SSC Annual Meeting, June 2009
Proceedings of the Survey Methods Section

SMALL AREA ESTIMATION: A BUSINESS SURVEY APPLICATION

W. Yung, S. Rubin-Bleuer and S. Landry[1]

ABSTRACT

Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH) combines data from a monthly payroll survey with administrative data from the Canada Revenue Agency (CRA). Using this administrative data allows SEPH to produce estimates for domains with small sample sizes. SEPH is currently investigating regression composite estimation as a means to further improve the efficiency of its estimates. Despite these efforts, there are still domains of interest for which reliable estimates can not be produced. In this presentation we propose an estimator that combines small area estimation methods with the regression composite estimator and illustrate it using data from SEPH.

KEY WORDS: Cross sectional and time series model; EBLUP

RÉSUMÉ

L'Enquête sur l'emploi, la rémunération et les heures de travail (EERH) de Statistiques Canada combine des données issues d'une enquête mensuelle sur la rémunération et des données administratives de l'Agence du revenu du Canada (ARC). En utilisant ces données, l'EERH peut produire des estimations – relativement aux domaines – en prenant des échantillons de petite taille. L'EERH évalue présentement l'estimateur composite de régression afin de déterminer s'il permettrait d'augmenter la précision de ses estimations. En dépit de ces efforts, il reste des domaines d'intérêt pour lesquels on ne dispose d'aucun estimateur fiable. Cette présentation se penche sur un estimateur qui allie les méthodes d'estimation de petits domaines à celles de l'estimateur composite de régression. Les données de l'EERH serviront d'étalon pour l'évaluation de cet estimateur.

MOTS CLÉS : Modèle transversal et de séries chronologiques; MPLSBE

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 survey 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 1 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 some background on SEPH will be presented and the current estimation strategy will be described. Small area estimation techniques investigated for SEPH will be described in section 3. Section 4 will present some results from a limited simulation study. Some conclusions and future work will be discussed in section 5.

2. SEPH BACKGROUND

In the past, SEPH was a large business survey with a monthly sample size 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:

[1] W. Yung, S. Rubin-Bleuer and S. Landry, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6, Wesley.Yung@statcan.gc.ca.

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 $E = \sum_U E_k$, $\hat{\boldsymbol{B}} = \hat{\boldsymbol{T}}^{-1} \sum_s w_k E_k \boldsymbol{x}_k^T y_k$ 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 (1) 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 $\sum_s w_k E_{pk} (y_k - \hat{y}_k)$ at the domain level given in equation (1). If only a small number of observations are 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 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

2.1 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, $i=1,...,m$. 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, $k=1,...,N_i$, $i=1,...,m$. 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,

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

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 $\bar{y}_{it}$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, $v_i \sim N(0, \sigma_v^2)$, $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim N(0, \sigma^2)$, $|\rho|<1$, 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, ℓ and $\boldsymbol{m}$, of known constants. 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,

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

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

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

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

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

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 $\boldsymbol{\delta} = (\sigma_v^2, \sigma^2)$.

2.2 Estimation of σ^2

To estimate σ^2, subtract $\overline{y}_{it-1}$ from $\overline{y}_{it}$ in (2) and set

$$z_{it} = \overline{y}_{it} - \overline{y}_{it-1} = (\overline{\mathbf{X}}_{it}^T - \overline{\mathbf{X}}_{it-1}^T)\boldsymbol{\beta} + \varepsilon_{it} + \delta_{it}, i = 1, \ldots, m, \tag{6}$$

where we use the following notation: $\delta_{it} = e_{it} - e_{it-1}$. Note that δ_{it} and ε_{it} are independent of each other, by the conditions of model (2). The variance of δ_{it} is given by $\phi_{it} = Var(\delta_{it}) = (1,-1)\mathbf{\Psi}_i \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\mathbf{D}_\phi = diag(\phi_{1t},\ldots,\phi_{mt})$.

