and twice will have no effect on the prediction performance. At the same time, in the conditional distribution, the individual components η_h and μ_h will become part of the intercept term of the regression equation, which explains why the predictors come out as "conditionally biased" in this case, cf. (5.26).

6.2 Unconditional prediction

From (5.22)-(5.25) and (5.29) we find that the unconditional variances of the prediction errors d_1 and d_2 can be written as

$$(6.2)\quad \begin{cases} \text{var}\,(d_1) = \tau^2\beta^2\left[\dfrac{(v_{11}+w_{11})^2}{n_1} + \dfrac{(v_{12}+w_{12})^2}{n_{12}} - H\right] + W_1\,, \\ \text{var}\,(d_2) = \tau^2\beta^2\left[\dfrac{(v_{21}+w_{21})^2}{n_{12}} + \dfrac{(v_{22}+w_{22})^2}{n_2} - H\right] + W_2\,, \end{cases}$$

where W_1 and W_2 are defined as in (3.39) and (3.40). These variances attain their minima, subject to (5.13), for the same choice of weights, (6.1), as in the corresponding problem of conditional prediction. Recalling (5.8), we find that (6.1) implies that the n_t+n_{12} observations on y_{ht} from period t are included with full weight in the predictor for this period, whereas the $H-n_t-n_{12}$ individuals unobserved are represented by the (estimated) value of $E(y_{ht})$ with x_{ht} set equal to its sample average, i.e.

$$(6.3)\qquad \hat{a}_{ht} = \hat{\alpha} + \hat{\beta}\,\frac{n_t\bar{x}_t(S_t)+n_{12}\bar{x}_t(S_{12})}{n_t+n_{12}} \qquad (t=1,2).$$

The optimal procedure for predicting Y_1 and Y_2 in the regression model is thus very similar to the optimal predictor in model I,(4.9).

Furthermore, the unconditional variance of d_Δ is

$$(6.4)\qquad \text{var}\,(d_\Delta) = \tau^2\beta^2\Big[\frac{(v_{11}+w_{11})^2}{n_1} + \frac{(v_{22}+w_{22})^2}{n_2} - 2(1-\rho_x)H$$
$$+ \frac{1}{n_{12}}\{(v_{12}+w_{12})^2 - 2\rho_x(v_{12}+w_{12})(v_{21}+w_{21})$$
$$+ (v_{21}+w_{21})^2\}\Big] + W_\Delta\,,$$

where W_Δ is given by (3.42). Obviously, minimization of this variance with respect to the v's and w's is not equivalent to minimization of W_Δ; i.e. the *distribution of the exogenous variable in the regression equation will affect the optimal choice of predictor of ΔY in this case*. Assume again, for simplicity, that the same number of individuals is observed in both periods, i.e. $n_1=n_2=n$. The values of v_{ts} and w_{ts} that minimize this variance is

$$(6.5)\quad \begin{cases} v_{11}=v_{22}=n(1-\rho), \\ v_{12}=v_{21}=n_{12}, \\ w_{11}=w_{22}=n[\dfrac{H(1-\rho_x)}{n(1-\rho_x)+n_{12}} - (1-\rho)], \\ w_{12}=w_{21}=n_{12}[\dfrac{H}{n(1-\rho_x)+n_{12}} - 1]. \end{cases}$$

Inserting these values in (5.9), we find that the optimal predictor can be written as

$$(6.6)\quad \Delta\tilde{Y} = n_{12}\Delta\bar{Y}(S_{12}) + n(1-\rho)\{\bar{Y}_2(S_2) - \bar{Y}_1(S_1)\} + n\rho\beta\{\bar{X}_2(S_2) - \bar{X}_1(S_1)\} + [H-n-n_{12}]\beta\Delta\tilde{\bar{X}} + n\rho_x\beta\Delta\tilde{\bar{X}},$$

where

$$(6.7)\quad \Delta\tilde{\bar{X}} = \frac{n(1-\rho_x)}{n(1-\rho_x)+n_{12}}[\bar{X}_2(S_2) - \bar{X}_1(S_1)] + \frac{n_{12}}{n(1-\rho_x)+n_{12}}\Delta\bar{X}(S_{12}).$$

This predictor implies that the individuals observed twice are given full weight, as in model I, cf. (4.10) (first term), whereas those observed once are represented by a weighted average of their observed value (second term) and the estimate of their expectation conditional on the values of x_{ht} from these individuals (third term), with weights equal to $1-\rho$ and ρ, respectively. Each individual not observed is represented by the estimate of the expected increase in y, $E(y_{h2}-y_{h1})$, with $x_{h2}-x_{h1}$ set equal to a $\Delta\tilde{\bar{X}}$, which is a weighted average of the predicted increase in x based on observations from all individuals in the sample (fourth term). The relative weights assigned to individuals observed once and twice in this average depend on ρ_x, the individual share in the total variance of x_{ht}, cf. (6.7). Finally, the fifth term in (6.6) "corrects" for using an inoptimal predictor of the increase in x in the third term of the prediction formula.

