TIME SERIES GENERALIZATIONS OF FAY-HERRIOT ESTIMATOR FOR SMALL AREAS

A.C. Singh, H.J. Mantel, and B.W. Thomas, Statistics Canada
A.C. Singh, 16-A, R.H. Coats Building, Statistics Canada, Ottawa K1A 0T6

KEY WORDS: Composite estimation; State space models; Kalman Filter; Best linear unbiased prediction.

1. INTRODUCTION

There exists a considerable body of research on small area estimation using cross-sectional survey data in conjunction with supplementary data obtained from census and administrative sources. A good collection of papers on this topic can be found in Platek, Rao, Särndal and Singh (1987). The basic idea underlying all small area methods is to borrow strength from other areas via a model containing auxiliary variables from the supplementary data. Recently time series methods are being employed for repeated surveys to develop improved estimators for small areas; see Choudhry and Rao (1989) and Pfeffermann and Burck (1990). It is interesting to note that after the initiative of Scott and Smith (1974) on the application of time series methods to survey data, there has been only lately a resurgence of interest in developing suitable estimates of aggregates from complex surveys repeated at regular time intervals; see e.g. Bell and Hilmer (1987), Binder and Dick (1989), Tiller (1989), and Pfeffermann (1991).

In this paper we consider some natural generalizations of the Fay Herriot (FH) estimator for small areas when a time series of direct small area estimates is available. The important work of Fay and Herriot (1979) shows how direct estimators can be smoothed by cross-sectional modelling of small area totals. The resulting estimators are composite estimators (i.e. convex combinations of direct and model-based synthetic estimators) and are also empirical best linear unbiased predictors (EBLUPs). With the use of structural models, we derive time series EBLUPs which combine both cross-sectional and time series data. The main purpose of this paper is to compare time series EBLUPs with cross-sectional estimators such as post-stratified domain, synthetic, FH and sample size dependent estimators.

An empirical study based on Monte Carlo simulations from real time series data obtained from Statistics Canada's biannual farm surveys was conducted to investigate potential gains in efficiency with time series EBLUPs. The main findings of the study are

(i) There can be substantial gains in efficiency with time series EBLUPs over cross-sectional estimators.

(ii) Within the class of time series methods considered in this paper, introduction of serial dependence in the random small area effects is found to be considerably more beneficial than dependence of the parameters of the synthetic component of the cross-sectional EBLUP (i.e. FH estimator).

(iii) Within the class of cross-sectional methods, the performance of FH estimator is best overall followed by that of sample size dependent estimator.

(iv) Although any smoothed version of the direct small area estimator is expected to be biased, the time series EBLUPs exhibit less bias in magnitude than other methods including FH estimator.

Section 2 contains a version of various cross-sectional methods for small area estimation. Time series EBLUPs are described in Section 3 and the details and results of the Monte Carlo comparative study are given in Section 4. Finally, some directions for future work are mentioned in the Section 5.

2. METHODS BASED ON CROSS-SECTIONAL DATA

In this section, we assume that information is available only for a particular point in time t, $t=1,\dots,T$. Let $\underset{\sim}{\theta}_t$ denote the vector of small area population totals $\theta_{kt}, k=1, \dots K$, at time t. Here we define briefly some well known small area estimators under the assumption that the underlying sampling design is stratified simple random; for more details, see Rao (1986). Särndal and Hidiroglou (1987) and Pfeffermann and Burck (1991) also contain a good survey of various small area estimators.

2.1 Method 1 (Expansion estimator)

This method of estimation is defined by

$$g_{1kt} = \sum_h (N_{ht}/n_{ht}) \sum_{j\in s_{hkt}} y_{hjt}, \tag{2.1}$$

where at time t, y_{hjt} is the jth observation in the h stratum, s_{hkt} denotes the set of n_{hkt} sample units falling in the kth small area in the hth stratum and n_{ht}, N_{ht} denote respectively the sample and population sizes for the hth stratum. The above estimator is generally unreliable because the random sample size n_{hkt} is likely to be small in expectation and could have high variability. Conditional on the realized sample size n_{hkt}, g_{1kt} is biased. However, unconditionally, it is unbiased for θ_{kt}.

2.2 Method 2 (Post-stratified domain estimator)

We will refer to this estimator also as the direct small area estimator. Suppose the population size N_{hkt} is known for each (h,k,t). The efficiency of estimator g_{1kt} could be improved by post-stratification. Suppose small areas themselves constitute post-strata within stratum h. We have

$$g_{2kt} = \sum_h (N_{hkt}/n_{hkt}) \sum_{j\in s_{hkt}} y_{hjt} = \sum_h N_{hkt}\bar{y}_{hkt}. \tag{2.2}$$

However, this estimator also may not be sufficiently reliable because of the possibility of n_{hkt}'s being small in expectation. If $n_{hkt}=0$, the above estimator is not defined. In practice, some ad hoc value such as 0 is often chosen for $\bar{y}_{hkt}$ when $n_{hkt}=0$. In the empirical study presented in this paper, we have set $\bar{y}_{hkt}$ as $(\bar{X}_{hkt}/\bar{X}_{ht})\,\bar{y}_{ht}$ whenever $n_{hkt}=0$, where X is a suitable covariable.