Let $P_{it} = (\bar{X}_{it} - \bar{X}_{it-1})$, $\mathbf{P} = (P_{1t},\ldots,P_{mt})^T$ and $\mathbf{z}_t = (z_{1t},\ldots,z_{mt})^T$, then (6) is a simple Fay-Herriot model given by

$z_{it} = P_{it}\beta_1 + \varepsilon_{it} + \delta_{it}$. Its vector analogue is $\boldsymbol{z} = \boldsymbol{P\beta_1} + \boldsymbol{\varepsilon} + \boldsymbol{\delta}$ with $\boldsymbol{\varepsilon} = (\varepsilon_{1t},\ldots,\varepsilon_{1m})^T \sim (0,\sigma^2\boldsymbol{I}_m)$ and $\boldsymbol{\delta} = (\delta_{1t},\ldots,\delta_{1m})^T \sim (0,\boldsymbol{D}_\phi)$. An unbiased and consistent estimator of the variance of ε_{it}, obtained by the method of moments, is

$$\hat{\sigma}^2 = \frac{\boldsymbol{z}^T\boldsymbol{M_P}\boldsymbol{z} - trace(\boldsymbol{D}_\phi\boldsymbol{M_P})}{m-1}, \tag{7}$$

where $\boldsymbol{M_P} = \boldsymbol{I_m} - \boldsymbol{P(P^T P)^{-1} P^T}$ is an idempotent matrix. Proofs of the consistency of $\hat{\sigma}^2$ are given in Rubin-Bleuer (2009).

2.3 Estimation of σ_v^2

To estimate σ_v^2, we fix the time period t and write equation (2) as

$$\bar{y}_{it} = \bar{\boldsymbol{X}}_{it}^T\boldsymbol{\beta} + v_i + w_{it},\quad w_{it} = u_{it} + e_{it}, \tag{8}$$

where $\bar{\boldsymbol{X}}_{it}$ is a 2×1 vector of covariates, $v_i \sim (0,\sigma_v^2)$, $w_{it} \sim (0,\ 2\sigma^2 + \Psi_i(t,t))$ with $\Psi_i(t,t) = Var(e_{it})$, for $i=1,\ldots,m$. Equation (8) represents a Fay-Herriot model of the form:

$$\bar{\boldsymbol{y}}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{w}_t,\quad \boldsymbol{X}_t = (\bar{\boldsymbol{X}}_{1t},\ldots,\bar{\boldsymbol{X}}_{mt})^T,\quad \bar{\boldsymbol{y}}_t = (\bar{y}_{1t},\ldots,\bar{y}_{mt})^T.$$

We now define $D_w = diag(2\sigma^2 + \Psi_1(t,t),\ldots,2\sigma^2 + \Psi_m(t,t))$ and $\boldsymbol{M}_{\boldsymbol{X}_t} = \boldsymbol{I}_m - \boldsymbol{X}_t(\boldsymbol{X}_t^T\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t^T$. Then, an unbiased and consistent estimator of the variance of the area effects σ_v^2 obtained by the method of moments is given by

$$\hat{\sigma}_v^2 = \frac{\bar{\boldsymbol{y}}_t^T\boldsymbol{M}_{\boldsymbol{X}_t}\bar{\boldsymbol{y}}_t - trace(\boldsymbol{D}_w\boldsymbol{M}_{\boldsymbol{X}_t})}{m-2}. \tag{9}$$

Proof of the consistency of $\hat{\sigma}_v^2$ is given in Rubin-Bleuer (2009).

2.4 Empirical Best Linear Unbiased Predictor (EBLUP) Estimator

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

$$\hat{\theta}_{it} = X_{it}^T\hat{\beta} + (\hat{\sigma}_v^2 + 2\hat{\sigma}^2, \hat{\sigma}_v^2 + \hat{\sigma}^2)\hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_{it}\hat{\boldsymbol{\beta}}) \tag{10}$$

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 $\hat{\boldsymbol{V}}_i = \boldsymbol{\Psi}_i + \hat{\sigma}_v^2\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \hat{\sigma}^2\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Now, suppose we wish to derive the MSE of $\hat{\boldsymbol{\theta}}$

$$\begin{aligned}\text{MSE}(\hat{\boldsymbol{\theta}}) &= E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 \\ &= E(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 + E(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^2 + E(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \text{MSE}(\tilde{\boldsymbol{\theta}}) + \text{E}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^2\end{aligned}$$

since one can show $E(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = 0$ if the v_i, e_{it} and u_i are all normally distributed (See Rao 2003, pg 103). To derive $\text{MSE}(\tilde{\boldsymbol{\theta}})$, we follow the Rao's (2003, pg 98) development,

