Small Area Estimation for Business Surveys

W. Yung[1], S. Rubin-Bleuer[2] and S. Landry[3]
[1]Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A-0T6
[2]Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A-0T6
[3]Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, Canada, K1A-0T6

Abstract

Small area estimation techniques are becoming increasingly used in survey applications to provide estimates for local areas of interest. These techniques combine direct survey estimates with auxiliary data to produce estimates for areas with very few sampled units. Most of the applications have been in social surveys where the areas of interest are geographical regions, with fewer applications to business surveys, although tax information is usually available as auxiliary data. Statistics Canada has been investigating small area estimation for its Survey of Employment, Payrolls and Hours (SEPH). Preliminary results have been encouraging but some issues with negative variance component estimates have been noted. In this presentation, alternative methods to estimate the variance components will be investigated and evaluated using a population generated from SEPH data.

Key Words: Small area estimation; Business surveys; Variance components

1. Introduction

Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH) is a monthly survey designed to produce estimates of levels and month-to-month trends of payrolls, employment, paid hours and earnings. The target population is composed of all employees in Canada except for those in a few select industries (ex. agriculture, fishing and trapping, etc.). The program makes extensive use of administrative data with the aid of a monthly survey. The administrative source is the Canada Revenue Agency's (CRA) Payroll Deduction Accounts (PD7) file, which includes the number of employees and the gross monthly payroll for the approximately one million employers in Canada. The administrative data is combined with data from the monthly Business Payroll Survey (BPS) through the use of the Generalized Regression (GREG) estimator. Taking advantage of the administrative data has allowed SEPH to produce quality estimates at a moderately detailed industry by province level. However, data users are asking for levels of detail for which the SEPH sample is unable to support estimates of reliable quality. To address these demands, SEPH is investigating the use of small area estimation techniques which would produce quality estimates at domains where there are very few sampled units.

In the following section we present some background on SEPH and we describe the current estimation strategy. In section 3 we describe small area estimation techniques and estimation of variance components we have investigated for SEPH. In section 4 we present some results from a limited simulation study. We discuss the results in section 5.

2. The Survey of Employment, Payrolls and Hours

In the past, SEPH was a large business survey with a monthly sample of approximately 60,000 businesses. In the mid 1990's, SEPH was redesigned to take advantage of the availability of administrative data (PD7) from CRA. By combining PD7 data with monthly survey data, the sample size was reduced to the current 15,000 units. The BPS uses a stratified random sample design with a 1/12 rotation each month. The variables collected by the survey include employment (E), gross monthly payroll (P) and summarized weekly earnings (SWE). From these variables, two additional ones are derived:

Average Weekly Earnings (AWE) = SWE/E and

Average Monthly Earnings (AME) = P/E.

Before defining the sample GREG estimator in the SEPH context, it will be useful to introduce the following notation. For the k^{th} establishment in any given month, let

- E_k be the number of employees from the PD7 file (E_{1k} is the number of employees from the BPS)
- P_k is the gross monthly payroll from the PD7 file (P_{1k} is the payroll from the BPS)
- SWE_k is the summarized weekly earnings (available from the BPS only)
- $AWE_k = SWE_k / E_{1k}$ is the average weekly earnings (available from the BPS only) and
- $AME_{1k} = P_{1k} / E_{1k}$ is the average monthly earnings based on the BPS and
- $AME_k = P_k / E_k$ is the average monthly earnings based on the PD7 file.

Now suppose the variable of interest is the *AWE*. It can be derived for respondents to the BPS and an estimate can be obtained using a Horvitz-Thompson (H-T) estimator. The H-T estimator can be improved upon by recognizing that *AWE* is correlated to the *AME* on the PD7 file and using a GREG estimator given by

$$\hat{Y}_G = \sum_s \tilde{w}_k g_k y_k$$

where $\tilde{w}_k = w_k E_k / \sum_U E_k$, w_k is the design weight obtained from the survey design, y_k is the *AWE* of the k^{th} unit in the BPS, $g_k = 1 + \boldsymbol{x}_k^T \hat{\boldsymbol{T}}^{-1} (\boldsymbol{X}^* - \hat{\boldsymbol{X}}^*)$, $\hat{\boldsymbol{T}} = \sum_s w_k E_k \boldsymbol{x}_k \boldsymbol{x}_k^T$, $\boldsymbol{x}_k^T = (1, \mathrm{AME}_k)$, $\boldsymbol{X}^* = \left(\sum_U E_k, \sum_U E_k AME_k \right)$ and $\hat{\boldsymbol{X}}^* = \sum_s w_k E_k \boldsymbol{x}_k$. Note that the estimate of *AWE* is a weighted sum of each unit's *AWE*, with the weight equal to employment. Also, note that $\hat{Y}_G$ can be expressed as

$$\hat{Y}_G = \frac{1}{E} \left(\sum_U E_k \hat{y}_k + \sum_s w_k E_k (y_k - \hat{y}_k) \right) \tag{1}$$

where $E = \sum_U E_k$, $\quad \hat{\boldsymbol{B}} = \hat{\boldsymbol{T}}^{-1} \sum_s w_k E_k \boldsymbol{x}_k^T y_k$ and $\hat{y}_k = \hat{\boldsymbol{B}}^T x_k$.

