Some Problems and Proposed Solutions in Developing a Small Area Estimation Product for Clients

A.C. Singh
Household Survey Methods Division, Statistics Canada, Ottawa K1A 0T6 avi.singh@statcan.ca

Abstract

With the goal of developing a small area estimation (SAE) product for clients, several user needs were identified. The basic area level mixed linear model was taken as the starting point upon which several enhancements were built to incorporate desired features based on user needs. One of the main research contributions of this SAE product development endeavor is a generalization of Kalman Filter (KF) to two or higher dimensions (corresponding to area and outcome variables besides time) for computational feasibility and statistical efficiency under multivariate modeling. For this purpose, the area and outcome variables are ranked in some ad hoc manner for rank-order to serve as a pseudo-time variable so that the usual theory of KF can be applied. Use of KF also gives rise to a versatile set of diagnostics based on innovations which are extremely useful to overcome the masking effect present in the usual BLUP residuals. This can be serious in the case of correlated random effects. In this paper, the new SAE product of Statistics Canada termed BUPF (signifying Best Unbiased Prediction via Filtering) is briefly described. It is observed that taking on the task of developing a product (the term product is defined as a menu-driven software system customized for clients and requiring minimal statistical training) unlike a tool (requiring statistical skills to operate) can be mutually beneficial for both researchers and clients. It helps researchers to take a disciplined path with deliverables, set priorities, and receive client input and validation at early stages, while it helps clients to have an understanding and confidence about the product and to plan for data gaps and customer needs in advance. The BUPF product was successfully pilot-tested with the 2003 Labour force Survey (LFS) data for producing annual averages of monthly employment total estimates for three digit occupation codes by province. Several other applications are planned in future in light of increasing user demands.

Key Words: Benchmarking; Client Input; Collapsing; Innovation Diagnostics; 2-dim Kalman Filtering; Multivariate Modeling; Smoothing; Statistical Products

1. Introduction

The present research is an outgrowth of the new initiative SPORD (Statistical Product Oriented R&D) for SAE currently underway at Statistics Canada for the period 2005-07. SPORD signifies a client customized menu-driven software system with several features depending on the client need. After initial discussions with the client (Labour Market Information Working Group of the Forum of Labour Market Ministers), several features of the SAE-product were established and it became apparent that none of the alternative available products (such as PROCMIXED in SAS, MLwiN, and WinBUGS) satisfy the methodology requirements of the desired product. With this in mind, we embarked on an intensive R&D, and were successful in completing a pilot test (Singh and Verret, 2006a) in Spring ’06 for the Labour Force Survey (LFS) for estimating provincial employment by 3-digit occupation codes under contract from the client.

The initial process of identifying client needs led to the following requirements:

1. The SAE methodology should take full account of the survey design to make it robust to model misspecification for nonignorable designs, and for obtaining design consistent estimates for validation by local area knowledge.
2. It should be amenable to good model diagnostics and should make Least amount of Assumptions in Modeling for Prediction (LAMP) in the interest of robustness to departures from them.
3. It should be able to produce estimates for areas with very few or no observations.
4. It should allow for multivariate modeling for internal consistency (i.e., SAE of a sum of outcome variables is sum of SAEs) as well as for improved efficiency for correlated outcome variables.
5. It should be computationally feasible for large multi-dimensional data involving time, area and outcome dimensions.
6. The estimates should be benchmarked to a set of reliable direct estimates for key large areas or suitable subgroups of small areas for robustification to potential model breakdowns, maintaining face validity, and avoiding possible over-shrinkage; as well as standard errors and confidence intervals associated with the estimates.

The LAMP requirement turned out to be an overarching principle for the product methodology requirements.

In order to meet the above requirements, it was decided for simplicity to start with the multivariate version of the basic area level mixed linear model of Fay and Herriot (1979) and build several enhancements. It is known that it takes full account of the survey design via design-based covariance matrix of the vector of observation errors in the direct estimates. The approximate assumption of normality of direct estimates with known covariance matrix except possibly for over/under dispersion can be reasonably well satisfied after collapsing and smoothing whenever necessary as suggested by Singh, Folsom, and Vaish (2005). The underlying framework is basically semiparametric with minimal assumptions, and the resulting estimates have the desirable property of being empirical best unbiased predictor as well as shrinkage-type estimator lying between direct and indirect estimates—something easily understandable by clients. Moreover, even for discrete outcome variables, the above area level linear mixed model framework is known to work well in general.

One of the main research contributions of this SAE product development endeavor is a generalization of Kalman Filter (KF) to two or higher dimensions (corresponding to area and outcome variables besides time) for computational feasibility and statistical efficiency for multivariate modeling. For this purpose, the area and outcome variables are ordered in some ad hoc manner for rank order to serve as a pseudo-time variable so that the usual theory of KF can be applied. Use of KF also gives rise to a versatile set of diagnostics based on innovations which are extremely useful to overcome the masking effect present in the usual BLUP residuals. The masking effect can be marked when the random area-specific effects have spatial auto-correlation. Other main contributions in the SAE product for enhancing the basic area level model correspond to smoothing of the error covariance structure while allowing for unknown over/under-dispersion, collapsing of areas with very few or no observations for estimating fixed parameters while providing area-specific estimates for random parameters, exact benchmarking, innovation diagnostics and covariate selection for linear mixed models based on both likelihood ratio test and variance component.

The organization of this paper is as follows. In Section 2, the nature of methodological problems in satisfying user requirements are examined, while in Section 3, proposed solutions for incorporating the desired features are described. Section 4 contains a brief description of the Statistics Canada's SAE product 'BUPF' currently being developed. Finally, Section 5 contains summary and directions of future work.

2. SAE Methodological Problems for Satisfying User Requirements

2.1 Full account of the survey design under weak assumptions

When the sample design is nonignorable for the (super-population) model, i.e., the model doesn't hold for the sampled observations, it is difficult in general to take full account of the survey design with the unit-level model without making strong parametric modeling assumptions; see e.g., the sample distribution and sample complement distribution approach of Pfeffermann and Sverchkov (2003). Although the approach of Pfeffermann and Sverchkov is quite ingenious and is optimal given the parametric assumptions, it requires extra modeling assumptions which may not be preferable by practitioners. Alternatively, with fixed effects models (linear or nonlinear), a reasonable alternative under semi-parametric assumptions (in the sense of first two moments) is to use optimal estimating functions (Godambe and Thompson, 1986; see also Binder 1983) involving sampling weights. This approach condenses the data in an optimal way but different from that based on sufficiency considerations. The above idea was extended by Singh, Folsom, and Vaish (2003) to nonlinear random effects models under a hierarchical Bayes framework with an approximate Gaussian likelihood. However, if the model is linear and at the area or aggregate level, then the frequentist framework under semiparametric assumptions can be used to take full account of the survey design by means of the design-based covariance matrix of direct estimates. This is similar to the method of Fay and Herriot (1979) and may be preferable in view of the overall LAMP principle underlying the methodology requirements.