The estimator g_{2ht} is both conditionally and unconditionally unbiased.

2.3 Method 3 (Synthetic estimator)

It is possible to define a more efficient estimator by assuming a model which allows for "borrowing strength" from other small areas. This gives rise to synthetic estimators. For instance, suppose different small area totals are connected via the auxiliary variable X_{kt} by a linear model as

$$\theta_{kt} = \beta_{1t} + \beta_{2t} X_{kt}, k=1,\dots K, \tag{2.3a}$$

or in matrix notation

$$\underset{\sim}{\theta}_t = F_t \underset{\sim}{\beta}_t, \tag{2.3b}$$

where $F_t = (F_{1t}, F_{2t}, \ldots, F_{Kt})'$, $F_t = (1, X_{Kt})'$. Now consider a model for the direct small area estimators g_{2kt}'s as

$$\underset{\sim}{g}_{2t} = F_t \underset{\sim}{\beta}_t + \underset{\sim}{\epsilon}_t \tag{2.4}$$

where $\underset{\sim}{g}_{2t} = (g_{21t}, \ldots, g_{2Kt})'$, $\underset{\sim}{\epsilon}_t = (\epsilon_{1t}, \ldots \epsilon_{Kt})'$, ϵ_{kt}'s are uncorrelated as k varies with mean 0 and variance v_{kt}.

Denoting by $\hat{\underset{\sim}{\beta}}_t$ the weighted least squares (WLS) estimate of $\underset{\sim}{\beta}_t$, we obtain the regression-synthetic estimator of θ_{kt} under the assumed model as

$$\underset{\sim}{g}_{3t} = F_t \hat{\underset{\sim}{\beta}}_t. \tag{2.5}$$

The above estimator could be heavily biased unless the model (2.3) is satisfied reasonably well.

2.4 Method 4 (EBLUP - empirical best linear unbiased predictor)

Using the empirical Bayes approach of Fay and Herriot (1979) or the more general best linear unbiased predictor (BLUP) approach; see e.g. Battese, Harter, and Fuller (1988), and Pfeffermann and Barnard (1991), the bias of the synthetic estimator can be reduced considerably by using a composite estimator. This is obtained as a convex combination of $\underset{\sim}{g}_{2t}$ and a somewhat modified $\underset{\sim}{g}_{3t}$. For this purpose, it is assumed that

$$\underset{\sim}{\theta}_t = F_t \underset{\sim}{\beta}_t + \underset{\sim}{a}_t, \tag{2.6}$$

where a_{kt}'s are uncorrelated random small area effects with mean 0 and variance w_{kt}. Thus we have a somewhat modified model for $\underset{\sim}{g}_{2t}$ as

$$\underset{\sim}{g}_{2t} = F_t \underset{\sim}{\beta}_t + \underset{\sim}{a}_t + \underset{\sim}{\epsilon}_t. \tag{2.7}$$

Here $\underset{\sim}{a}_t$ is also assumed to be uncorrelated with $\underset{\sim}{\epsilon}_t$. Let g^*_{3t} denote the modified synthetic estimator of $\underset{\sim}{\theta}_t$ under (2.7). The BLUP of $\underset{\sim}{\theta}_t$ under the model defined by (2.6) and (2.7) is

$$\begin{aligned} \underset{\sim}{g}_{4t} &= \underset{\sim}{g}^*_{3t} + \Lambda_t(\underset{\sim}{g}_{2t} - \underset{\sim}{g}^*_{3t}) \\ &= \Lambda_t \underset{\sim}{g}_{2t} + (I-\Lambda_t)\, \underset{\sim}{g}^*_{3t} \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} \Lambda_t &= (V_t^{-1} + W_t^{-1})^{-1} V_t^{-1} = W_t U_t^{-1}, \\ I-\Lambda_t &= (V_t^{-1} + W_t^{-1})^{-1} W_t^{-1} = V_t U_t^{-1}, \\ U_t &= V_t + W_t,\ V_t = \mathrm{diag}(v_{1t}, \ldots v_{Kt}), \\ W_t &= \mathrm{diag}(w_{1t}, \ldots, w_{Kt}). \end{aligned} \tag{2.9}$$