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{t}(\boldsymbol{\delta}, \boldsymbol{y}) = \boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) + \boldsymbol{d}^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

where $\boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y})$ is the BLUP estimator when $\boldsymbol{\beta}$ is known

$$\boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) = \ell^T \boldsymbol{\beta} + \boldsymbol{b}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

with $\boldsymbol{b}^T = \boldsymbol{m}^T \boldsymbol{G}\boldsymbol{Z}^T \boldsymbol{V}^{-1}$ and $\boldsymbol{d}^T = \ell^T - \boldsymbol{b}^T \boldsymbol{X}$. Since $\boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) - \boldsymbol{\mu}$ and $\boldsymbol{d}^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ are uncorrelated;

$$\begin{aligned}\text{MSE}(\boldsymbol{t}(\boldsymbol{\delta}, \boldsymbol{y})) &= \text{MSE}(\boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y})) + \text{Var}(\boldsymbol{d}^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &= g_1(\boldsymbol{\delta}) + g_2(\boldsymbol{\delta})\end{aligned}$$

where

$$\begin{aligned}g_1(\boldsymbol{\delta}) &= \text{Var}(\boldsymbol{t}^*(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{y}) - \boldsymbol{\mu}) \\ &= \boldsymbol{m}^T(\boldsymbol{G} - \boldsymbol{G}\boldsymbol{Z}^T \boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{G})\boldsymbol{m}\end{aligned}$$

and $g_2(\boldsymbol{\delta}) = \boldsymbol{d}^T(\boldsymbol{X}^T \boldsymbol{V}^{-1}\boldsymbol{X})\boldsymbol{d}$.

Given the block diagonal structure of the SEPH model, we have that the estimator of AWE for the i^{th} area at time t is given by

$$\tilde{\theta}_{it} = t_i(\boldsymbol{\delta}_i, \boldsymbol{y}_i) = \overline{\boldsymbol{X}}_{it}^T \tilde{\boldsymbol{\beta}} + \boldsymbol{m}_i^T \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}})$$

and $\text{MSE}(\hat{\theta}_{it}) = g_{1it}(\boldsymbol{\delta}) + g_{2it}(\boldsymbol{\delta})$ where

$$\begin{aligned}g_{1it}(\boldsymbol{\delta}) &= \boldsymbol{m}_i^T(\boldsymbol{G}_i - \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \boldsymbol{G}_i)\boldsymbol{m}_i, \\ g_{2it}(\boldsymbol{\delta}) &= \boldsymbol{d}_{it}^T(\boldsymbol{X}^T \boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{d}_{it}, \\ \boldsymbol{d}_{it}^T &= \overline{\boldsymbol{X}}_{it}^T - \boldsymbol{b}_i^T \boldsymbol{X}_i\end{aligned}$$

and $\boldsymbol{b}_i^T = \boldsymbol{m}_i \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}$.

Noting that $\boldsymbol{m}_i^T = (1,0)\boldsymbol{Z}_i$ and recalling that $\boldsymbol{V}_i = \boldsymbol{\psi}_i + \boldsymbol{Z}_i \boldsymbol{G}_i \boldsymbol{Z}_i^T$, $g_{1it}(\boldsymbol{\delta})$ can be expressed, after some algebra, as

$$\begin{aligned}g_{1it}(\boldsymbol{\delta}) &= (1,0)\boldsymbol{Z}_i(\boldsymbol{G}_i - \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \boldsymbol{G}_i)\boldsymbol{Z}_i^T(1,0)^T \\ &= (1,0)(\boldsymbol{\psi}_i(\boldsymbol{I}_2 - \boldsymbol{V}_i^{-1}\boldsymbol{\psi}_i))(1,0)^T.\end{aligned}$$