When implementing the GREG estimator, the population is commonly divided into mutually exclusive and exhaustive groups called model groups. The GREG estimator is implemented within each model group and then simply summed over the model groups to obtain estimates of population parameters. It is well known that the GREG estimator is approximately unbiased for population parameters and even at domains lower than the model group level. However, it does need to have a large enough sample to reliably estimate the adjustment $\sum_s w_k E_k (y_k - \hat{y}_k)$ in equation (1).

If only a small number of observations is available at the domain level, the sample GREG estimator remains approximately unbiased but becomes inefficient in terms of sampling variability. The potential instability of the sample GREG estimator was observed in previous studies. An obvious way to improve the stability is to increase the sample size to ensure that all domains of interest have a sufficient sample. However, given the budgetary constraints that Statistics Canada is facing, it is unlikely that such an increase would be approved. In an attempt to improve the precision of the GREG estimator in SEPH, small area estimation techniques for domains where there are very few observations have been investigated. In the next section we present some work on using small area estimation methods in the SEPH context.

3. Small Area Estimation

3.1 Best Linear Unbiased Predictor (BLUP) Estimator

Given the rotation in the SEPH design, the small area model will borrow strength across time and across areas or domains. We adopt the following notation: i represents the domain of interest, t the time period of interest and k the establishment. We denote by m the number of domains within a model group, each containing N_i establishments, $i=1,\ldots,m$. The AWE at time t for a domain i is given by the weighted average $\bar{Y}_{it} = \sum_{k=1}^{N_i} E_{ikt} AWE_{ikt} \Big/ \sum_{k=1}^{N_i} E_{ikt}$ where E_{ikt} is the number of employees at time t and AWE_{ikt} is the average weekly earnings in establishment k of domain i, $k = 1,\ldots,N_i$, $i = 1,\ldots,m$. Let $\bar{X}_{it} = AME_{it} = \sum_{k=1}^{N_i} E_{ikt} AME_{ikt} \Big/ \sum_{k=1}^{N_i} E_{ikt}$ be the average monthly earnings in domain i, at time t, $t=1,2$. Let $\bar{y}_{it}$ be the GREG estimator of average weekly earnings at month t for domain i. We define the vector of covariates as $\bar{\mathbf{X}}_{it} = (1, \bar{X}_{it})'$.

The small area model considered for SEPH is the Rao-Yu (Rao 2003, 158-160) model that combines a time series component with a cross-sectional one,

$$
\begin{aligned}
\bar{y}_{it} &= \theta_{it} + e_{it} \\
\theta_{it} &= \bar{\boldsymbol{X}}_{it}^T \boldsymbol{\beta} + v_i + u_{it}
\end{aligned}
\tag{2}
$$

where θ_{it} is the small area mean at time t ($t=1, \ldots, \tau$) for small area i ($i=1, \ldots, m$), e_{it} is the sampling error corresponding to the direct estimates $\bar{y}_{it}$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$,

$v_i \sim N(0,\sigma_v^2)$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim N(0,\sigma^2)$, $|\rho|\le 1$, and $\Psi_i = \mathrm{Var}(\boldsymbol{e}_i)$ (assumed known). Note that the area random effects, v_i, model errors, ε_{it}, and sampling errors, e_{it}, are all independent of each other. In the context of SEPH, we assume a random walk model where $\rho=1$ and only two time points (i.e. $\tau=2$).

Note that model (2) is a special case of the general linear mixed model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{3}$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of sample observations, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known matrices of full rank and $\boldsymbol{v}$ and $\boldsymbol{e}$ are independently distributed with means $\boldsymbol{0}$ and covariance matrices $\boldsymbol{G}$ and $\boldsymbol{R}$ depending on some variance parameters $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_q)^T$ (see Rao, 2003). We denote the variance covariance matrix of $\boldsymbol{y}$ as

$$\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{V} = \boldsymbol{V}(\boldsymbol{\delta}) = \boldsymbol{R} + \boldsymbol{ZGZ}^T.$$

The small area parameters of interest are expressed as $\theta = \ell^T\boldsymbol{\beta} + \boldsymbol{m}^T\boldsymbol{v}$ for specified vectors of known constants ℓ and $\boldsymbol{m}$. For a known $\boldsymbol{\delta}$, the BLUP estimator of θ is given by

$$\tilde{\theta} = t(\boldsymbol{\delta},\boldsymbol{y}) = \ell^T\tilde{\boldsymbol{\beta}} + \boldsymbol{m}^T\boldsymbol{GZ}^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \tag{4}$$

where $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$.

Model (2) can be expressed in the general linear mixed model form as follows. Let,

$$\boldsymbol{y} = \underset{1\le i\le m}{\mathrm{col}}(\boldsymbol{y}_i) = (\boldsymbol{y}_1^T,\ldots,\boldsymbol{y}_m^T)^T, \quad \boldsymbol{X} = \underset{1\le i\le m}{\mathrm{col}}(\boldsymbol{X}_1^T,\ldots,\boldsymbol{X}_m^T)^T$$
$$\boldsymbol{Z} = \underset{1\le i\le m}{\mathrm{diag}}(\boldsymbol{Z}_i), \boldsymbol{v} = \underset{1\le i\le m}{\mathrm{col}}(\boldsymbol{v}_i), \quad \boldsymbol{e} = \underset{1\le i\le m}{\mathrm{col}}(\boldsymbol{e}_i)$$