More specifically, let the direct design-based estimator of the population total $T_{y,d}$ of the outcome variable y for the domain d be $t_{y,d}^{dir} = \sum_{k\in U_d} y_k w_k 1_{k\in s}$ based on the sample s where w_k is the calibrated sampling weight for the sampled unit k and U_d is the finite domain d universe, let $T_{x,d}$ be a column vector of q auxiliary variables with known population totals and $T_{c,d}$ be a D-vector of zero counts except for the *d*th position the known population count N_d for domain *d*. Now, for sample *s*, the survey weighted estimating functions for the fixed parameters β (*q* in all) and the area-specific random parameters $\eta_d, d = 1,...,D$ for the linear mixed area-level superpopulation model

$$y_{d,k} = A'_{x,d}\beta + \eta_d + \varepsilon_{d,k}\, , A_{x,d} = T_{x,d} / N_d\, ,$$
$$\varepsilon_{d,k} \sim NID(0,\sigma_\varepsilon^2)\, , \eta_d \sim NID(0,\sigma_\eta^2) \quad (2.1)$$

with $\varepsilon_{d,k}$ being independent of η_d, are given by

$$\psi_\beta = \sum_d A_{x,d} \sum_{k\in U_d} (y_{d,k} - A'_{x,d}\beta - \eta_d) w_k 1_{k\in s} \quad (2.2a)$$

$$\psi_{\eta,d} = \sum_{k\in U_d} (y_{d,k} - A'_{x,d}\beta - \eta_d) w_k 1_{k\in s} \quad (2.2b)$$

where only contributions from the observed data, and not from the unobserved prior information about the random effects are included. This shows that the sample data can be condensed into summary statistics $\{t_{y,d}^{dir}, d = 1,...,D\}$ assuming that $\hat{N}_d = N_d$, i.e., the domains are used as post-strata for weight calibration. The above derivation of summary statistics via EFs would be useful in motivating the idea of collapsing when dealing with areas with few or no observations; see Subsection (2.3) below. Now, the model taking account of the survey design can be written in terms of the summary statistics as $t_{y,d}^{dir} = T'_{x,d}\beta + T'_{c,d}\eta + e_d$, or in vector form,

$$t_y^{dir} = T(x)\beta + T(c)\eta + e \qquad (2.3)$$

where $e \sim (0,V),\ T(c)\eta \sim N(0,\Gamma)$ $T(c) = diag\{N_d\}$, V is the design-based covariance matrix of t_y and Γ is the diagonal matrix $T(c)\Gamma_\eta T(c)', \Gamma_\eta = \sigma_\eta^2 I$. Here it is assumed that $T_y \approx T(x)\beta + T(c)\eta$. Note that in the presence of spatial correlation, Γ will not be diagonal. Observe that the reduced model (2.3) is analogous to the two stage empirical Bayes model of Fay-Herriot (see Rao, 2003, Ch 7, for a good review) given by

Sampling model : $t_y^{dir} = T_y + e$ (2.4a)

Linking model : $T_y = T(x)\beta + T(c)\eta$ (2.4b)

2.2 Simple Interpretability of SAEs

Under the aggregate-level model (2.3), the usual estimator t_y^{sae} has the desirable optimality of being EBLUP (empirical best linear unbiased predictor or the stronger property of EBUP under normality) as well as the shrinkage-type estimator lying between direct (t_y^{dir}) and indirect ($T(x)\hat{\beta}$) estimates; here the term indirect is used for the model-based estimator (also known as the synthetic estimator) if there were no available data from the domains. The above shrinkage property makes the estimator easily understandable and interpretable in practice, and thus may be preferable for the desired product; the estimator t_y^{sae} is given by

$$t_y^{sae} = T(x)\hat{\beta} + T(c)\hat{\eta} \qquad (2.5a)$$

$$\hat{\beta} = (T(x)'W^{-1}T(x))^{-1}T(x)'W^{-1}t_y^{dir},$$

$$T(c)\hat{\eta} = \Gamma W^{-1}(t_y^{dir} - T(x)\hat{\beta}) \qquad (2.5b)$$

where $W = V + \Gamma$, and a consistent estimator such as REML $\hat{\sigma}_\eta^2$ is substituted in computing Γ .

2.3 Estimation for Areas with Few or No Observations

For the linear mixed model (2.3), it is assumed that the design-based covariance matrix V is known or has a stable estimate $\hat{V}$. This is clearly violated if some areas have few or no observations. To overcome this problem, Singh et al. (2003) suggested collapsing of areas based on some subject matter considerations or finding collapsing partners with similar random effects; for this purpose they suggested finding provisional estimates of random effects using individual observations while ignoring the design. This solution was motivated from the basic estimating functions (2.2a &b) and was based on the observation that the reduced model after collapsing of estimating functions should have, in general, sufficient degrees of freedom for estimating fixed first and second order parameters (β's$,\sigma_\eta^2$) while the random effects, although outnumbering the collapsed set of summary statistics, can still be estimated separately using BLUP theory because of the prior information; see Appendix I. This follows from the BLUP EFs given by (re: Henderson et al. 1959, Rao 2003, Ch 6, pp. 97)

$$\begin{pmatrix} T(x)' & O \\ T(c)' & I \end{pmatrix} \begin{pmatrix} V & O \\ O & \Gamma_\eta \end{pmatrix}^{-1} \begin{pmatrix} t_y^{dir} - T(x)\beta - T(c)\eta \\ 0 - \eta \end{pmatrix} = 0$$

However, it would be desirable to have an objective criterion to choose between collapsing partners. It may be noted that in the original two stage formulation (2.4a&b) of Fay-Herriot, it does not follow that collapsing would still lead to separate estimates for each area because in the first stage itself, collapsing would reduce the dimension of the small area parameter vector T_y. Also dropping the areas with zero or small sample sizes instead of collapsing would violate the BLUP principle as the corresponding SAEs would consist of only the model-based component. Moreover, we can't have built-in benchmarks (see subsection 2.5) if some areas are dropped from modeling.