Turning to $g_{2it}(\boldsymbol{\delta})$, noting that $\overline{\boldsymbol{X}}_{it}^{T} = (1,0)\boldsymbol{X}_i$ gives

$$\begin{aligned}\boldsymbol{d}_{it}^{T} &= \overline{\boldsymbol{X}}_{it}^{T} - \boldsymbol{m}_i^{T}\boldsymbol{G}_i\boldsymbol{Z}_i^{T}\boldsymbol{V}_i^{-1}\boldsymbol{X}_i \\ &= (1,0)\boldsymbol{\psi}_i\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\end{aligned}$$

and

$$g_{2it}(\boldsymbol{\delta}) = (1,0)\boldsymbol{\psi}_i\boldsymbol{V}_i^{-1}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{\psi}_i(1,0)^T.$$

It remains to evaluate $E(\hat{\theta}_{iT} - \tilde{\theta}_{iT})^2$. Following Prasad and Rao (1990),

$$E(\hat{\theta}_{it} - \tilde{\theta}_{it})^2 \approx tr\left[\boldsymbol{\varDelta}^T\boldsymbol{V}\boldsymbol{\varDelta}\Sigma^*\right] \equiv g_{3it}(\sigma^2, \sigma_v^2)$$

where $\boldsymbol{\Sigma}^*$ is the 2 × 2 covariance matrix of the unbiased estimators $\tilde{\sigma}^2$ and $\tilde{\sigma}_v^2$. Further, $\boldsymbol{\varDelta} = \left(\partial\boldsymbol{b}/\partial\sigma^2, \partial\boldsymbol{b}/\partial\sigma_v^2\right)$ where **b** is a $2m$ x 1 vector given as

$$\boldsymbol{b}^T = (0,\ldots,0,\boldsymbol{b}_i',0,\ldots,0)$$

with $\boldsymbol{b}_i' = \begin{pmatrix}\sigma_v^2 + 2\sigma^2 \\ \sigma_v^2 + \sigma^2\end{pmatrix}^T V_i^{-1}$. We have $\boldsymbol{\varDelta}^T\boldsymbol{V}\boldsymbol{\varDelta} = \boldsymbol{\varDelta}_i^T\boldsymbol{V}_i\boldsymbol{\varDelta}_i$ where $\boldsymbol{\varDelta}_i = \left(\partial\boldsymbol{b}_i/\partial\sigma_v^2, \partial\boldsymbol{b}_i/\partial\sigma^2\right)$ with $\boldsymbol{b}_i = \boldsymbol{V}_i^{-1}\boldsymbol{A}$ and . Differentiating $\boldsymbol{b}_i$ with respect to σ^2 we obtain

$$\partial\boldsymbol{b}_i/\partial\sigma^2 = \partial\boldsymbol{V}_i^{-1}/\partial\sigma^2\,\boldsymbol{A} + \boldsymbol{V}_i^{-1}\,\partial\boldsymbol{A}/\partial\sigma^2$$

where $\partial\boldsymbol{V}_i^{-1}/\partial\sigma^2 = -\boldsymbol{V}_i^{-1}\left(\partial\boldsymbol{V}_i/\partial\sigma^2\right)\boldsymbol{V}_i^{-1}$. Recalling that

$$\begin{aligned}\boldsymbol{V}_i &= \boldsymbol{\psi}_i + \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i^T \\ &= \boldsymbol{\psi}_i + \begin{pmatrix}\sigma_v^2 + 2\sigma^2 & \sigma_v^2 + \sigma^2 \\ \sigma_v^2 + \sigma^2 & \sigma_v^2 + \sigma^2\end{pmatrix}.\end{aligned}$$

and that $\boldsymbol{\psi}_i$ is independent of σ^2 we have $\dfrac{\partial\boldsymbol{V}_i}{\partial\sigma^2} = \begin{pmatrix}2 & 1 \\ 1 & 1\end{pmatrix}$. Also, $\dfrac{\partial\boldsymbol{A}_i}{\partial\sigma^2} = \begin{pmatrix}2 \\ 1\end{pmatrix}$ and $\dfrac{\partial\boldsymbol{A}_i}{\partial\sigma_v^2} = \begin{pmatrix}1 \\ 1\end{pmatrix}$ which gives,

$$\begin{aligned}\frac{\partial\boldsymbol{b}_i}{\partial\sigma^2} &= -\boldsymbol{V}_i^{-1}\begin{pmatrix}2 & 1 \\ 1 & 1\end{pmatrix}\boldsymbol{V}_i^{-1}\boldsymbol{A} + \boldsymbol{V}_i^{-1}\begin{pmatrix}2 \\ 1\end{pmatrix} \\ &= \boldsymbol{V}_i^{-1}\left[\begin{pmatrix}2 \\ 1\end{pmatrix} - \begin{pmatrix}2 & 1 \\ 1 & 1\end{pmatrix}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right].\end{aligned}$$