where $\boldsymbol{y}_i = (\bar{y}_{it},\bar{y}_{it-1})^T$, $\overline{\boldsymbol{X}}_{it} = (1,\overline{X}_{it})^T$, $\boldsymbol{v}_i = (v_i,u_{it},u_{it-1})^T$, $\boldsymbol{e}_i = (e_{it},e_{it-1})^T$, $\boldsymbol{X}_i = \begin{pmatrix}\overline{\boldsymbol{X}}_{it}^T \\ \overline{\boldsymbol{X}}_{it-1}^T\end{pmatrix}$ and $\boldsymbol{Z}_i = (\boldsymbol{1}_2, \boldsymbol{I}_2)$, $\boldsymbol{1}_2$ is a 2x1 vector of 1's and $\boldsymbol{I}_2$ is the identity matrix of order 2. Further, $\boldsymbol{R} = \underset{1\le i\le m}{\mathrm{diag}}(\boldsymbol{R}_i)$ and $\boldsymbol{G} = \underset{1\le i\le m}{\mathrm{diag}}(\boldsymbol{G}_i)$ so that $Var(\boldsymbol{y})=\boldsymbol{V}$ has a block diagonal structure $\boldsymbol{V} = \underset{1\le i\le m}{\mathrm{diag}}(\boldsymbol{V}_i)$ where $\boldsymbol{V}_i = \boldsymbol{R}_i + \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i^T$ with

$$\boldsymbol{R}_i = \mathrm{Var}(\boldsymbol{e}_i) = \boldsymbol{\Psi}_i$$
$$\boldsymbol{G}_i = \mathrm{Var}(\boldsymbol{v}_i) = \begin{pmatrix}\sigma_v^2 & \boldsymbol{0}^T \\ \boldsymbol{0} & \sigma^2\boldsymbol{\Lambda}\end{pmatrix}$$

and $\boldsymbol{\Lambda}$ is the 2x2 covariance matrix of $\boldsymbol{u}_i = (u_{it},u_{it-1})$. Assuming a random walk model, the $(t,s)^{\text{th}}$ element of $\boldsymbol{\Lambda}$ is given by $\Lambda_{t,s} = Cov(u_{it},u_{is}) = \min(t,s)$ for t, s=1, 2. That is,

$$\boldsymbol{\Lambda} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Given the block diagonal covariance structure, the general liner mixed model (3) can be decomposed into *m* sub-models,

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{v}_i + \boldsymbol{e}_i$$

and the BLUP estimator, equation (4), reduces to

$$\tilde{\boldsymbol{\theta}}_i = \boldsymbol{t}_i(\boldsymbol{\delta}_i \boldsymbol{y}_i) = \ell_i^T \tilde{\boldsymbol{\beta}} + \boldsymbol{m}_i^T \tilde{\boldsymbol{v}}_i$$

where $\tilde{\boldsymbol{v}}_i = \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}})$. If we are interested in the small area parameter for the current occasion, *t*, then $\ell_i = \overline{\boldsymbol{X}}_{it}$ and $\boldsymbol{m}_i^T = (1, 1, 0)$ and we have

$$\tilde{\theta}_{it} = \overline{\boldsymbol{X}}_{it}^T \tilde{\boldsymbol{\beta}} + (\sigma_v^2 + 2\sigma^2, \sigma_v^2 + \sigma^2) \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}}) \tag{5}$$

with $\tilde{\boldsymbol{\beta}} = \left(\sum_i \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right)^{-1} \sum_i \boldsymbol{X}_i \boldsymbol{V}_i^{-1} \boldsymbol{y}_i$. If we let $\boldsymbol{V}_i^{-1} = \begin{pmatrix} V_i^{11} & V_i^{12} \\ V_i^{21} & V_i^{22} \end{pmatrix}$, equation (5) can be expressed as

$$\begin{aligned}
\tilde{\theta}_{it} &= \overline{X}_{it}^T \tilde{\boldsymbol{\beta}} + \\
&\quad \left[(\sigma_v^2 + 2\sigma^2) V_i^{11} + (\sigma_v^2 + \sigma^2) V_i^{22}, (\sigma_v^2 + 2\sigma^2) V_i^{12} + (\sigma_v^2 + \sigma^2) V_i^{22} \right] (\mathbf{y}_i - \mathbf{X}_i \tilde{\boldsymbol{\beta}}) \\
&= \overline{X}_{it}^T \tilde{\boldsymbol{\beta}} + \gamma_{iT} (\bar{y}_{it} - \overline{X}_{it}^T \tilde{\boldsymbol{\beta}}) + \gamma_{iT-1} (\bar{y}_{it-1} - \overline{\mathbf{X}}_{it-1}^T \tilde{\boldsymbol{\beta}}) \\
&= \gamma_{it} \bar{y}_{it} + (1 - \gamma_{it}) \overline{\mathbf{X}}_{it}^T \tilde{\boldsymbol{\beta}} + \gamma_{it-1} (\bar{y}_{it-1} - \overline{\mathbf{X}}_{it-1}^T \tilde{\boldsymbol{\beta}})
\end{aligned}$$

where $\gamma_{it} = (\sigma_v^2 + 2\sigma^2) V_i^{11} + (\sigma_v^2 + \sigma^2) V_i^{22}$ and $\gamma_{it-1} = (\sigma_v^2 + 2\sigma^2) V_i^{12} + (\sigma_v^2 + \sigma^2) V_i^{22}$. Note that the BLUP estimator (5) depends on the vector of variance components $\boldsymbol{\delta} = (\sigma_v^2, \sigma^2)$.