The expression (2.8) follows from the general results on linear models with random effects, see e.g., Rao (1973, p. 267 and Harville (1976). The BLUP or BLUE of $F_t \underset{\sim}{\beta}_t$ is g^*_{3t} and BLUP of $\underset{\sim}{a}_t$ is $\Lambda_t(\underset{\sim}{g}_{2t} - \underset{\sim}{g}^*_{3t})$. It may be of interest to note that the formula for BLUP does not change regardless of known or unknown $\underset{\sim}{\beta}_t$. However, its MSE does change as expected due to estimation of $\underset{\sim}{\beta}_t$. It can be shown that,

$$\begin{aligned} \mathrm{MSE}(\underset{\sim}{g}_{4t} - \underset{\sim}{\theta}_t \mid \underset{\sim}{\beta}_t \text{ known}) &= W_t U_t^{-1} V_t \\ &= V_t - V_t U_t^{-1} V_t, \end{aligned} \tag{2.10}$$

and

$$\begin{aligned} &\mathrm{MSE}(\underset{\sim}{g}_{4t} - \underset{\sim}{\theta}_t \mid \underset{\sim}{\beta}_t \text{ unknown}) \\ &= \mathrm{MSE}\{(A_t + W_t U_t^{-1}(I-A_t))\, \underset{\sim}{\epsilon}_t \\ &\quad - V_t U_t^{-1}(I-A_t)\, \underset{\sim}{a}_t\} \end{aligned} \tag{2.11}$$

where $A_t = F_t(F_t' U_t^{-1} F_t)^{-1} F_t' U_t^{-1}$. The MSE matrix of (2.11) can be easily obtained from MSEs OF $\underset{\sim}{\epsilon}_t$ and $\underset{\sim}{a}_t$.

When V_t and W_t are replaced by their estimates, the estimator $\underset{\sim}{g}_{4t}$ is termed EBLUP. Note that the model (2.6) is more realistic than (2.3), and therefore, the performance of $\underset{\sim}{g}_{4t}$ is expected to be quite favourable. The estimator $\underset{\sim}{g}_{4t}$ approaches $\underset{\sim}{g}_{2t}$ when v_{kt}'s get small, i.e. when n_{hkt}'s become large. However, it remains biased in general, conditional on $\underset{\sim}{\theta}_t$.

2.5 Method 5 (Sample Size dependent estimator)

An alternative composite estimator which can considerably attenuate bias of the synthetic estimator $\underset{\sim}{g}_{3t}$ as compared to the EBLUP $\underset{\sim}{g}_{4t}$ is given by the sample size dependent estimator of Drew, Singh, and Choudhry (1982). It is defined as

$$\underset{\sim}{g}_{5t} = \Delta_t \underset{\sim}{g}_{2t} + (I-\Delta_t)\, \underset{\sim}{g}_{3t}, \tag{2.12}$$

where $\Delta_t = \mathrm{diag}(\delta_{1t}, \ldots, \delta_{Kt})$,

$$\delta_{kt} = \begin{cases} 1 & \text{if } \sum_h \hat{N}_{hkt} \geq \lambda \sum_h N_{hkt}, \\ \sum_h \hat{N}_{hkt}/\lambda \sum_h N_{hkt} & \text{otherwise} \end{cases} \tag{2.13}$$

$\hat{N}_{hkt}$ being $n_{hkt}(N_{ht}/n_{ht})$, and the parameter λ is chosen in an ad hoc manner as a way of controlling the contribution of the synthetic component. In practice, λ is generally chosen as 1, 1.5 or 2. The above estimator takes account of the realized sample size n_{hkt}'s and if these are deemed to be sufficiently large according to the condition in (2.13), then it does not rely on the synthetic estimator. This property is somewhat similar to that of $\underset{\sim}{g}_{4t}$. However, the condition in (2.13) could be satisfied even if some or all n_{hkt}'s are small, and then unlike $\underset{\sim}{g}_{4t}$, the above estimator fails to borrow strength from other small areas even though $\underset{\sim}{g}_{2t}$ is unreliable.

3. METHODS BASED ON POOLED CROSS-SECTIONAL AND TIME SERIES DATA

Suppose information is available for several time points, $t=1\ldots T$, in the form of direct small area estimators $\underset{\sim}{g}_{2t}$ and also the small area population totals for the auxiliary variable. We will now introduce some estimators which generalize the Fay-Herriot estimator $\underset{\sim}{g}_{4t}$ in different ways by taking account of the serial dependence of the direct estimates $\{\underset{\sim}{g}_{2t} : t=1\ldots T\}$. Recall that for the Fay-Herriot estimator, the model for $\underset{\sim}{\theta}_t$ has two components, namely, the trend component $F_t \underset{\sim}{\beta}_t$ and the area component $\underset{\sim}{a}_t$. The estimator $\underset{\sim}{g}_{4t}$ borrows strengh over areas for each t and is given by the sum of two components, each being BLUP (BLUE) for the corresponding random (fixed) effect, i.e.,

$$\underset{\sim}{g}_{4t} = F_t \hat{\underset{\sim}{\beta}}_t + \hat{\underset{\sim}{a}}_t. \tag{3.1}$$