Similarly, one can show

$$\frac{\partial\boldsymbol{b}_i}{\partial\sigma_v^2} = \boldsymbol{V}_i^{-1}\left[\begin{pmatrix}1 \\ 1\end{pmatrix} - \begin{pmatrix}1 & 1 \\ 1 & 1\end{pmatrix}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right].$$

The matrix $\boldsymbol{\varDelta}_i$ simplifies to

$$\boldsymbol{\Delta}_i = \left(\boldsymbol{V}_i^{-1}\left[\boldsymbol{I}_2 - \boldsymbol{J}_2\boldsymbol{V}_i^{-1}\boldsymbol{A}\right], \boldsymbol{V}_i^{-1}\left[\boldsymbol{\lambda}_1 - \boldsymbol{\Lambda}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right]\right),$$

where $\boldsymbol{\lambda}_1^T$ is the first row of $\boldsymbol{\Lambda}_i$, $\boldsymbol{I}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{J}_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. The elements of the 2×2 matrix $\boldsymbol{B}_i = \boldsymbol{\Delta}_i^T \boldsymbol{V}_i \boldsymbol{\Delta}_i$ are,

$$\begin{aligned} b_{11} &= \left[\boldsymbol{I}_2 - \boldsymbol{J}_2\boldsymbol{V}_i^{-1}\boldsymbol{A}\right]^T \boldsymbol{V}_i^{-1}\left[\boldsymbol{I}_2 - \boldsymbol{J}_2\boldsymbol{V}_i^{-1}\boldsymbol{A}\right] \\ b_{22} &= \left[\boldsymbol{\lambda}_1 - \boldsymbol{\Lambda}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right]^T \boldsymbol{V}_i^{-1}\left[\boldsymbol{\lambda}_1 - \boldsymbol{\Lambda}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right] \\ b_{12} &= b_{21} = \left[\boldsymbol{I}_2 - \boldsymbol{J}_2\boldsymbol{V}_i^{-1}\boldsymbol{A}\right]^T \boldsymbol{V}_i^{-1}\left[\boldsymbol{\lambda}_1 - \boldsymbol{\Lambda}\boldsymbol{V}_i^{-1}\boldsymbol{A}\right]. \end{aligned}$$

The MSE of the EBLUP is given as

$$\text{MSE}(\hat{\theta}_{it}) = g_{1it}(\boldsymbol{\delta}) + g_{2it}(\boldsymbol{\delta}) + g_{3it}(\boldsymbol{\delta})$$

and can be estimated by (following Prasad and Rao, 1990)

$$\widehat{\text{MSE}}(\hat{\theta}_{it}) = g_{1it}(\hat{\boldsymbol{\delta}}) + g_{2it}(\hat{\boldsymbol{\delta}}) + 2g_{3it}(\hat{\boldsymbol{\delta}})$$

where $g_{1it}(\hat{\boldsymbol{\delta}})$, $g_{2it}(\hat{\boldsymbol{\delta}})$ and $g_{3it}(\hat{\boldsymbol{\delta}})$ are obtained from $g_{1it}(\boldsymbol{\delta})$, $g_{2it}(\boldsymbol{\delta})$ and $g_{3it}(\boldsymbol{\delta})$ with $\boldsymbol{\delta}$ replaced by $\hat{\boldsymbol{\delta}}$.

3. EMPIRICAL ILLUSTRATION

A simulated population based on actual SEPH data was used to compare the GREG and the proposed EBLUP estimator. 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 were defined at a level higher than the detailed industry level as some detailed industry domains had very few sampled units. The GREG estimates were then used as the direct estimates in the small area model.