The Empirical Best Linear Unbiased Predictor, or EBLUP, is obtained by estimating $\boldsymbol{\delta}$ with $\hat{\boldsymbol{\delta}} = (\hat{\sigma}_v^2, \hat{\sigma}^2)$ and substituting it into the BLUP estimator, *i.e.*, $\hat{\boldsymbol{\theta}} = \boldsymbol{t}(\hat{\boldsymbol{\delta}}, \boldsymbol{y})$. Thus for SEPH, the EBLUP of θ_{it} is

$$\hat{\theta}_{it} = X_{it}^T \hat{\beta} + (\hat{\sigma}_v^2 + 2\hat{\sigma}^2, \hat{\sigma}_v^2 + \hat{\sigma}^2) \hat{\boldsymbol{V}}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_{it} \hat{\boldsymbol{\beta}})$$

where $\hat{\boldsymbol{\beta}} = \left(\sum_i \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_i \boldsymbol{X}_i \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{y}_i \right)$ and $\hat{\boldsymbol{V}}_i = \boldsymbol{\Psi}_i + \hat{\sigma}_v^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \hat{\sigma}^2 \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$

3.2 Estimation of Variance Components

In order to use the EBLUP estimator, estimates of the variance components, $\boldsymbol{\delta}=(\sigma_v^2,\sigma^2)$ are required. Yung, Rubin-Bleuer and Landry (2009) derived method of moments estimators for σ_v^2 and σ^2 for the Rao-Yu model. A common problem with estimating the variance components is that some estimates can be negative. The method of moments variance estimators are generally slower in convergence than other methods and consequently, they often yield negative estimates. When negative estimates were obtained Yung, Rubin-Bleuer and Landry (2009) replaced them with an estimate of the variance of the respective variance component as suggested by Wang and Fuller (2003). Details on estimating the variance of σ_v^2 and σ^2 under a method of moments approach are given in Rubin-Bleuer (2009). EBLUP estimators for the Rao-Yu model were calculated and showed some gains over the direct GREG SEPH estimator. Given these gains, despite the large number of negative estimates, it was felt that further improvements could be achieved with 'better' estimates of the variance components. Thus two alternative approaches were investigated: the Restricted Maximum Likelihood (REML) and the Adjusted Maximum Likelihood (AML) approaches.

3.2.1 Restricted Maximum Likelihood (REML)

In order to estimate the variance components σ_v^2 and σ^2 using REML, we express the model as a combination of fixed effects and pure random effects by setting $u_{i2}=u_{i1}+\varepsilon_{i2}$, $u_{i1}=\varepsilon_{i1}$ and $\varepsilon_{it}\sim(0,\sigma^2)$, $t=1,2$. Hence

$$\mathbf{y}_i=\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i^*\mathbf{b}_i+\mathbf{e}_i, \tag{6}$$

where

$$\mathbf{v}_i=\begin{pmatrix} v_i \\ u_{i2} \\ u_{i1}\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix} v_i \\ \varepsilon_{i2} \\ \varepsilon_{i1}\end{pmatrix},$$

$\mathbf{b}_i=(v_i,\varepsilon_{i2},\varepsilon_{i1})^T$ and

$$\mathbf{Z}_i^*=\mathbf{Z}_i\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}=\begin{pmatrix}1&1&1\\1&0&1\end{pmatrix}.$$

Thus the variance of $\boldsymbol{y}_i$, $i=1,\ldots m$, is

$$\boldsymbol{V}_i=Var(\boldsymbol{y}_i)=\boldsymbol{Z}_i^*Diag(\sigma_v^2,\sigma^2,\sigma^2)\boldsymbol{Z}_i^{*T}+\boldsymbol{\Psi}_i.$$

The variance of $\mathbf{y}=(\mathbf{y}_1^T,\ldots,\mathbf{y}_m^T)^T$ is then $\boldsymbol{V}=BD(\boldsymbol{V}_1,\ldots,\boldsymbol{V}_m)$ where BD stands for Block Diagonal matrix.

A shortcoming of maximum likelihood estimation is that in estimating the variance components it does not take into account the degrees of freedom that are involved in the estimation of the fixed effects. Restricted Maximum Likelihood Estimation (REML) estimates the variance components based on residuals after fitting the ordinary least squares to just the fixed effects part of the model. Under the assumption of normality, the restricted log likelihood function is

$$L_R(\boldsymbol{\sigma}, \boldsymbol{y}) = -\frac{1}{2}\log(\det \boldsymbol{V}) - \frac{1}{2}\log(\det \boldsymbol{A}^T \boldsymbol{V}\boldsymbol{A}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

where $\boldsymbol{\sigma} = (\sigma_v^2, \sigma^2)$, A is any $2m \times (2m-p)$ matrix of full rank orthogonal to the $2m \times p$ matrix X. The partial derivatives of the restricted likelihood function are:

$$s_v(\underset{\sim}{\boldsymbol{\sigma}}, \mathbf{y}) = \partial L_R(\underset{\sim}{\boldsymbol{\sigma}}, \mathbf{y}) / \partial \sigma_v^2 = -\frac{1}{2} tr\left[\mathbf{P}\mathbf{V}_v\right] + \frac{1}{2}\mathbf{y}^T \mathbf{P}\mathbf{V}_v\mathbf{P}\mathbf{y},$$

and

$$s_\varepsilon(\underset{\sim}{\boldsymbol{\sigma}}, \mathbf{y}) = \partial L_R(\underset{\sim}{\boldsymbol{\sigma}}, \mathbf{y}) / \partial \sigma^2 = -\frac{1}{2} tr\left[\mathbf{P}\mathbf{V}_\varepsilon\right] + \frac{1}{2}\mathbf{y}^T \mathbf{P}\mathbf{V}_\varepsilon\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$,

$$\boldsymbol{V}_v = \partial \boldsymbol{V} / \partial \sigma_v^2 = BD\left\{\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right\}$$

and

$$\mathbf{V}_\varepsilon = \partial \mathbf{V} / \partial \sigma^2 = BD\left\{\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\right\}.$$

The REML estimator of $\boldsymbol{\sigma}$ is obtained iteratively using the "scoring" algorithm:

$$\boldsymbol{\sigma}^{(a+1)} = \boldsymbol{\sigma}^{(a)} + \boldsymbol{I}^{-1}(\boldsymbol{\sigma}^{(a)})\begin{pmatrix} s_v(\boldsymbol{\sigma}^{(a)}, \boldsymbol{y}) \\ s_\varepsilon(\boldsymbol{\sigma}^{(a)}, \boldsymbol{y}) \end{pmatrix}, \tag{7}$$

where $\boldsymbol{\sigma}^{(a)}$ is the estimate of $\boldsymbol{\sigma}$ at the a^{th} iteration and $\boldsymbol{I}(\boldsymbol{\sigma}^{(a)})$ is the matrix of expected second derivatives of $-L_R(\boldsymbol{\sigma}, \boldsymbol{y})$ given by

$$\mathbf{I}(\boldsymbol{\sigma}) = \begin{pmatrix} I_{vv} & I_{v\varepsilon} \\ I_{\varepsilon v} & I_{\varepsilon\varepsilon} \end{pmatrix},$$

where

$$I_{v\varepsilon} = \frac{1}{2} tr\left[\mathbf{P}\mathbf{V}_v\mathbf{P}\mathbf{V}_\varepsilon\right],$$

$I_{vv} = \frac{1}{2} tr\left[\mathbf{P}\mathbf{V}_v\mathbf{P}\mathbf{V}_v\right]$ and $I_{\varepsilon\varepsilon} = \frac{1}{2} tr\left[\mathbf{P}\mathbf{V}_\varepsilon\mathbf{P}\mathbf{V}_\varepsilon\right]$.

3.2.2 Adjusted Maximum Likelihood (AML)

Although the REML estimator of $\boldsymbol{\sigma}$ converges much faster than the estimator by the method of moments, it is also subject to negative variance estimates when the number of areas is not sufficiently large. Li and Lahiri (2010) proposed an Adjusted Maximum Likelihood (AML) method for the Fay-Herriot model, which produces only positive estimates of the variance components. The proposed method maximizes an adjusted likelihood defined as the product of the variance component to be estimated and a standard likelihood.

We now present the AML estimator for the Rao-Yu model under consideration. Using model (6) we have, under normality, the adjusted log likelihood function for σ_v^2 and σ^2

$$L_A(\boldsymbol{\sigma},\mathbf{y}) = -\frac{1}{2}\log(\det\mathbf{V}) - \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\,\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \log\sigma_v^2 + \log\sigma^2.$$

To solve for σ_v^2 and σ^2, we have

$$\begin{aligned} s_v(\boldsymbol{\sigma},\mathbf{y}) &= \partial L_A(\boldsymbol{\sigma},\mathbf{y})/\partial\sigma_v^2 \\ &= -\frac{1}{2}tr[\mathbf{V}^{-1}\mathbf{V}_v] + \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\,\mathbf{V}^{-1}\mathbf{V}_v\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{1}{\sigma_v^2}, \end{aligned}$$

$$\begin{aligned} s_\varepsilon(\boldsymbol{\sigma},\mathbf{y}) &= \partial L_A(\boldsymbol{\sigma},\mathbf{y})/\partial\sigma^2 \\ &= -\frac{1}{2}tr[\mathbf{V}^{-1}\mathbf{V}_\varepsilon] + \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\,\mathbf{V}^{-1}\mathbf{V}_\varepsilon\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \frac{1}{\sigma^2}, \end{aligned}$$

where $\mathbf{V}_v$ and $\mathbf{V}_\varepsilon$ are defined in Section 3.2.1.