Note that a suitable collapsing procedure would also make it legit to assume approximate normality of revised observation errors $\tilde{e}_d$ from (2.3). With approximate normality, it would be possible to develop several model diagnostics as discussed in Subsection 2.7. Moreover, it will give rise to a stronger optimality of EBUP for SAEs and allow us to use a simplified estimate of MSE adjusted for estimation of variance components along the lines of Prasad and Rao (1990) and Datta et al. (1992). It may also be noted that collapsing has the added benefit of stabilizing the error covariance structure.

2.4 Specification of the Design-based Error Covariance Structure

Even after suitable collapsing, the estimated error covariance structure $\hat{V}$ may not be stable enough. Suppose the estimated covariance matrix V^* based on the working assumption of simple random sampling or that of ignorability of the design is stable. This can be reasonably well ensured in practice by suitable collapsing such that each area or collapsed subgroup of areas has a minimum number of observations. Instead of modeling both the elements of $E(\hat{V})$ and the error covariance structure of $\hat{V} - E(\hat{V})$, which may require stronger assumptions, it may be easier to just specify the mean function $E(\hat{V})$ which may be known from a priori considerations. This mean function can be used to smooth $\hat{V}$ to make it stable and thus treat it as approximately known. In fact, it is known that the estimated design effects (deffs) defined as the ratio of estimated true design-based variance and that under the working assumption tend to have common means over a subgroup of statistics. Using this empirical observation, Singh et al. (2005) proposed a smoothed $\hat{V}$ by using a common mean model for estimated generalized deff (or g-deffs, defined as the eigenvalues $\hat{\lambda}_d$'s of $V^{*-1}\hat{V}$, see Rao and Scott, 1981) over suitable subgroups. More specifically, let the Cholesky decomposition of V^* be $A'A$ where A is an upper triangular matrix. Now, observe that the g-deffs $\hat{\lambda}_d$'s are the eigenvalues of $(A')^{-1}\hat{V}(A)^{-1}$ and letting P_d's to be the corresponding eigenvectors, we have

$$\hat{V} = \sum_d \hat{\lambda}_d (A'P_d)(A'P_d)',$$
$$V^* = \sum_d (A'P_d)(A'P_d)' \qquad (2.6)$$

The smoothed estimate $\tilde{V}$ of V is then obtained by replacing $\hat{\lambda}_d$'s by averages $\tilde{\lambda}_g$'s over suitable subgroups g. Although the above method of smoothing is found to work well in limited simulation studies, it may be biased if the common mean model for estimated eigenvalues is not correct. Also there may be omitted random effects in the linear mixed model (2.3). It would therefore be desirable to correct for such over- or under-dispersion.

2.5 Benchmarking to reliable direct estimates

Typically, benchmarked SAEs are obtained post-modeling by using a regression or raking type adjustment to SAEs; see e.g., Fay and Herriot (1979), Battese, Harter, and Fuller (1988), Pfeffermann and Barnard (1991), and Rao and Choudhry (1995). With this kind of benchmarking (corresponding to parameter constraints on D small area parameters) defined independently of the q fixed parameters β's from the original model (2.3), there is an inherent incompatibility of the two specifications. Consequently, the resulting benchmarked estimates are no longer optimal. The corresponding MSE is increased by an extra term, $(t_{y,d}^{sae} - t_{y,d}^{bsae})^2$, bsae denoting the benchmarked SAE; the adjustment to the MSE being motivated from the asymptotic hierarchical Bayes equivalence to EBUP, see e.g., Datta et al. (1992). Moreover, approximate normal confidence intervals for areas with sufficient realized sample sizes remain applicable for regression-type adjustments but not necessarily for raking-type adjustments due to nonlinearity although raking may be preferable to avoid negative estimates. As an alternative, it would be desirable to have built-in benchmarks in the model specification (2.3) to protect against possible model breakdowns. It is interesting to note that since the benchmarks are random and based on the same data used for SAE modeling, the problem doesn't conform to a Bayesian framework with parameter constraints. On the other hand, a frequentist solution may be feasible.

2.6 Computational Feasibility of SAEs

Even for the simple linear mixed models for SAE, the computation of BUP requires inversion of the matrix W which may cause numerical instability or may not be computationally feasible for large number of domains (as in the case of large multi-dimensional domains defined by time, area and outcome variables) with an arbitrary covariance matrix V. It would be desirable to have a recursive method of computing BUP which updates it as data corresponding to each domain is incorporated. It is shown in the next section that the computationally efficient technique of Kalman Filtering (KF) can be adapted under certain conditions to the present problem after a suitable orthogonalizing transformation whether or not the data has a time dimension. Numerical stability is achieved because each updating step involves only scalar operations of multiplication and division.

2.7 Model Selection and Diagnostics

Besides the usual stepwise procedures of covariate selection for linear models without random effects, it would be desirable to bring in mixed model features in selecting covariates for fixed effects. For model diagnostics, it is known that the usual residual diagnostics for regression models with fixed effects may suffer from masking effect due to correlations induced by estimation of common fixed parameters. Also effect of some extreme observations may not manifest itself in the plot of standardized residuals. In the presence of correlated random effects, the correlations between BLUP residuals may not be negligible. In these situations, it is desirable to use an alternative set of residuals that are uncorrelated with each other. It is also known that innovations from KF provide such a set of orthogonal residuals. In the next section, we consider how these can be used to provide a versatile set of model diagnostics.

2.8 Computational Feasibility of estimating Variance Components and MSE adjusted for them

There may also be the problems of computational infeasibility as well as numerical instability with large multi-dimensional data over time, area, and outcome variables when estimating variance components by using REML and estimating MSE of SAEs adjusted for estimation of variance components; see Rao (2003, Ch 6) for computational formulas. It would be desirable as in the case of BUPs to use a recursive algorithm such as the one based on KF for this purpose. In the next section, we show this can indeed be done.

3. Proposed Solutions for Incorporating Desired Features in the SAE Product

3.1 Features of full accountability of design and simple interpretability of SAE under LAMP

In view of the methodology requirements 2.1 and 2.2, it was decided to use the basic aggregate level model (2.3) and the associated EBLUP estimator t_y^{sae} obtained under a semi-parametric framework. Although there is no built-in parameter restrictions to avoid SAEs being outside the range (e.g., employment proportions must be between 0 and 1), the shrinkage estimator is generally known to work well even with discrete outcome variables except possibly for rare outcomes. It would, however, be useful in future to generalize the proposed methodology to nonlinear models.