Methods based on time series data could, however, borrow strength over time as well. There are several ways one could build serial dependence in the series $\{\underset{\sim}{g}_{2t}\}$. We introduce three estimators $\underset{\sim}{g}_{6t}$, $\underset{\sim}{g}_{7t}$ and $\underset{\sim}{g}_{8t}$ corresponding to three interesting scenarios which are motivated from specific structural models for serial dependence.

3.1 Method 6 (Time Series EBLUP-I)

In this case, the structural time series model for the direct small area estimates $\{\underset{\sim}{g}_{2t} : t=1,..T\}$ is specified by the following state space model.

Let α_t denote $(\beta_t', a_t')'$ and H_t denote (F_t, I).

Observation Equation

$$\begin{aligned} g_{2t} &= \theta_t + \epsilon_t \\ \theta_t &= F_t \beta_t + a_t \equiv H_t \alpha_t \end{aligned} \tag{3.2a}$$

Transition Equation

$$\alpha_t = G_t \alpha_{t-1} + \zeta_t \tag{3.2b}$$

where

$$G_t = \begin{pmatrix} G_t^{(1)} & 0 \\ 0 & 0 \end{pmatrix}, \; \zeta_t = \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix}, \tag{3.2c}$$

along with the usual assumptions about random errors, i.e., ϵ_t ζ_t are all mutually uncorrelated, ζ_t is uncorrelated with α_s. for $s<t$, and that $\epsilon_t \sim (0, V_t)$, $\zeta_t \sim (0, \Gamma_T)$ where Γ_t = block diag (B_t, W_t). The covariance matrices V_t, B_t, and W_t are generally diagonal. If β_t evolves according to a random walk, then $G_t^{(1)} = I$. The second diagonal submatrix of G_t is zero because a_t's are assumed to be serially independent.

The estimator g_{6T} is BLUP of θ_t given all the direct estimates up to time T. To find g_{6T}, first we will find BLUP $\tilde{\alpha}_t$ of α_T, from which BLUP of θ_T can be simply obtained as $H_t \tilde{\alpha}_T$. Since α_t's are connected over time according to the transition equation, it is possible, albeit cumbersome, to get $\tilde{\alpha}_T$ directly from the theory of linear models with random effects for the complete data . However, it could be convenient to compute it recursively using Kalmar Filter (KF). The recursion algorithim for obtaining $\tilde{\alpha}_{t|s}$ is given as follows.

A time $t-1$, let $\tilde{\alpha}_{t-1}$ and P_{t-1} denote respectively the BLUP of α_t and it MSE i.e.,

$$\tilde{\alpha}_{t-1} - \alpha_{t-1} \sim (0, P_{t-1}) \tag{3.3}$$

Therefore, the BLUP $\tilde{\alpha}_{t|t-1}$ of α_t and its MSE $P_{t|t-1}$ based on $t-1$ observations is given by (in view of the relation 3.2b),

$$\tilde{\alpha}_{t|t-1} = G_t \tilde{\alpha}_{t-1}, \; P_{t|t-1} = G_t P_{t-1} G_t' + Q_t \tag{3.4}$$

Now, combining data at t, i.e., g_{2t} with $\tilde{\alpha}_{t|t-1}$, one can get BLUP $\tilde{\alpha}_t$ and its MSE as

$$\begin{aligned} \tilde{\alpha}_t &= \tilde{\alpha}_{t|t-1} + \\ & P_{t|t-1} H_t' (H_t P_{t|t-1} H_t' + V_t)^{-1} (g_{2t} - H_t \tilde{\alpha}_{t|t-1}) \end{aligned} \tag{3.5a}$$

and

$$\begin{aligned} \mathrm{MSE}(\tilde{\alpha}_t - \alpha_t) &\equiv P_t = P_{t|t-1} - \\ & P_{t|t-1} H_t' (H_t P_{t|t-1} H_t' + V_t)^{-1} H_t P_{t|t-1} \end{aligned} \tag{3.5b}$$

In the usual *KF* terminology, (3.4) and (3.5) specify respectively the prior and posterior distributions of α_t once the data at time t becomes available. Note that here distributions are specified only up to first two moments which is of course sufficient for linear Bays estimation. The results (3.5a) and (3.5b) respectively give the posterior mean and variance of α_t given data up to time t.