In order to apply the scoring algorithm (7), one needs the matrix of expected second derivatives of $-L_A(\boldsymbol{\sigma},\boldsymbol{y})$. Algebraic manipulation yields

$$-\partial s_v(\boldsymbol{\sigma},\boldsymbol{y})/\partial\sigma_v^2 = \frac{1}{2}tr[\boldsymbol{V}^{-1}\boldsymbol{V}_v\boldsymbol{V}^{-1}\boldsymbol{V}_v] - (\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}\boldsymbol{V}_v\boldsymbol{V}^{-1}\boldsymbol{V}_v\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) - \left(\frac{1}{\sigma_v^2}\right)^2,$$

and

$$-\partial s_\varepsilon(\boldsymbol{\sigma},\boldsymbol{y})/\partial\sigma^2 = \frac{1}{2}tr[\boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon\boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon] - (\boldsymbol{y}-\boldsymbol{X\beta})^T\boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon\boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}) - \left(\frac{1}{\sigma^2}\right)^2.$$

Taking expectations gives

$$\begin{aligned} I_{vv} &= E\left(-\partial s_v(\boldsymbol{\sigma},\boldsymbol{y})/\partial\sigma_v^2\right) \\ &= \frac{1}{2}tr[\boldsymbol{V}^{-1}\boldsymbol{V}_v\boldsymbol{V}^{-1}\boldsymbol{V}_v] + \left(\frac{1}{\sigma_v^2}\right)^2, \end{aligned}$$

$$\begin{aligned} I_{v\varepsilon} = I_{\varepsilon v} &= E\left(-\partial s_v(\boldsymbol{\sigma}, \boldsymbol{y}) / \partial \sigma_v^2\right) \\ &= \frac{1}{2} tr[\boldsymbol{V}^{-1}\boldsymbol{V}_v \boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon] \end{aligned}$$

and

$$\begin{aligned} I_{\varepsilon\varepsilon} &= E\left(-\partial s_\varepsilon(\boldsymbol{\sigma}, \boldsymbol{y}) / \partial \sigma^2\right) \\ &= \frac{1}{2} tr[\boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon \boldsymbol{V}^{-1}\boldsymbol{V}_\varepsilon] + \left(\frac{1}{\sigma^2}\right)^2. \end{aligned}$$

The AML estimate of $\boldsymbol{\sigma}$ is then obtained iteratively as the solution to

$$\boldsymbol{\sigma}^{(a+1)} = \boldsymbol{\sigma}^{(a)} + \boldsymbol{I}^{-1}(\boldsymbol{\sigma}^{(a)}) \begin{pmatrix} s_v(\boldsymbol{\sigma}^{(a)}, \boldsymbol{y}) \\ s_\varepsilon(\boldsymbol{\sigma}^{(a)}, \boldsymbol{y}) \end{pmatrix}.$$

Rubin-Bleuer, Yung and Landry (2010) have shown that the AML estimates are positive and are consistent for $\boldsymbol{\sigma}$ as the number of small areas goes to infinity.

4. Evaluation of Variance Components Estimators

In Section 3 we presented three methods of estimating the variance components: the method of moments with the Wang-Fuller approach for negative variance estimates, the REML approach and the proposed AML approach. Here we compare the design-based properties of the corresponding small area estimators and the direct GREG estimator.

A simulated population was created using on actual SEPH data. The population contained approximately 1,000,000 units and had values for AME, AWE and the number of employees for each unit. Using the BPS sample design, samples were drawn from this simulated population and GREG estimators were computed for detailed industry groups within each province and territory. The model groups for the GREG estimators were defined at a level higher than the detailed industry level as some detailed industry domains had very few sampled units. GREG estimates $\bar{y}_{it}$, $i = 1,...,m$, $t = 1, 2$, calculated at small area levels were then used as direct estimates to fit small area model (2).

The EBLUP estimator depends on estimates for ψ_i, σ_v^2 and σ^2. In practice, estimates of the variance of the sampling errors ψ_i, are typically smoothed by fitting regression models to the design based estimates of the variance of the sampling errors. Since the population is available in our empirical study, ψ_i was calculated by drawing 1,000 independent samples and then averaging over the samples. The variance components σ_v^2 and σ^2 were estimated using the three methods described above.

EBLUP estimates were obtained for sub-industries within four major industries for all provinces and territories. The number of sub-industries varied with the industry and the geographic region (province or territory). We present the results for two major industries: Construction and Real Estate and Rental and Leasing. In the Construction industry the sample sizes of the 26 ‘small areas’ varied from 1 to 77 observations, while they varied from 1 to 22 for the Real Estate and Rental and Leasing industry across the 28 small areas.

Design-based relative biases (RB) and "coefficients of variation" (CV) were calculated for the GREG estimator and each of the EBLUP estimators based on the three different methods of estimating the variance components: 1,000 independent samples were drawn from the simulated SEPH population and the following formulas were used,

$$RB(\hat{\theta}_{it}) = \frac{1}{1000} \sum_{b=1}^{1000} \left(\hat{\theta}_{it}^{(b)} - \theta_{it}\right) / \theta_{it}$$

and

$$CV(\hat{\theta}_{it}) = \sqrt{\frac{1}{1000} \sum_{b=1}^{1000} \left(\hat{\theta}_{it}^{(b)} - \theta_{it}\right)^2} \Big/ \theta_{it}$$

where $\hat{\theta}_{it}^{(b)}$ is the estimator of the small area AWE for the i^{th} area at time t and for the b^{th} sample and θ_{it} is the true small area AWE at time t for area i, $i=1, \ldots m$.