3.2 Feature of Objective Collapsing Criterion

For the requirement 2.3, we propose to use the objective criterion of minimizing the estimated variance (σ_η^2) of the model error in choosing between alternative areas for collapsing partners. This criterion is data-driven and is the proportionality constant in the signal to noise ratio ($N_d^2\sigma_\eta^2 / V_d$) appearing as an argument in the leading term of the MSE expression of EBLUP. We also propose to use collapsing if necessary to ensure approximate normality of the observation error. This will give rise to the stronger optimality property of EBUP for SAEs along with other benefits mentioned in the subsection 2.3.

3.3 Feature of Over/Under dispersion

For the requirement 2.4, we propose to introduce a scalar over-dispersion parameter (σ_0^2) as a multiplicative factor of the error covariance structure V as well as the random effects covariance Γ. Such a specification is known to work well in general to capture unaccounted clustering effects; see e.g., McCullagh and Nelder (1989, pp. 121-4). Note that in the presence of the overdispersion parameter, the objective criterion for collapsing should be modified to $\sigma_0^2\sigma_\eta^2$.

3.4 Feature of Built-in Benchmarks

For the requirement 2.5, we propose to enlarge the model (2.3) by adding fixed effects corresponding to suitably transformed indicator covariates for desired benchmark subgroups of areas. Here it is assumed that the total number of domains after collapsing, if any, is much larger than the total number of revised fixed effects (including those for benchmarks) in order to have sufficient degrees of freedom for consistent estimation of variance components. To see how exact benchmarking works, it is observed that if the additional covariates $V1_b$ (where the dth element of 1_b is 1 if $d \in U_b$, U_b denotes the bth large area or subgroup of small areas for benchmarking, $b = 1,\dots,B$, B being the total number of benchmarks), are included in the model (2.3), then the benchmark condition is satisfied if

$$1_b'(t_y^{dir} - T(x)'\hat{\beta} - T(c)'\hat{\eta}) \\ 1_b'VW^{-1}(t_y^{dir} - T(x)'\hat{\beta}) = 0 \qquad (3.1)$$

using the result (2.5b). The above relation is indeed true in view of the estimating equation for β given by $T(x)W^{-1}(t_y^{dir} - T(x)'\hat{\beta}) = 0$ and the fact that the new covariates $V_d 1_{d\in U_b}$ are elements of the q-vector $T_{x,d}$ for each d. It is interesting to note that the above result on benchmarking is similar to the one reported by Maiti (2006). While the above result on exact benchmarking is appealing as it preserves the optimality of SAEs in a wider class, it requires introduction of additional covariates as a function of the covariance matrix V which may be deemed as an artifact rather than being legitimate predictors. As an alternative, it would be useful in future to investigate ways to modify the BLUP estimating functions themselves so that benchmark constraints are built-in.

3.5 Feature of Recursive Computation via Multi-dimensional Kalman Filtering

For the requirement 2.6, we rank-order the domains in an ad hoc manner (e.g., in the order of decreasing effective domain sample size) so that the domain rank serves as a pseudo-time variable for the state-space modeling of state parameters $\theta_d = \theta = (\beta', \eta')'$ consisting of all the fixed and random parameters and has dimension $q + D$. We can now define the state space framework as

Measurement Eqn : $t^{dir}_{y,d} = F_d\theta_d + e_d$ (3.2a)

Transition Equation : $\theta_d = \theta_{d-1}$ (3.2b)

where the covariate matrix F_d is simply $(T'_{x,d}, T'_{c,d})$ where $T_{c,d}$ is the vector of zero counts except for the domain population N_d as the *d*th element. Note that the state vector is defined to be time-invariant implying that the transition error in (3.2b) is identically zero. Now two conditions are needed for using Kalman filtering : First, conditional independence of measurement errors e_d given the state vector θ_d; this can be satisfied by making an orthogonalizing transformation (such as the one based on Gram-Schmidt) on the vector t^{dir}_y to make the measurement error covariance matrix V diagonal. Second, Markovian dependence of the state vectors θ_d over d. This is satisfied trivially as it is time-invariant.

The process of KF is an alternative computationally efficient way of obtaining BUP $\hat{\theta}_{(D|D)}$ and the corresponding naive MSE $\Sigma_{\theta(D|D)}$; see Appendix II for details. Unlike the usual KF, here we don't need the smoothing steps after filtering because fixed-interval smoothing is built-in as the state vector is time-invariant. To compute BUP, we need prediction and updating equations or equivalently prior and posterior distributions via innovations when the variance components $(\sigma^2_0, \sigma^2_\eta)$ are assumed known. The initial proper prior is defined by sacrificing the first few observations, i.e., as many as the number q of fixed parameters $\beta's$. At $d = q+1$, the prediction step corresponds to the prior of θ_{q+1} based on past q observations and is given by

$$\left[\theta_{q+1}\left|q\right.\right] = N_{q+D}\left(\hat{\theta}_{q+1|q}, \Sigma_{\theta(q+1|q)}\right) \quad (3.3a)$$

while the innovation step provides the next independent piece of information in the new observation $t_{y,q+1}$ at $d = q+1$, and is given by

$$\nu_{q+1} = t^o_{y,q+1} - F^o_{q+1}\hat{\theta}_{(q+1|q)} \quad (3.3b)$$

where the superscript 'o' denotes that the orthogonalizing transformation has been applied, and finally, the updating step corresponds to the posterior of θ_{q+1}, and is given by

$$\left[\theta_{q+1}\left|q+1\right.\right] = N_{q+D}\left(\hat{\theta}_{(q+1|q+1)}, \Sigma_{\theta(q+1|q+1)}\right) \quad (3.3c)$$

The above steps of prediction, innovation, and updating are continued recursively until we reach $d = D$, and obtain

$$\left[\theta_D\left|D\right.\right] = N_{q+D}\left(\hat{\theta}_{(D|D)}, \Sigma_{\theta(D|D)}\right).$$

If the linear mixed model is multivariate, we propose a two-dimensional KF for computing BUPs and MSE. The related reference of Fay (1987) and Datta et al. (1992, 1998) consider the usual BUP approach without filtering. Here it is assumed that $D \gg q + J$ where J is the dimension of the multivariate outcome. For the two-dim KF with the domain subscript $d = 1,...,D$, and the multivariate outcome subscript $j = 1,..., J$, we first order all the data points lexicographically (say) as