The above recursive algorithm or *KF* can be started at the initial time $t=1$ by noting that $\tilde{\alpha}_1$ and P_1 are given by (2.8) and (2.11) respectively, i.e. the corresponding expressions for *FH* estimator at time $t=1$. The recursion is continued until $t=T$ to obtain $\tilde{\alpha}_T$ and the MSE matrix P_T. This, in turn, yields BLUP of θ_T as $t=1$ and its MSE as $H_T P_T H_T'$.

We will now illustrate method of moments for estimating model parameters in the special case when there is only one auxiliary variable X_{ht}, i.e. $F_t = (F_{1t}, \ldots F_{kt})'$, $F_{kt} = (1, X_{kt})'$, $\beta_t = (\beta_{1t}, \beta_{2t})'$, $w_{kt} = \tau^2$ and when β_t follows a random walk i.e. $G_t^{(1)} = I$. Let $B_t = \mathrm{diag}(\gamma_1^2, \gamma_2^2)$. Now, unknown parameters $\tau^2_1 \gamma_1^2, \gamma_2^2$ can be estimated by the method of moments as follows. The parameter τ^2 is obtained as the solution of

$$\textstyle\sum_{t=1}^{T} \sum_{k=1}^{k} (g_{2kt} - F_{kt}' \hat{\beta}_t)^2 / (v_{kt} + w_{kt}) = T(k-2) \tag{3.6}$$

If there is no positive solution, we set τ^2 as zero. Here $\hat{\beta}_t$ denotes the WLS estimate of β_t based on only the cross-sectional data at t. This is analogous to the method used in Fay and Herriot (1979) for cross-sectional data. An estimate of γ_i^2 can be obtained by solving (for $i=1,2$)

$$\textstyle\sum_{t=2}^{T} (\hat{\beta}_{it} - \hat{\beta}_{i,t-1})^2 / (\gamma_i^2 + d_{ii}^{(t)}) = T-1 \tag{3.7}$$

where $d_{ii}^{(t)}$ is the (i,i)th element of $(F_{t-1}' U_{t-1}^{-1} F_{t-1})^{-1} + (F_t' U_t^{-1} F_t)^{-1}$.

When the above estimators of model parameters are substituted in the expression for $H_T \tilde{\alpha}_T$ we get the time series EBLUP-I estimator g_{6T} at time T.

3.2 Method 7 (Time Series EBLUP-II)

The equations for state space model for this case are similar to 3.2(a) and (b) except that the transition matrix G_t and the covariance matrix Γ_t are different. We have two cases.

3.2.1 Case 1 First suppose β_t's fixed and time-invariant but a_t's are serially dependent. Then the matrices G_t and Γ_t are given by

$$G_t = \begin{pmatrix} I & 0 \\ 0 & G_t^{(2)} \end{pmatrix}, \; \Gamma_t = \text{block diag } (0, Q_t) \tag{3.8}$$

For a given choice of Q_t, the *KF* can be run as in method 6 with the initial values $\tilde{\alpha}_1$ and P_1 at $t=1$ obtained from the *FH* estimator at $t=1$. If a_t is assumed to evolve according to a random walk, then $G_t^{(2)} = I$. Moreover, if Q_t is taken as $\nu^2 I$, then the only unknown parameter ν^2 can be estimated from an equation similar to (3.6). We will denote by g_{7t} the EBLUP obtained in this case when the parameter estimate is substituted. Also we will denote by g_{7t}^{a} the estimator in the special case when the common value of β_t is assumed known.

3.2.2 Case 2 Here we assume that β_t's are fixed but different for different time points. The area effects α_t evolve over time as before. The matrices G_t *and* Γ_t are

$$G_t = \begin{pmatrix} I & 0 \\ 0 & G_t^{(2)} \end{pmatrix}, \; \Gamma_t = \text{block diag}\{mI, Q_t\} \tag{3.9}$$

where m is a large integer. The expression of $\tilde{\alpha}_T$ *and* P_T obtained from the *KF* in this case approximately give the correct formulas as $m \to \infty$. The time series EBLUP in this case will be denoted by g_{7t}^{b}.

3.3 Method 8 (Time Series EBLUP-II)

As was the case with method 7, the equations for the state space model are similar to 3.2(a) and (b) except that the two matrices G_t and Γ_t are different. We have

$$G_t = \begin{pmatrix} G_t^{(1)} & 0 \\ 0 & G_t^{(2)} \end{pmatrix}, \; \Gamma_t = \text{block diag } \{B_t, Q_t\}, \tag{3.10}$$

If β_t and $\underset{\sim}{\alpha}_t$'s follow the random walk-process, then both $G_t^{(1)}$ and $G_t^{(2)}$ are identity matrices. Moreover, as before, if $B_t = \mathrm{diag}\{\gamma_1^2, \gamma_2^2\}$ and $Q_t = \nu^2 I$, then the model parameters ν^2, γ_1^2, γ_2^2 can be estimated in an analogous manner by the method of moments. The resulting EBLUP of $\underset{\sim}{\theta}_T$ will be denoted by $\underset{\sim}{g}_{8t}$.