The average relative biases of the four estimators for the small areas are given in figures 1 and 2. Along the horizontal axis appears a four digit code followed by the sample size of the small area. The four digit code consists of a two digit province code and a two digit code indicating the detailed sub-industry. In terms of relative bias, there does not appear to be any discernible patterns. The relative biases of the EBLUP estimators tend to be similar to that of the GREG estimator when sample sizes are small (less than 15 to 20) and are typically larger than that of the GREG estimator when sample sizes tend to be larger. This is not unexpected as the GREG estimator is asymptotically unbiased for the small area totals as the sample sizes become large. Comparing the EBLUB estimators, one sees that there is very little difference between the three methods used to calculate $\boldsymbol{\sigma}$. The REML and the AML have very similar relative biases and the method of moments estimator has a slightly smaller relative bias.

The CVs for the four estimators are given in figures 3 and 4. From the figures, one can see that the three EBLUP estimators tend to have smaller CVs than the GREG estimator, in particular when the sample sizes are small. Despite the additional biases in the EBLUP estimators when the sample sizes are large, enough strength is obtained from the model so that the CVs are smaller. When comparing the EBLUP estimators, we do not see many differences indicating that they are quite robust to negative estimates of the variance components. The REML and AML based estimators behave very similarly, while the method of moments based estimator tends to have a slightly higher CV. We do point out that in one of the industries studied, the AML did perform significantly better than the REML and the method of moments based estimators.

5. Conclusion

The GREG estimator currently in use by SEPH has well known large sample properties. However, with more and more demand for estimates at very detailed industry levels, sample sizes are not always large enough to support the GREG estimator. Small area estimation techniques have shown some gains in efficiency over the GREG estimator in domains with small sample sizes. One problem that small area methods have, are negative estimators of the variance components. Three methods of estimating the variance components were compared using a simulated population based on the SEPH survey. The comparison showed that although many negative estimates of the variance components were obtained for the method of moments estimator, there was little effect in terms of relative bias and CV.

REFERENCES

Li, H. and Lahiri, P. (2010), An Adjusted Maximum Likelihood Method for Solving Small Area Estimation Problems, *Journal of Multivariate Analysis*, **101**, 882-892.

Rao, J.N.K. (2003), *Small Area Estimation*, New Jersey: John Wiley and Sons, Inc.

Rubin-Bleuer, S. (2009). Estimation of the Variance Components in Time Series and Cross-Sectional Data Using the Method of Moments. Internal Statistics Canada document.

Rubin-Bleuer, S., Yung, W. and Landry, S. (2010), Adjusted Maximum Likelihood Method for A Small Area Model Accounting for Time and Area Effects, Internal Statistics Canada document.

Wang, J. and Fuller, W.A. (2003). The Mean Square Error of Small Area Predictors Constructed with Estimated Area Variances, *Journal of the American Statistical Association*, **98**, 716-723.

You, Y. (1999), Hierarchical Bayes and Related Methods for Model-Based Small Area Estimation, Unpublished Ph.D. Thesis, Carleton University, Ottawa, Canada.

Yung, W., Rubin-Bleuer, S. and Landry, S. (2009). Small Area Estimation: A Business Survey Application, Proceedings of the Survey Research Methods Section, Statistical Society of Canada.