$(1,1) \to (1,2) \to ...(1, J) \to (2,1) \to ...(D, J)$, and then apply an orthogonalizing transformation to make the $DJ \times DJ$ matrix V diagonal. The basic idea of the two-dim KF entails two sets of transition equations, one over the outcome dimension given the area, and the other over areas given the outcome. Now sacrifice the first few observation vectors to get the prior $\left[\theta_{(q+1,1)}\left|(q, J)\right.\right]$ at $(q+1,1)$, and continue filtering until we get

$$\left[\theta_{(D,J)}\left|(D, J)\right.\right] = N_{(q+D)J}\left(\hat{\theta}_{(D,J)|(D,J)}, \Sigma_{\theta(D,J|D,J)}\right) \quad (3.4)$$

If we need to combine longitudinal and cross-sectional data for multivariate SAE, then we can as before order the data $t_{(y,r,d,j)}$, $r = 1,..., R; d = 1,..., D; j = 1,..., J$ over three dimensions suitably, and then use a three-dim KF. Here one of the dimensions is in fact time, and so we will first use that to order data over occasions $r = 1,..., R$, and then within each *r*, we can order (*d* ,*j*) as before.

3.6 Features of Variance Component-based Covariate Selection and Innovation Diagnostics

For model selection requirement 2.7, we divide the covariates into groups such as one-factor effects and two-factor effects in a hierarchical manner. The model is then enlarged by adding groups in a stepwise manner using the likelihood ratio test. Now before adding the next group, factors within a group are ranked based on the significance probability under a backward procedure for dropping factors. Next at each step of inclusion of the most significant factor, the decision is based on whether there is appreciable decrease in the estimated variance component. Thus the mixed model feature is brought into the covariate selection.

For the model diagnostics requirement, we propose to use innovations $\nu_d : d = q+1,...D$ of (3.3b) suitably standardized for several diagnostics (see e.g., Harvey, 1989, Ch 5, p. 236) such as chi-square test for over/under dispersion, random pattern in the standardized innovation plots over areas for different orderings of areas, Q-Q plot for normality, Ch-square tests for goodness of fit and for model adequacy, and CUSUM test for any systematic over/under prediction as well as for any structural change. For measuring impact of outlier observations, case deletion-type diagnostics could be performed. For this purpose, first randomly assign observations deemed to be outliers toward the end of ordering and then check the innovation plot to see if they continue to be outliers. Also check coefficient sensitivity (for fixed regression coefficients and variance components) by using partial data (or innovations) when observations deemed as outliers are placed at the end of ordering. Finally, prediction or post-sample diagnostics, somewhat similar to cross-validation, can be performed. For this purpose, we can take 100 or so random reorderings of domains or observations and compute (one-step ahead) prediction errors

$$t_{y,d} - (T'_{x,d}\hat{\beta}_{(d|d-1)} + T'_{c,d}\hat{\eta}_{(d|d-1)}),\ d = q+1, ..., D \tag{3.5}$$

and then plot the average absolute relative prediction errors against sample sizes used for prediction to check for a decreasing trend.

3.7 Features of KF recursions for estimating variance components and MSE of SAEs

For the requirement 2.8, we first show how REML can be computed using innovations. Here for the sake of simplicity, we will assume that the overdispersion parameter σ_0^2 is known. The proposed method is similar to the one used by Sallas and Harville (1988). First we sacrifice the first q observations to get a proper initial prior and then define the normal log-likelihood of (D-q) innovations as

$$\log L = -\frac{1}{2}\left[\sum_{d=q+1}^{D} \frac{\nu_d^2}{\sigma_{\nu(d)}^2} + \sum_{d=q+1}^{D} \log \sigma_{\nu(d)}^2\right] + const. \tag{3.6}$$

where $\sigma_{\nu(d)}^2$ is the variance of ν_d. Now REML is simply the MLE of σ_η^2 based on the above likelihood obtained by Newton-Raphson iterations. The gradient and Hessian at each iteration can be obtained recursively using the standard KF theory. Note that a suitable approximation to the Hessian can be obtained from only first derivatives of innovations and their variances.

Finally, we propose to compute estimated MSE of SAEs adjusted for variance component estimation using KF. This is an area of considerable activity; important related references in the non-KF context are Kackar and Harville (1984), Prasad and Rao (1990), Datta et al. (1992), Singh, Stukel, and Pfeffermann (1998), Datta and Lahiri (2000), while in the KF context is Quenneville and Singh (2001). It can be shown that under regularity conditions of Datta and Lahiri (2000), the second order adjusted MSE of $t_y^{sae} = F\hat{\theta}_{(D|D)}^{bup}$ is given by

$$MSE(t_y^{sae}) \doteq F\Sigma_{\theta(D|D)}F' + \left[F\left(\frac{\partial}{\partial\sigma_\eta^2}\hat{\theta}_{(D|D)}\right)\left(\frac{\partial}{\partial\sigma_\eta^2}\hat{\theta}_{(D|D)}\right)' F'\right] I^{-1}(\sigma_\eta^2) \tag{3.7}$$

where *F* is the *D x(q+D)* matrix $(T(x), T(c))$.

The required derivatives can be easily computed using KF recursions. The inverse of the information matrix is already available from the Newton-Raphson iterations for REML. For estimating MSE, the estimate of the variance component is plugged in as well as the second term in (3.7) is multiplied by a factor of *2* to account for downward bias analogous to the original result of Prasad and Rao (1990).

4. Statistics Canada's SAE Product BUPF : A Brief Description

The SAE product (termed BUPF to signify best unbiased prediction via filtering) of Statistics Canada is designed to meet a number of user requirements. A pilot test of an initial version was successfully conducted for producing employment estimates at the three digit occupation code for LFS. The SAE product consists of twelve linked modules which can be run interactively. In the following, these modules are described briefly in terms of the LFS application mentioned above.