In order to use equation (10), estimates for ψ_i, σ_v^2 and σ^2 are needed. In practice, estimates of the variance of the sampling errors, ψ_i, are typically smoothed by fitting regression models to the estimated sampling errors. Since the population is available in our empirical study, ψ_i was estimated by drawing 1,000 independent samples and then averaging over the samples. The variance components σ_v^2 and σ^2 were estimated using equations (7) and (9). A problem that occurs when estimating the variance components is that some estimates can be negative. Wang and Fuller (2003) suggest replacing the negative estimates with an estimate of the variance of σ_v^2 and σ^2 to get around this problem. Details on estimating the variance of σ_v^2 and σ^2 are available in Rubin-Bleuer (2009). With the appropriate estimates of σ_v^2 and σ^2, the EBLUP estimator is given by (10).

The EBLUP estimator was calculated 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 Other Services (except Public Administration). In the Construction industry the sample sizes of the 'small areas' varied from 1 to 77 observations, while they varied from 1 to 22 for the Other Services industry. In order to compare the design properties of the estimators, 1,000 independent samples were drawn and the estimated relative bias, the variance and the MSE were averaged over the samples. The average relative biases of the two 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. Both the GREG and the EBLUP estimators are quite variable; especially with the sample sizes are very small (less than 15). For the Construction industry, the two estimators become essentially equivalent as the sample gets large (n = 36 or 77).

Turning to the estimated variance/MSE, we see that, generally, the EBLUP estimator performs better than the GREG (see figures 3 and 4). The GREG estimator is quite variable in domains where the sample sizes are very small but seems to settle down when n becomes moderately large (n >15). The EBLUP estimator appears to be less affected by sample size but there are still some domains where the GREG estimator outperforms it despite the small sample size.

4. CONCLUSIONS AND FUTURE WORK

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. A small area estimator combining time series and cross sectional components is proposed to allow estimation for domains with very few sample units. While the proposed estimator is shown to be more precise that the GREG estimator through an empirical study, there is some work that needs to be done. Using the method of moments approach, many negative estimates of the variance components were obtained. Other estimation methods should be investigated to see if the number of negative estimates could be reduced. In the empirical study, the variance of the sampling errors, ψ_i's, were estimated over repeated samples. In real life applications this would not be possible and some smoothing would be necessary. The proposed estimator should be evaluated using a smoothed estimate of the ψ_i's. The consistency of the MSE estimator needs to be proved and other estimation methods, perhaps the bootstrap, should be investigated.

REFERENCES

Prasad, N.G.N. and Rao, J.N.K. (1990), The Estimation of the Mean Squared Error of Small-area Estimators, *Journal of the American Statistical Association*, **85**, 163-171

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.

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.