We see that *observations on x_{ht} and y_{ht} from all individuals - whether observed once or twice - are elements in the optimal predictor of ΔY in the general case where $0 \leq \rho < 1$ and $0 \leq \rho_x < 1$.* In certain particular cases, however, we will only make use of information on *either* the y's *or* the x's from the individuals observed once, but we will always need all information from those observed twice. The following examples illustrate this point:

$$\underline{\rho=\rho_x=1}: \Delta\tilde{Y} = n_{12}\Delta\bar{Y}(S_{12}) + n\beta\{\bar{X}_2(S_2)-\bar{X}_1(S_1)\} + [H-n_{12}]\beta\Delta\bar{X}(S_{12}),$$

$$\underline{\rho=1, \rho_x=0}: \Delta\tilde{Y} = n_{12}\Delta\bar{Y}(S_{12}) + n\beta\{\bar{X}_2(S_2)-\bar{X}_1(S_1)\}$$

$$+ [\frac{H}{n+n_{12}} - 1]\beta n[\bar{X}_2(S_2)-\bar{X}_1(S_1)] + n_{12}\Delta\bar{X}(S_{12})\} ,$$

$$\underline{\rho=0, \rho_x=1}: \Delta\tilde{Y} = n_{12}\Delta\bar{Y}(S_{12}) + n\{\bar{Y}_2(S_2)-\bar{Y}_1(S_1)\} + [H-n_{12}]\beta\Delta\bar{X}(S_{12}).$$

The larger is ρ, the less useful will be the observations on y_{ht} from the individuals observed only once, the larger is ρ_x, the less useful will be the observations on x_{ht} from the same individuals.

The crucial role played by ρ and ρ_x in the optimal predictor of ΔY can be explained in a slightly different way. From (6.5) it follows that

$$\frac{v_{11}}{v_{12}} = \frac{v_{22}}{v_{21}} = \frac{n}{n_{12}} (1-\rho),$$

$$\frac{v_{11}+w_{11}}{v_{12}+w_{12}} = \frac{v_{22}+w_{22}}{v_{21}+w_{21}} = \frac{n}{n_{12}} (1-\rho_x),$$

i.e. the relative weight given to *observations on y_{ht}* from individuals observed once and twice depends on ρ only, whereas the relative weight assigned jointly to *observations on y_{ht} and estimates of $E(y_{ht})$ based on the x_{ht} observed* for the same individuals depends on ρ_x only.

Let, as before, $N = n+n_{12}$ and $c = n_{12}/N$. The minimum value of var (d_Δ) can then be written as

$$(6.8) \qquad \text{var}\ (d_\Delta)^{\min} = 2\tau^2\beta^2(1-\rho_x)H[\frac{H}{N}\cdot\frac{1}{1-\rho_x+\rho_x c} - 1] + 2\sigma^2(1-\rho)[H-N(1-\rho+\rho c)].$$

Since both terms in this expression are decreasing functions of c if either ρ_x or ρ is positive, we can always obtain a better prediction performance by increasing the share of the sample which is observed twice. The minimum value, for c=1, is $2\tau^2\beta^2(1-\rho_x)H(H/N-1) + 2\sigma^2(1-\rho)(H-N)$.

Let var $(d_\Delta)^*$ denote the value of var (d_Δ) when all individuals are given the same weight in the prediction formula, i.e. when using (6.1). We find

$$(6.9) \qquad \text{var}\ (d_\Delta)^* = \text{var}\ (d_\Delta)^{\min} + 2\tau^2\beta^2H\frac{\rho_x c(1-c)}{1-\rho_x+\rho_x c} + 2\sigma^2\rho^2(1-c)N.$$

The loss of efficiency is larger the larger is ρ_x and ρ and the smaller is c.