Module 1 (Data Specification) In this module, the user is asked to define SAE parameters such as annual averages of monthly total employed by 3-digit occupation codes by province, target variables such as employment status, and target domains such as 3-digit occupation codes by province; the target domain may be multi-dimensional, e.g., the two dimensions of occupation and province in this case. Next the user is asked to define direct data sources such as the cross-sectional 2003 LFS data for province of Newfoundland and Labrador, and indirect data source for each direct estimate such as again the cross-sectional 2003 LFS data for provinces other than the target province of Newfoundland and Labrador, or LFS data from past years for the same target province. Next information about auxiliary sources is needed such as demographic counts at province or subprovincial level, taxfiler counts by gender, and employment insurance beneficiaries data by gender and occupation. Finally, small area modeling domains (may be different from target domains) are specified such as province or subprovince; note that the number of modeling domains should be sufficiently large for adequate modeling purposes. Also, modeling variables are specified such as the outcome of total employed by occupation codes and covariates such as the demographic counts, taxfiler counts, and EI beneficiary counts, all at the modeling domain level.

Module 2 (Task Specification) Here the user is asked to choose outcome variables for a given modeling problem such as employment for the code H52-Printing Operators and commercial divers, auxiliary variables among possible ones such as the taxfiler counts by ER, and specify the two variance-covariance matrices for the error covariance structure in the direct estimates, one under the actual design, and the other under the working design of simple random sampling.

Module 3(Benchmark Constraints) Here the user is asked to choose subgroups of modeling domains such as the overall national subgroup whose direct estimates are desired to be equal to the sum of the SAEs of the domains in the subgroup. The sample size in the subgroup should be large enough for the direct estimate to be precise.

Module 4 (Domain Collapsing) Here the user is asked, based on subject matter considerations, how to choose collapsing partners for domains with very small or no sample size or with zero estimates due rareness of the outcome of interest. Technically speaking, the partners should be chosen with similar domain-specific random effects. In practice, however, areas with similar economic activity and geographic proximity might serve as good candidates for collapsing partners. The objective criterion of minimizing the estimated variance of the model error is used to choose between potential collapsing partners.

Module 5 (Variance Smoothing) This module computes estimated true design-based error covariance structure and working design-based covariance for direct estimates corresponding to collapsed domains. Then suitable eigen-values and eigenvectors are computed for variance smoothing.

Module 6 (Model Selection) Here the model with an initial set of covariates is enlarged to account for benchmarks. Scatter plots are used to select the initial set of covariates. Then likelihood ratio tests combined with variance component estimates are used to choose between covariates.

Module 7 (Variance Components) In this module, innovations from Module 8 are used to define the likelihood ratio and then REML (restricted maximum likelihood) estimates of variance components are computed. This module is rerun if the model is revised under Module 6.

Module 8 (Innovation Sequence) Given the selected model from Module 6, and given variance components from Module 7 and with a flat prior on fixed parameters, this module computes innovation in sequence and corresponding variances. The Kalman filter is used after the linear mixed model is cast into a state-space framework.

Module 9 (Model Diagnostics) Here, for a given ordering of domains such as that given by decreasing order of realized domain sample sizes, standardized innovations are plotted to check for trends and potential outliers. Also CUSUM (cumulative sum) of standardized innovations are used to test for any structural changes.

Module 10 (SAE) This uses a two-step approach. First, filtering (from Module 8) is used to compute innovations for estimating variance components and, second, state parameters consisting of fixed and random effects are estimated. SAEs are then computed as BUPs (best unbiased predictors) via filtering, hence the name BUPF for the SAE system.

Module 11 (Evaluation of SAEs) Here, SAEs are checked for how well they satisfy the benchmarks as well as internal and external validation (whenever possible based on local area knowledge) of SAEs are performed. For internal validation, large areas are randomly split to create pseudo small areas and their SAEs are checked against large area direct estimates. Also effect of potential outliers identified from innovation plots is analyzed by comparing SAEs with and without outliers.

Module 12 (Overall Summary) This consists of point estimates (direct, indirect and SAE), their SE adjusted for the downward bias due to estimated variance components, and plots before and after modeling. Diagnostic measures such as CUSUM tests and R^2-type goodness-of-fit measures are provided with cautionary remarks if any. Also if necessary, results for both with and without outliers are presented.

After each module is run, user is asked to review the choices and results and confirm to proceed to the next module. The modular nature of the BUPF system allows for all modules to be interconnected and any module can be rerun at any stage of model building.

5. Summary and Future Work

In this paper, a methodology based on best unbiased prediction via filtering for SAE under linear mixed models at the area level was proposed for computational and statistical efficiencies with several desirable features such as benchmarking for robustification to model breakdowns, face validity, and avoiding overshrinkage, collapsing to produce SAEs for areas with few or zero observations, filtering for versatile diagnostics, and multivariate modeling for internal consistency and improved efficiency. It is based on a suitable generalization of Kalman Filter to two and higher dimensions when the data is not necessarily a time series. The alpha version of the BUPF product is planned to be completed in Fall '06 while a more efficient beta version in Spring '07. Next year, applications of the BUPF product are being planned for production of SAEs for the Labour Force Survey and the Canadian Community Health Survey .

In future, it would be useful to develop a generalization of BUPF to the case of nonlinear models so that suitable range restrictions on SAE parameters could be built-in; see Singh and Verret (2006b) for some preliminary results. It would also be useful to investigate ways to modify estimating functions for BLUP so that benchmark constraints are automatically satisfied without enlarging the model.

Acknowledgments : The author is grateful to late Dr. M.P. Singh, Jack Gambino, Don Royce, Jean-Louis Tambay and Jon Rao for their crucial support and encouragement at the initial stages of this research endeavor. Thanks are also due to Francois Verret and Claude Nadeau of Statistics Canada for several useful discussions. This research was partially supported by Statistics Canada's Methodology Block Fund, Labour Statistics Division, FLMM-LMI working Group, and a grant from NSERC of Canada held at Carleton University under an adjunct research professorship.

Appendix I

Separate Estimates of Random Effects for Small Areas in the Collapsed Subgroup

Collapsing can be viewed as subdividing the target population into a smaller number of domains such that sample sizes for the revised domains are sufficiently large for smoothing variance of revised direct estimates as well as their approximate normality for effective model diagnostics. Suppose collapsing is done based on prior considerations, and thus is not data-driven. This is often reasonable in practice for repeated surveys because past data are available. Then despite having domains with few or no observations, the direct estimate of the collapsed domain would be approximately unbiased and normal under general conditions. Now, the small area modeling with collapsed domains does preserve the identities of area-specific random effects. Therefore, separate BLUP estimates can be obtained (assuming for simplicity that the covariance V is diagonal) by simply apportioning the observed residual $t^{dir}_{y,\tilde{d}} - T_{x,\tilde{d}}\hat{\beta}$ (where $\tilde{d}$ denotes the collapsed domains consisting of small area modeling domains $d \,\&\, d'$, say,) to random effects $N_d\eta_d \,\&\, N_{d'}\eta_{d'}$ relative to their contributions to the variance similar to the usual formula (2.5b).