4. MONTE CARLO STUDY

The cross-sectional and time series methods were compared empirically by means of a Monte Carlo simulation from a real time series obtained from Statistics Canada's biannual farm surveys, namely, the National Farm Survey (in June) and the January Farm Survey. Due to the redesign after the census of Agriculture in 1986, the survey data for the six time points starting with the summer of 1988 were employed to create a population for simulation purposes. To this, data from the census year 1986 was also added. Thus information at one more time point was available although this resulted in a 3-point gap in the series. The parameter of interest was taken as the total number of cattle and calves for each crop district. For simplicity, independent stratified random samples were drawn for each occasion from the pseudo-population although the farm surveys use rotating panels over time. The auxiliary variable used in the model was the ratio-adjusted census '86 value of the total cattle and calves for each small area.

4.1 Design of the simulation experiment

First we need to construct a pseudo-population from the survey data over six time points (June'88, Jan'89, ..., Jan'91). It was decided to avoid variability due to changes in the underlying population over time by retaining only those farms which responded to all the six occasions.

The total count of farm units was found to be 1160 which represented a total of over 40,000 farms as a result of appropriate sampling design weighting. For the pseudo-population, we replicated the 1160 farm units proportional to their sampling weight so that the total size N of the population was brought down to a manageable number of 10362 for micro-computer simulation.

The pseudo-population of 10362 units was stratified into four take-some and one take-all strata using Census'86 count data on cattle and calves as the stratification variable. The total sample size was 1036 (about 10% sampling rate) and the size of the take all stratum was 11. A total of 5000 simulations were performed. For each simulation, samples were drawn independently for each time point using a stratified simple random sampling without replacement. The 5000 simulations were conducted in 2500 sets of 2 simulations where each set corresponds to a different vector of realized sample sizes in the twelve small areas within each stratum. This was required to compute certain conditional evaluation measures as described in the next subsection, see also Särndal and Hidiroglou (1989).

4.2 Evaluation Measures

Suppose m simulations are performed in which m_1 sets of different vectors of realized sample sizes in domains (h,k) are replicated m_2 times. The following measures can be used for comparing performance of different estimators at time T. Let i vary from 1 to m_1 and j from 1 to m_2.

(i) Absolute Relative Bias

$$\mathrm{ARB}_k = \left|\left(m^{-1}\sum_i\sum_j (\mathrm{est})_{ijk} - (\mathrm{true})_k\right) / (\mathrm{true})_k\right| \tag{4.1}$$

The average of ARB_k over areas k will be denoted by $\overline{\mathrm{ARB}}$.

(ii) Root Mean Square Conditional Relative Bias

$$\mathrm{RMSCRB}_k = \left\{m_1^{-1}\sum_i\left(m_2^{-1}\sum_j (\mathrm{est})_{ijk} - (\mathrm{true}_k)\right)^2 / (\mathrm{true}_k)^2 - B\right\}^{1/2} \tag{4.2a}$$

$$B = m^{-1}\sum_i\left[\sum_j (\mathrm{est})^2_{ijk}/m_2 - \left(\sum_j (\mathrm{est})_{ijk}/m_2\right)^2\right] / (\mathrm{true}_k)^2 \tag{4.2b}$$

The correction term B adjusts for the bias in the first term. The average of RMSCRB_k over areas k will be denoted by $\overline{\mathrm{RMSCRB}}$..

(iii) Mean Absolute Relative Error

$$\mathrm{MARE}_k = m^{-1}\sum_i\sum_j \left|(\mathrm{est})_{ijk} - (\mathrm{true})_k\right| / (\mathrm{true})_k. \tag{4.3}$$

and $\overline{\mathrm{MARE}}$ denotes the average of MARE_k over areas.

(iv) Root Mean Squre Error

$$\mathrm{RMSE}_k = \left\{m^{-1}\sum_i\sum_j \left((\mathrm{est})_{ijk} - (\mathrm{true})_k\right)^2\right\}^{1/2} \tag{4.4}$$

and $\overline{\mathrm{RMSE}}$ as before denotes the average over areas.

(v) Relative Root Mean Square Error

$$\mathrm{RRMSE}_k = \mathrm{RMSE}_k \,/\, (\mathrm{true})_k. \tag{4.5}$$

Again, we can define $\overline{\mathrm{RRMSE}}$ as before.