7. CONCLUDING REMARKS

In this paper, we have been particularly concerned with the interplay between the sampling design and the covariance structure of the data vector when predicting an aggregate variable y from sampling survey data. One conclusion is that the optimal choice of predictor, i.e. the one that minimizes the variance of the prediction error, will not, in general, be the same when predicting the aggregate level of y and when the purpose is to predict its aggregate change. In the latter case, in contrast to the first, information on the relative share of the individuals which are observed twice as well as on the share of the variance of y which is due to individual differences, play a crucial role in the optimal prediction formula. Hence, these parameters become key parameters when assessing the potential gain which could be obtained by changing the sampling design. This is by no means a point of academic interest only. An empirical study of consumer demand in Norwegian households based on rotating panel data from the years 1975-1977, gave estimates of the individual share of the total disturbance variances which extended from zero to about 0.7. For 22 of 28 commodity groups - accounting for about 85 per cent of the budget of the average consumer - the estimates were significantly different from zero. (Biørn and Jansen (1982, section 7.5).)

Furthermore, we have shown how observations on a variable x which is related to y through a linear regression equation may be used to improve the predictor of the latter variable. In this case, ρ_x, the individual share of the variance of x, turns out to be a crucial parameter in determining the optimal predictor for the change in y.

Another conclusion is that when individual specific components are present, we can always improve our predictor of the change in y by increasing the share of the individuals which are observed twice, given the total sample size. The variance of the prediction error will then take its lowest value when all individuals are observed twice, and in that case - and only then - will there be no conflict between the optimal choice of predictor for the level of y and for its change. It should be recalled, however, that this conclusion rests on our simplifying assumption that errors in the estimated structural coefficients (i.e. $\hat{a}_1$ and $\hat{a}_2$ in model I, $\hat{\alpha}$ and $\hat{\beta}$ in model II) can be neglected. It may well be modified in small sample situations when such errors are taken into account. If, for instance, we can increase the spread in the data by increasing the share of the individuals which are observed once, we may obtain better estimates of the structural coefficients, which in turn may lead to the conclusion that a design with some degree of rotation may be the best compromise design for prediction purposes.

This problem deserves further research. However, as the algebra seems to become rather messy, Monte Carlo experiments may be the only feasible approach. The models we have considered here are the simplest possible, and more general situations may be well worth investigation. An obvious extension would be, with basis in the general framework outlined in section 2, to consider a situation with more than two periods involved and in which some individuals are observed more than twice. Another interesting generalisation might be a situation in which there exists summary information on the regressor variable x for (some of) the individuals outside the sample, in addition to the joint observations on y and x from those included in the sample.

---oOo---

NOTES

1) Confer Biørn (1981, p. 17).

2) See Biørn (1981, pp. 26-27).

3) No such correlation would exist, however, if the estimates of a_1 and a_2 were based on data from independently drawn samples which were non-overlapping with S_1, S_2, or S_{12}.

4) Not surprisingly, we find that W_1, W_2, and W_Δ coincide with V_1, V_2, and V_Δ when $v_{11}=k_1H$, $v_{12}=(1-k_1)H$, $v_{22}=k_2H$, $v_{21}=(1-k_2)H$.

5) Since λ is rather insensitive with respect to the value of H/N, provided it is not too small (less than 50 say), the figures in table 2 are valid approximations to the exact λ over most of the relevant range of H/N.

6) These ratios overstate the gain which can be obtained in practical situations, since a_1 and a_2 will have to be estimated from the data.

7) For simplicity, we confine attention to one regression variable only. The generalization to multiple regression models is straightforward.

8) Assumptions (2.5) and (2.6) then hold conditionally on the x's, which, of course, also implies that they hold marginally.

9) We implicitly assume that α, β, ρ, and σ^2 are not parameters in the distribution of the x's, so that the ML estimates can be obtained by maximizing the conditional density.

10) Note that x_{ht} in this model is generated by the same kind of mechanism as y_{ht} in model I, cf. (2.5)-(2.6).

REFERENCES

Biørn, E. (1981):

Estimating Seemingly Unrelated Regression Models from Incomplete Cross-Section/Time-Series Data. Reports from the Central Bureau of Statistics of Norway, 81/33. (Oslo: Central Bureau of Statistics, 1981.)

Biørn, E. and Jansen, E. S. (1982):

Econometrics of Incomplete Cross-Section/Time-Series Data: Consumer Demand in Norwegian Households 1975-1977. Social Economic Studies 52. (Oslo: Central Bureau of Statistics, 1982.)

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1979):

Prediction Theory for Finite Populations when Model-Based and Design-Based Principles are Combined. Scandinavian Journal of Statistics, 6(1979), 97-106.

Royall, R. M. (1970):

On Finite Population Sampling Theory Under Certain Regression Models. Biometrika, 57(1970), 377-387.