More specifically, suppose the total variance (i.e., including both model and the observation error) for the direct estimate $t^{dir}_{y,\tilde{d}}$ about the model mean $T_{x,\tilde{d}}\beta$ is $(N_d^2 + N_{d'}^2)\sigma_\eta^2 + V_{\tilde{d}}$, then the BLUP estimates of $N_d\eta_d \,\&\, N_{d'}\eta_{d'}$ based on the reduced model after collapsing are obtained as

$$N_d\hat{\eta}_d = N_d^2\sigma_\eta^2[(N_d^2 + N_{d'}^2)\sigma_\eta^2 + V_d]^{-1}(t^{dir}_{y,\tilde{d}} - T'_{x,\tilde{d}}\hat{\beta}) \quad (A.1)$$

$$N_{d'}\hat{\eta}_{d'} = N_{d'}^2\sigma_\eta^2[(N_d^2 + N_{d'}^2)\sigma_\eta^2 + V_d]^{-1}(t^{dir}_{y,\tilde{d}} - T'_{x,\tilde{d}}\hat{\beta}) \quad (A.2)$$

It should be noted that the above estimates are optimal only in the reduced class, and would be suboptimal had there been no collapsing.

Appendix II

BLUP via KF

Using the formulation of extended least squares, the BLUP equations can alternatively be expressed as

$$X'_*\Sigma_*^{-1}(y_* - X_*\theta) = 0 \quad (A.3)$$

where the *2D-vector* $y_* = (t'^{dir}_y, 0)'$, the *(q+D)-vector* $\theta = (\beta', \eta')'$ while the *2D x (q+D)* matrix X_* and the *2D x 2D* matrix Σ_* are given by

$$X_* = \begin{pmatrix} T(x) & T(c) \\ O & I \end{pmatrix}, \ \Sigma_* = \begin{pmatrix} V & O \\ O & \Gamma_\eta \end{pmatrix} \quad (A.4)$$

It follows that the BLUP $\hat{\theta}^{BLUP}$ and the corresponding (naïve) $MSE(\hat{\theta}^{BLUP})$ are given by

$$\hat{\theta}^{BLUP} = \left(X'_*\Sigma_*^{-1}X_*\right)^{-1} X'_*\Sigma_*^{-1}y_* \quad (A.5a)$$

$$MSE(\hat{\theta}^{BLUP}) = \left(X'_*\Sigma_*^{-1}X_*\right)^{-1} \quad (A.5b)$$

Now the SAEs are given by $t^{sae}_y = F\hat{\theta}^{BLUP}$ where F is defined in (3.7) ,and the corresponding MSE can be obtained as

$$F\left(X'_*\Sigma_*^{-1}X_*\right)^{-1}F' =$$
$$(V - VW^{-1}V) + VW^{-1}T(x)\left(T(x)'W^{-1}T(x)\right)^{-1}T(x)'W^{-1}V \quad (A.6)$$

To obtain the above BLUP and its MSE via KF, first use an orthogonal transformation on the vector t^{dir}_y to obtain t^o_y such that the covariance matrix V^o becomes diagonal which in turn makes W^o diagonal. In the formulation of extended least squares, the observation vector now becomes y_*^o, the covariate matrix X_*^o, and the covariance matrix Σ_*^o. Clearly, after the transformation, the BLUP estimator of θ and its MSE don't change. Now to start the KF, rank order the domains such that the initial *(q+D) x (q+D)* covariate matrix for the first *q* observations is non-singular, and then for the initial distribution of θ, we have

$$\hat{\theta}_{q|q} = \left(X'^o_{*[q]} \Sigma^{o^{-1}}_{*[q]} X^o_{*[q]} \right)^{-1} X'^o_{*[q]} \Sigma^{o^{-1}}_{*[q]} y^o_{*[q]} = X^{o^{-1}}_{*[q]} y^o_{*[q]} \quad \text{(A.7a)}$$

$$\Sigma_{\theta(q|q)} = \left(X'^o_{*[q]} \Sigma^{o^{-1}}_{*[q]} X^o_{*[q]} \right)^{-1} = (X^o_{*[q]})^{-1} \Sigma^o_{*[q]} (X'^o_{*[q]})^{-1} \quad \text{(A.7b)}$$

where the subscript *[q]* simply denotes that only the first *q* observations are incorporated. Next, at *d=q+1*, we define

$$y^o_{*[q+1]} = \begin{pmatrix} y^o_{*[q]} \\ t^o_{y,q+1} \end{pmatrix}, \quad X^o_{*[q+1]} = \begin{pmatrix} X^o_{*[q]} \\ F^o_{q+1} \end{pmatrix} \quad \text{(A.8a)}$$

$$\Sigma^o_{*[q+1,q+1]} = \begin{pmatrix} \Sigma^o_{*[q,q]} & O \\ O & w^o_{(q+1,q+1)} \end{pmatrix} \quad \text{(A.8b)}$$

where $w^o_{(q+1,q+1)}$ is the *(q+1)st* diagonal element of $W^o_{(q+1,q+1)}$. Now the BLUP and its MSE using the first *(q+1)* observations are, after some algebra, given by (here the diagonal structure of the total covariance matrix W^o is crucial),

$$\hat{\theta}_{(q+1|q+1)} = \left(X'^o_{*[q+1]} \Sigma^{o^{-1}}_{*[q+1]} X^o_{*[q+1]} \right)^{-1} X'^o_{*[q+1]} \Sigma^{o^{-1}}_{*[q+1]} y^o_{*[q+1]} = \hat{\theta}_{(q|q)} + \Sigma_{\theta(q|q)} F'^o_{q+1} \sigma^{2^{-1}}_{\nu(q+1)} \nu_{q+1} \quad \text{(A.9a)}$$

$$\Sigma_{\theta(q+1|q+1)} = \left(X'^o_{*[q+1]} \Sigma^{o^{-1}}_{*[q+1]} X^o_{*[q+1]} \right)^{-1} = \Sigma_{\theta(q|q)} - \Sigma_{\theta(q|q)} F'^o_{q+1} \sigma^{2^{-1}}_{\nu(q+1)} F^o_{q+1} \Sigma_{\theta(q|q)} \quad \text{(A.9b)}$$

where ν_{q+1} is defined by (3.3b), and its variance by

$$\sigma^2_{\nu(q+1)} = w^o_{(q+1,q+1)} + F^o_{q+1} \Sigma^{o^{-1}}_{\theta(q|q)} F'^o_{q+1}. \quad \text{(A.10)}$$