4.3 Estimators used in the Comparative Study

There were thirteen estimators included in the study, namely, *M1-M8* corresponding to $\underset{\sim}{g}_{1T}$ to $\underset{\sim}{g}_{8T}$, *M3a - M5a, M7a* corresponding $\underset{\sim}{g}^a_{3T}$ to $\underset{\sim}{g}^a_{5T}$ and $\underset{\sim}{g}^a_{7T}$ when $\underset{\sim}{\beta}$ is assumed known, and finally *M7b* corresponding $\underset{\sim}{g}^b_{7T}$, - see section 3 for the definition of the estimators. We used a simple linear regression model for the synthetic component with the auxiliary variable defined as

$$X_{kt} = (\hat{\theta}_t/\theta_1)\,\theta_{k1} \tag{4.6}$$

where θ_{k1}, θ_1 respectively denote the population totals for small area k and the province at t=1, i.e. at Census'86. The estimator $\hat{\theta}_t$ denotes the post-stratified estimator of θ_t from the farm survey at time t at the province level. Thus, X_{kt} is simply a ratio-adjusted synthetic variable. The variances of error components in the regression model were assumed to be constant over areas. For time series models, it was assumed that the serial dependence was generated by a random walk. The above type of model assumptions have been successfully used in many applications and the main reason for our choice was considerations of simplicity. It was hoped, however, that the chosen models might be adequate for our purpose and might illustrate the desired differential gains with different types of models.

Since the Census'86 data was included in the time series, the direct estimate $\underset{\sim}{g}_{21}$ corresponds to Census'86 and therefore the survey error $\underset{\sim}{\epsilon}_1$ would be identically $\underset{\sim}{0}$. Moreover, from the definition of X_{kt}, it follows that a reasonable choice of (β_1, β_2) would be (0,1) which implies that $\underset{\sim}{a}_1$ must be $\underset{\sim}{0}$. Thus the covariance matrices B_t and W_t at t=1 are null and therefore, the distribution of $\underset{\sim}{\alpha}_t$ at t=1 would not require estimation as was suggested in section 3. The above modification in the initial distribution of α_t is natural in view of the extra information available from the census. Moreover, since the direct estimates $\tilde{y}_{2t}$ were not available for $t=2,3,4$; equations for estimating various model parameters were modified accordingly. For methods *M3a* - *M5a* and *M7a*, the value of β_t for $t\geq 2$ was fixed at (0,1) i.e. the same value corresponding to $t=1$.

4.4 Empirical Results

The main results are shown in Figures 1 to 4. Figure 1 shows plots of the five evaluation measures averaged over small areas relative to the *FH* (*M4*) value. There is a clear pattern in the behaviour of various measures across different estimators. The direct estimator *M2* does very well with respect to $\overline{\text{ARB}}$ and $\overline{\text{RMSCRE}}$. The time series methods *M7* (also *M7a*; *M7b*) and *M8* perform somewhat worse than *M2* with regard to bias, but overall they perform best. The *FH* estimator *M4* (*M4a*), sample size dependent estimator *M5* (and *M5a*) and the first time series method *M6* are almost at par. Both the expansion estimator *M1* and the synthetic estimator *M3* (and *M3a*) have very large conditional biases. We have not shown the Monte Carlo standard errors in the figures but they are all found to be quite negligible. Figures 2 to 4 show plots of RMSE_k for small areas divided into three size groups, namely low, medium and high, based on the ranking of their true population totals at time T.

The main conclusions are listed in Section 1 and will not be repeated here.

5. CONCLUDING REMARKS

It was seen by means of a simulation study that small area estimation methods obtained by combining both cross-sectional and time series data could substantially improve performance of estimators based only on cross-sectional data. The models for the study were chosen on general considerations. However, in practice, suitable diagnostics similar to those employed in Pfeffermann and Barnard (1991) should be performed before any model-based method can be recommended. It should also be noted that the small area estimators can be modified to make them robust to misspecification of the underlying model; see e.g. the constraints used in Fay and Herriot (1979)and an alternative approach suggested by Pfeffermann and Burck (1990). Further extension of the methods presented in this paper to the more realistic case of correlated sampling errors is currently being investigated.

ACKNOWLEDGEMENT

We would like to thank Jon Rao and Danny Pfeffermann for useful discussions and comments on an earlier version of this paper. The first author's research was supported in part by a grant from Natural Sciences and Engineering Research Council of Canada held at Carleton University, Ottawa. Thanks are also due to Judy Clarke, Christine Larabie and Carmen Lacroix for their efficient processing the manuscript.

REFERENCES

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error-components model for prediction of country crop areas using survey and satelite data. Journal of the American Statistical Association, 83, 28-36.

Bell, W.R., and Hillmer, S.C. (1987). Time series methods for survey estimation. Proceeding of the Statistics Association, Sec. Survey Res. Meth., 83-92.