Figure 1 Relative Bias – Construction Industry

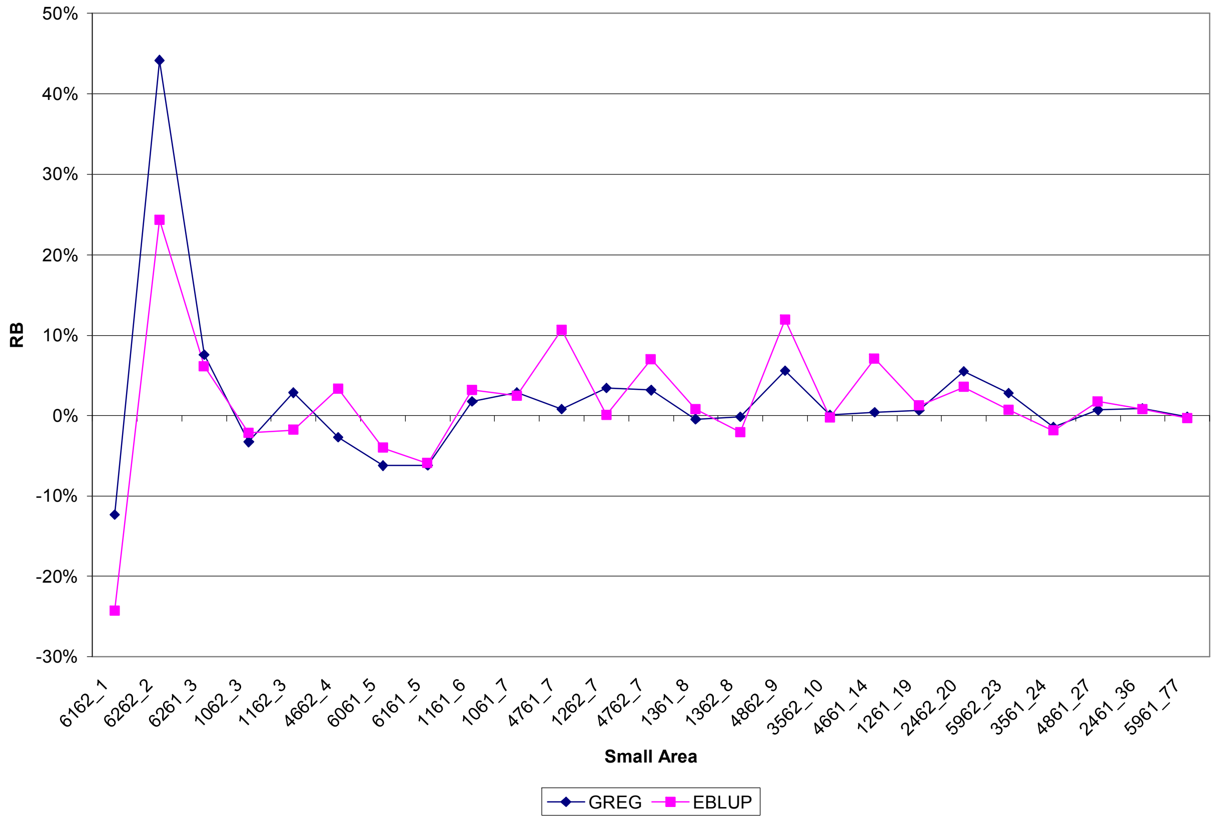

Figure 2 Relative Bias – Other Services (except Public Administration)