Figure 1. Relative Bias- Construction Industry

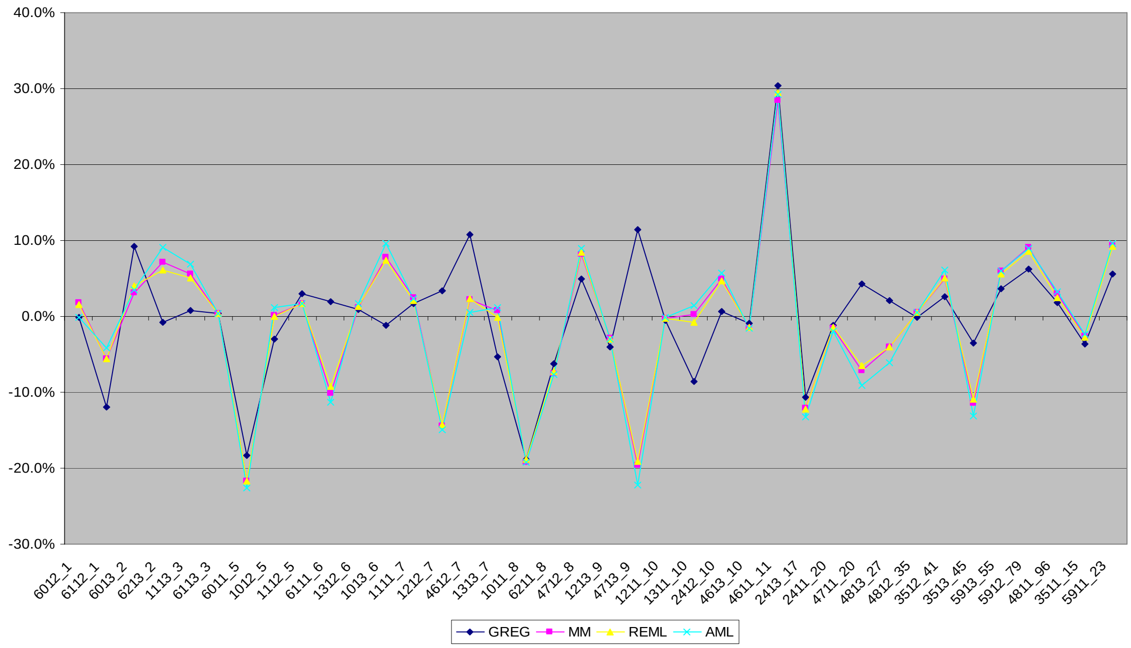

Figure 2. Relative Bias- Real Estate and Rentals and Leasing Industry

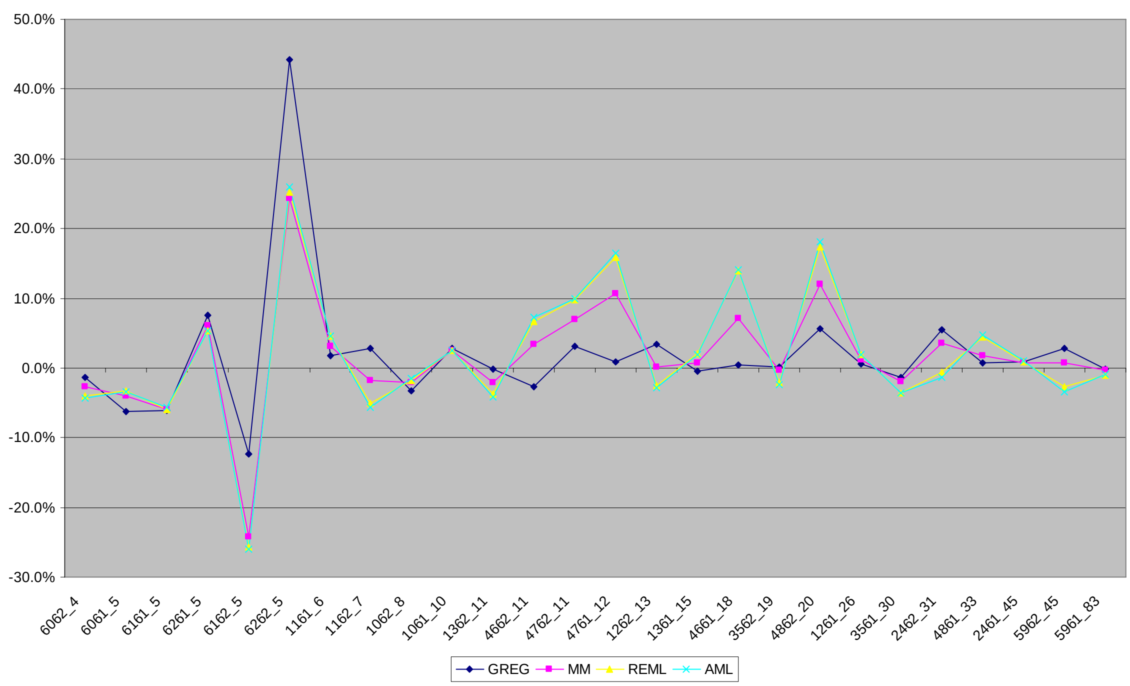

Figure 3. CV- Construction Industry

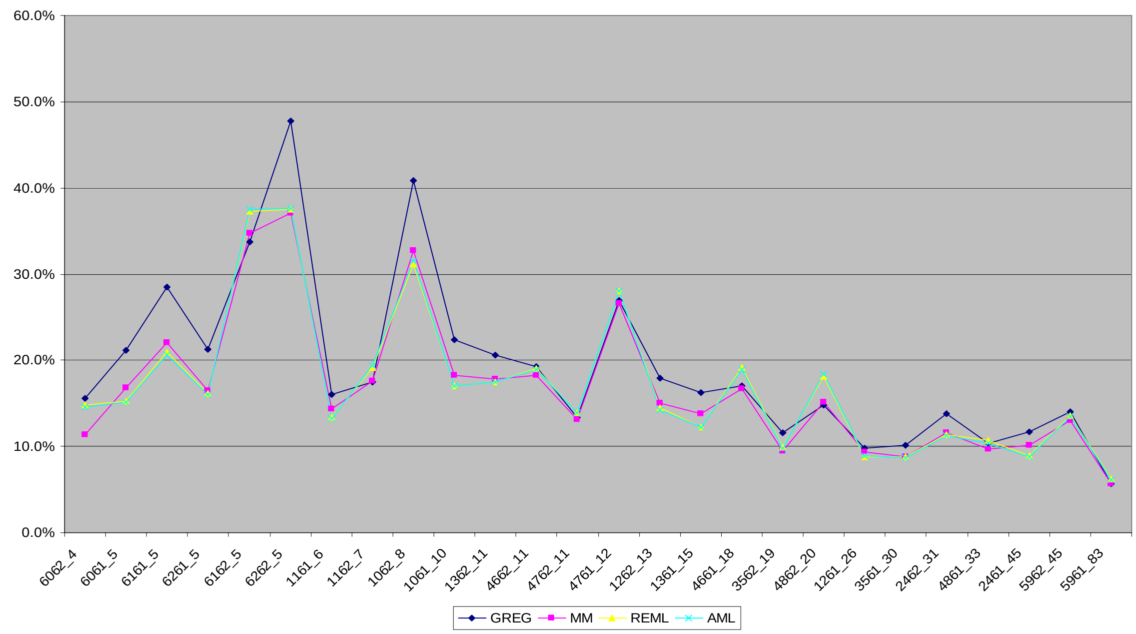

Figure 4. CV- Real Estate and Rentals and Leasing Industry