Binder, D.A., and Dick, J.P. (1989). Modelling and estimation for repeated surveys. Survey Methodology, 15, 29-45.

Choudhry, G.H., and Rao, J.N.K. (1988). Evaluation of small area estimators: an empirical study. Paper presented at the Symposium or Small Area Statistics, New Orleans, Aug. 26-27.

Choudhry, G.H., and Rao, J.N.K. (1989). Small area estimation using models that combine time series and cross-sectional data. Proceedings of the Statistics Canada Symposium on Analysis of Data in Time, eds. A.C. Singh and P Whitridge, 67-74.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. Journal American Statistics Association, 74, 269-277.

Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. Journal Business Economics Statistics, 9, 163-175.

Pfeffermann, D., and Barnard, C.H. (1991). Some new estimators for small area means with application to the assessment of farmland values. Journal Business Economics Statistics, 9, 73-84.

Pfeffermann, D., and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. Survey Methodology, 16, 217-237.

Platek, R., Rao, J.N.K., Särndal, C.E., and Singh, M.P. eds (1987). Small Area Statistics: An International Symposium; New York; John Wiley & Sons.

Rao, J.N.K. (1986). Synthetic estimators, SPREE and best model-based predictors of small area means. Technical Report, Laboratory in Statistics and Probability, Carleton University, Ottawa.

Särndal, C.E., and Hidiroglou, M.A. (1989). Small domain estimation: a conditional analysis. Journal American Statistical Association, 69, 676-678.

Scott, A.J., and Smith, T.M.F. (1974). Analysis of repeated surveys using time series methods. Journal American Statistical Association, 69, 674-678.

Tiller, R. (1989). A Kalman filter approach to labor force estimation using survey data. Paper presented to the Annual American Statistical Association Meeting, Washington, D.C.

Figure 1: evaluation measures relative to Fay–Herriot

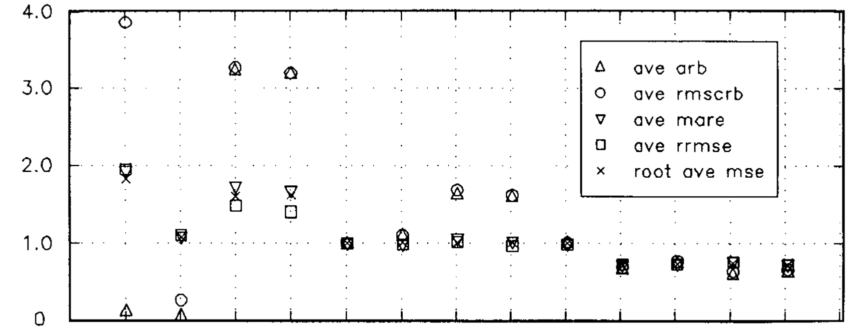

Figure 2: root mean squared errors, small small areas

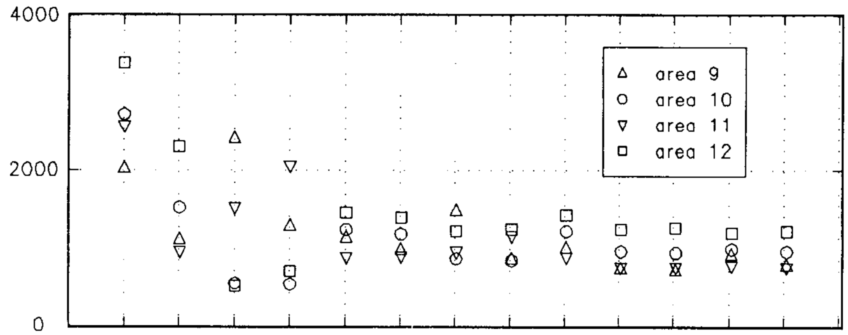

Figure 3: root mean squared errors, medium small areas

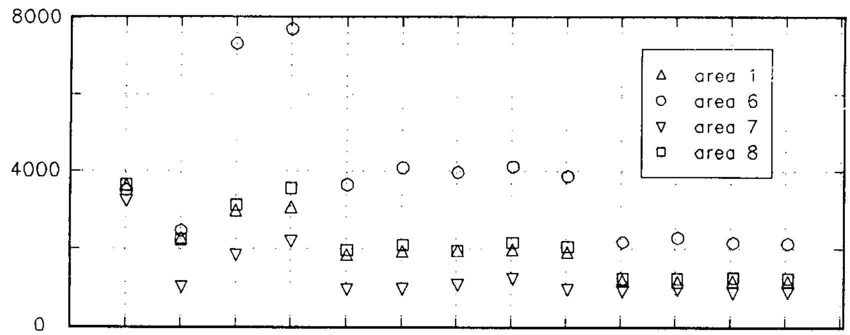

Figure 4: root mean squared errors, large small areas