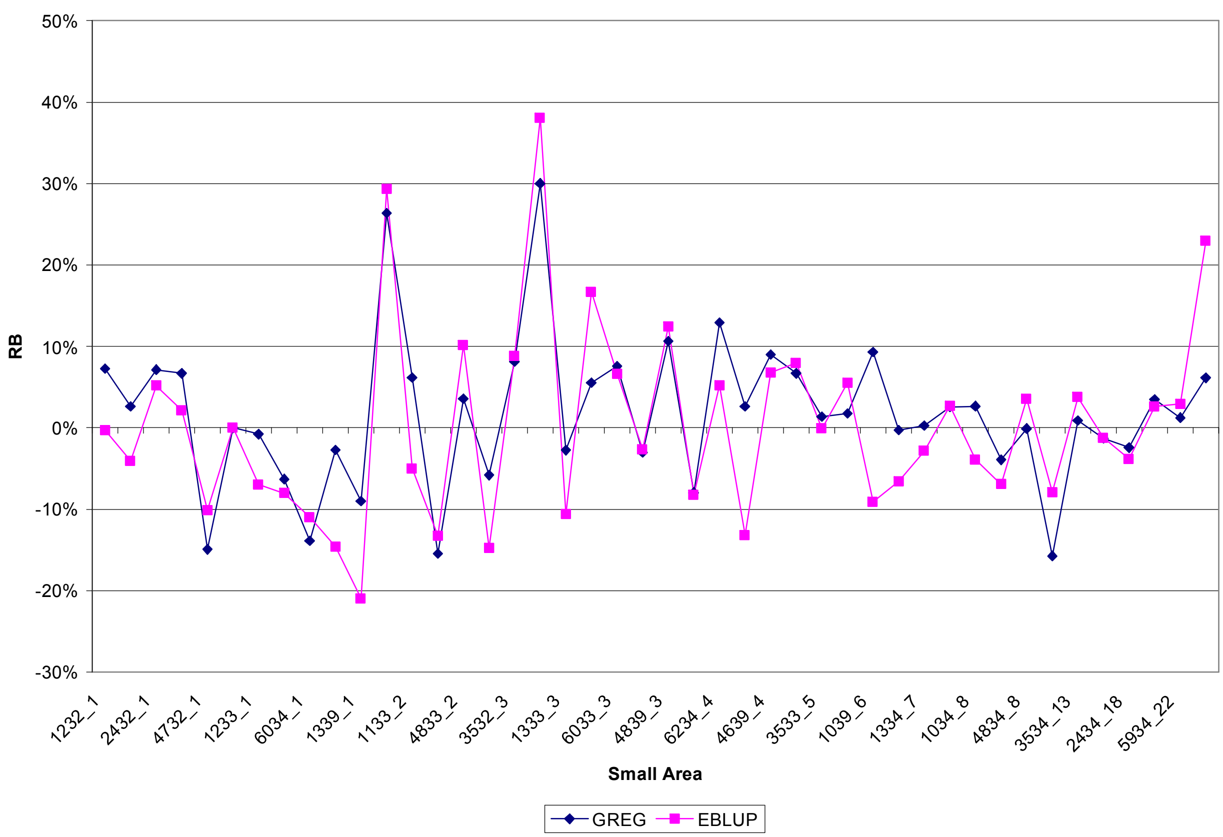

Figure 3 Variance/MSE – Construction

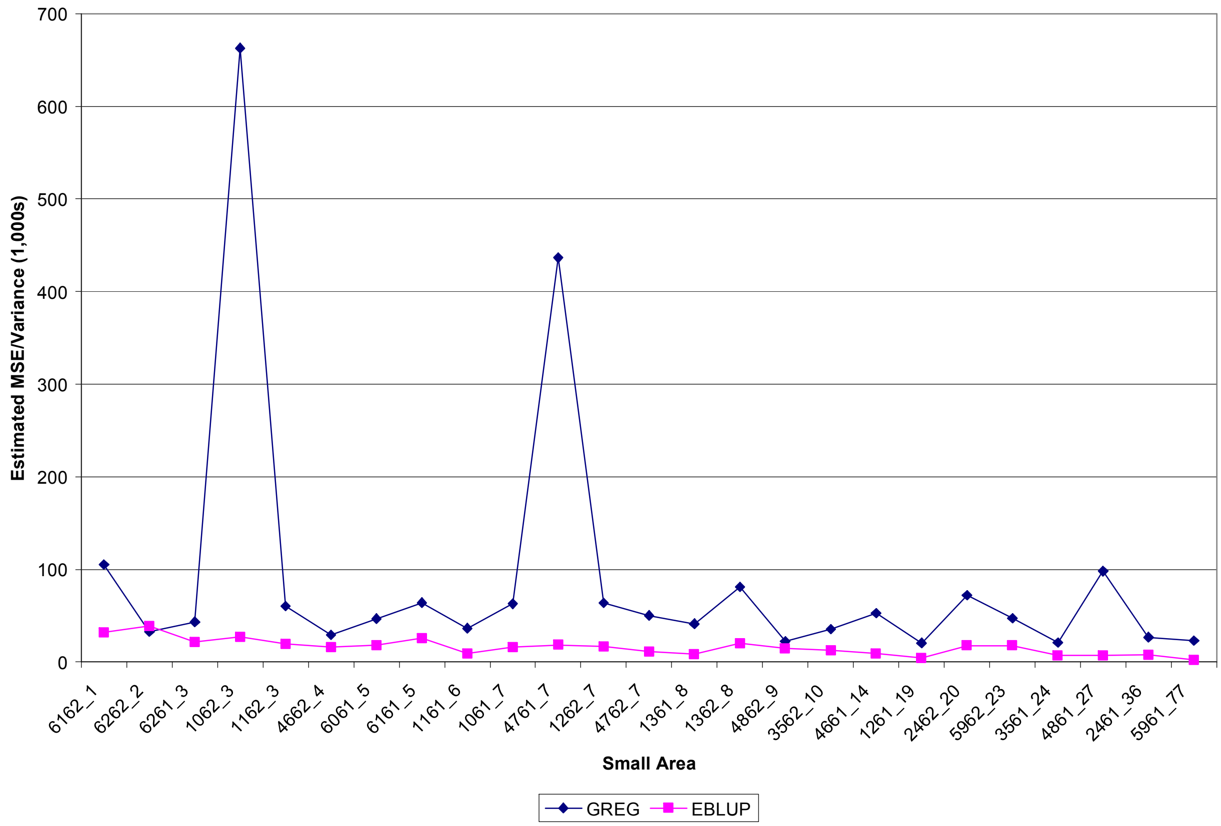

Figure 4 Variance/MSE – Other Services (except Public Administration)