The above recursions are continued to obtain $\hat{\theta}^{BLUP}$ as $\hat{\theta}_{(D|D)}$ and its MSE as $\Sigma_{\theta(D|D)}$. The algebra behind (A.9a &b) is briefly outlined below.

$$\hat{\theta}_{(q+1|q+1)} = \left(\Sigma^{-1}_{\theta(q|q)} + F'^o_{q+1} w^{o^{-1}}_{(q+1,q+1)} F^o_{q+1} \right)^{-1} \times \left(X'^o_{*[q]} \Sigma^{o^{-1}}_{*[q]} y^o_{*[q]} + F'^o_{q+1} w^{o^{-1}}_{(q+1,q+1)} t^o_{y,q+1} \right) \quad \text{(A.11)}$$

The matrix inverse in the first term on the R.H.S of (A.11) can be expressed using the inverse partitioned matrix formula as $\Sigma_{\theta(q+1|q+1)}$ (this is how (A.9b) can be obtained), and so we can simplify (A.11) as

$$\hat{\theta}_{(q|q)} + \Sigma_{\theta(q|q)} F'^o_{q+1} \sigma^{2^{-1}}_{\nu(q+1)} \times \{(\sigma^2_{\nu(q+1)} - F^o_{q+1} \Sigma_{\theta(q|q)} F'^o_{q+1}) w^{o^{-1}}_{(q+1,q+1)} t^o_{y.q+1} - F^0_{q+1} \hat{\theta}_{(q|q)}\} \quad \text{(A.12)}$$

which reduces to (A.9a) as desired.

References

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *JASA*, 83, 28-36.

Datta, G.S., and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10, 613-627.

Datta, G.S., Ghosh, M., Huang, E.T., Isaki, C.T., Schultz, L.K., and Tsay, J.H. (1992). Hierarchical and Empirical Bayes Methods for adjustment of Census Undercount: The1988 Missouri Dress Rehearsal data. *Survey Methodology*, 18, 95-108.

Datta, G.S., Day, B., and Maiti, T.P. (1998). Multivariate Bayesian Small Area Estimation: An application to survey and Satellite data. *Sankhya, Ser A*, 60, 1-19.

Fay, R.E. (1987). Application of multivariate regression to small domain estimation. In Platek, Rao, Sarndal, and Singh (Eds.) *Small Area Statistics*, New York, Wiley, 91-102.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to Census data. *Jour. Amer. Statist. Assoc.*, 74, 269-277.

Godambe, V.P., and Thompson, M.E. (1986). Parameters of superpopulation and survey population: their relationship and estimation. *Int. Statist. Rev*. 54, 127-38.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge Univ. Press.

Henderson, C.R., Kempthorne, O, Searle, S.R., and von Krosigk, C.N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 13, 192-218.

Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *JASA*, 79, 853-862.

Maiti, T.P (2006). Personal Communication. JSM '06, Seattle.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models*, Second Ed., Chapman and Hall.

Pfeffermann, D. (2002). Small Area Estimation—New Developments and Directions, *International Statistical Review*, 70, 125-143.

Pfeffermann, D. and Sverchkov, M. (2003). Fitting generalized linear models under informative sampling. *Analysis of Survey Data*, (eds. R.L. Chambers and C.J. Skinner), New York: John Wiley, 175-194.

Pfeffermann, D., and Barnard , C. (1991). Some new estimators of small area means with applications to the assessment of farmland values. *Jour. Bus. Eco. Statist.*, 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.

Prasad, N.G.N, and Rao, J.N.K. (1990). The estimation of mean squared error of small area estimators. JASA, 85, 163-171.

Quenneville, B., and Singh, A.C. (2001). Bayesian prediction mean square error for state space models with estimated parameters. *Jour. Time Series Analysis*, 21, No. 2, 219-236.

Rao, J.N.K. (2003). *Small Area Estimation*, New York: John Wiley.

Rao, J.N.K., and Choudhry, G.H. (1995). Small area estimation: Overview and empirical study, in B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott (eds.), *Business Survey Methods*, New York: Wiley, pp. 527-542.

Rao, J.N.K., and Yu, M. (1994). Small area estimation by combining time series and cross sectional data. *Canadian Jour. Statist*., 22, 511-528.

Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *JASA*, 76, 221-30.

Sallas, W.M., and Harville, D.A. (1988). Noninformative priors and restricted likelihood estimation in the Kalman Filter. *Bayesian Analysis of Time Series and Dynamic Models* (Ed. James C. Spall), New York: Marcel Dekker Inc., 477-508.

Singh, A.C., and Verret, F. (2006a). Statistics Canada's Small area Estimation Product and Its Pilot Test with the Labour Force Survey data. *Report prepared under contract for FLMM_LMI Working Group,* Canada, 36 pages.

Singh, A.C. and Verret, F. (2006b). Linear Nonlinear Mixed Models for Small Area Estimation from Survey Data. *Proceedings of the Statistics Canada Symposium*, to appear.

Singh, A.C., Mantel, H.J., and Thomas, B.W. (1994). Time series EBLUPs for small areas using survey data. *Survey Methodology*, 20, 33-43.

Singh, A.C., Stukel, D.M., and Pfeffermann, D. (1998). Bayesian versus Frequentist measures of error in small area estimation. *Jour. Roy. Statist. Soc*., Ser B., 377-396.

Singh, A.C., Folsom, R.E., Jr., and Vaish, A.K. (2003). Hierarchical Bayes Small Area Estimation for Survey Data by EFGL : the Method of Estimating Function-Based Gaussian Likelihood, *Federal Committee on Statistical Methods Statistical Policy Working paper* #36, pp. 47-74 www.fcsm.gov/reports.

Singh, A.C., Folsom, R.E., Jr., and Vaish, A.K. (2005). Small area modeling for survey data with smoothed error covariance structure via generalized design effects. *Federal Committee on Statistical Methods Conference Proceedings*, Washington, D.C. www.fcsm.gov/reports.