# SMALL AREA ESTIMATION BY COMBINING TIME SERIES AND CROSS-SECTIONAL DATA

**J.N.K. Rao and Mingyu Yu, Carleton University**
**J.N.K. Rao, Dept. of Math & Stats, Carleton Univ., Ottawa, Canada, K1S 5B6**

**Key Words: Autoregressive process, Best linear unbiased predictor, Cross-sectional data, Hirarchical Bayes, Gibbs sampling, Two-stage estimation.**

## SUMMARY

A model involving random effects and autocorrelated errors is proposed for small area estimation, using both time series and cross-sectional data. This model is an extension of the well-known Fay-Herriot model for cross-sectional data. A two-stage estimator (predictor) of a small area mean at a given time point is obtained under the proposed model, by first deriving the best linear unbiased predictor (BLUP) assuming that the variance components and the autocorrelations that determine the variance-covariance matrix are known, and then replacing them with their consistent estimators. Extending the approach of Prasad and Rao (1986, 1990) for the Fay-Herriot model, an estimator of the mean square error of the two-stage estimator, correct to a second-order approximation, is obtained. A hierarchical Bayes approach, using Gibbs sampling, is also outlined.

## 1. INTRODUCTION

Small area statistics are needed in formulating policies and programs, in allocation of government founds, and in regional programs, etc. Demand for reliable small area statistics has steadily increased in recent years which prompted considerable research on efficient small area estimation.

Direct small area estimators from survey data fail to borrow strength from related small areas since they are based solely on the sample data associated with the corresponding areas. As a result, they are likely to yield unacceptably large standard errors unless the sample size for the small area is reasonably large. Alternative estimators that borrow strength from related small areas are therefore needed to improve efficiency. Such estimators are based on either implicit or explicit models which provide a link to related small areas through supplementary data such as administrative records and recent census counts.

Most of the research on small area estimation was focused on cross-sectional data at a given point in time. Rao (1986) has given an account of this research. Estimators proposed in the literature include (a) synthetic estimators (Gonzalez, 1973; Ericksen, 1974), structure preserving estimators (Purcell and Kish, 1980); (b) sample-size dependent estimators (Drew et al., 1982; Särndal and Hidiroglou, 1989); (c) empirical Bayes estimators (Fay and Herriot, 1979; Ghosh and Lahiri, 1987), empirical best linear unbiased predictors (Prasad and Rao, 1986 and 1990; Battese et al., 1988); (d) hierarchical Bayes estimators (Datta and Ghosh, 1991).

Scott and Smith (1974) and Jones (1980) used time series methods to develop efficient estimators of aggregates (e.g., overall means) from repeated surveys, by combining the direct survey estimates over time. Tiller (1989) used the Kalman filter to combine a current period state-wide estimate from the Current Population Survey with past estimators for the same state and auxiliary data from the unemployment insurance system and the Current Employment Statistics payroll survey. However, neither Scott and Smith (1974) nor Tiller (1989) considered small area estimation by combining time series and cross-sectional data.

The main purpose of this paper is to propose cross-sectional and time series models with random effects and autocorrelated errors, and to obtain empirical best linear unbiased predictors and associated standard errors for small areas at each time point using these models. Section 2 reviews some work on regression synthetic estimators and empirical Bayes estimators obtained from cross-sectional data at a given point in time. Cross-section and time series models are considered in Section 3, and an extension of the Fay-Herriot (1979) model is proposed. Two-stage estimators (empirical best linear unbiased predictors) of small area means are given in Section 4, and an estimator of mean square error (MSE) of a two-stage estimator, correct to a second-order approximation, is obtained in Section 5. Finally, Section 6 outlines a hierarchical Bayes approach, using Gibbs sampling, to obtain the posterior mean and the posterior variance of a small area mean at a given time point.

## 2. CROSS-SECTIONAL ESTIMATORS

### 2.1. Regression synthetic estimators

Let $y_{it}$ be the direct survey estimator of $i$-th small area mean at time point $t$, say $\theta_{it}$ ($i = 1, \ldots, m$; $t = 1, \ldots, T$). We assume that $y_{it}$ is unbiased for $\theta_{it}$, i.e., $y_{it} = \theta_{it} + e_{it}$, where the $e_{it}$'s are sampling errors with $E(e_{it}) = 0$, given $\theta_{it}$. We assume that a vector of concomitant variables, $\underset{\sim}{x}_{it} = (x_{it1}, \ldots, x_{itp})'$ related to $\theta_{it}$ is available such that $\theta_{it} = \underset{\sim}{x}'_{it}\underset{\sim}{\beta}_t$, where $\underset{\sim}{\beta}_t = (\beta_{t1}, \ldots, \beta_{tp})'$ is the vector of regression coefficients. A regression synthetic estimator of $\theta_{it}$, based solely on the cross-sectional data $\{(y_{it}, \underset{\sim}{x}_{it}),\ i = 1, \ldots, m\}$ for time $t$, is then given by

$$\tilde{\theta}_{it\,(reg)} = \underset{\sim}{x}'_{it}\,\underset{\sim}{\tilde{\beta}}_t \tag{2.1}$$

where $\underset{\sim}{\tilde{\beta}}_t$ is the ordinary least squares estimator of $\underset{\sim}{\beta}_t$ obtained from the combined model

$$y_{it} = \underset{\sim}{x}'_{it}\,\underset{\sim}{\beta}_t + e_{it}, \quad i = 1, \ldots, m.$$

Alternatively, we can use the generalized least squares estimator of $\beta$ if the estimated covariance matrix of $\underset{\sim}{\tilde{y}}_t = (y_{1t}, \ldots, y_{mt})'$ is available.

Synthetic estimators like (2.1) could lead to substantial biases since they do not give a weight to the direct estimator $y_{it}$. On the other hand, empirical Bayes or two-stage estimators give proper weights to the survey estimator and the synthetic estimator, and as a result lead to smaller biases relative to synthetic estimators.

### 2.2. Empirical Bayes or two-stage estimators

Following Fay and Herriot (1979), we introduce uncertainty into the model $\theta_{it} = \underset{\sim}{x}'_{it}\underset{\sim}{\beta}_t$ as follows:

$$\theta_{it} = \underset{\sim}{x}'_{it}\,\underset{\sim}{\beta}_t + v_{it},$$

where the $v_{it}$'s are independent random variables, for each $t$, with mean 0 and unknown variance $\sigma^2_{vt}$. For sampling errors, we assume that the $e_{it}$'s are independent normal variables with $E(e_{it}) = 0$ and $V(e_{it}) = \sigma^2_{it}$, where $\sigma^2_{it}$ is known. The combined model is then given by

$$y_{it} = \underset{\sim}{x}'_{it}\beta_t + v_{it} + e_{it}. \tag{2.2}$$

Under this model, the empirical Bayes estimator (or the two-stage estimator) of $\theta_{it}$ is given as a weighted sum of the direct estimator $y_{it}$ and the regression synthetic estimator $\hat{\theta}_{it\,(reg)} = \underset{\sim}{x}'_{it}\hat{\beta}_t$:

$$\hat{\theta}_{it}(\hat{\sigma}^2_{vt}, \underset{\sim}{\tilde{y}}_t) = w_{it}y_{it} + (1 - w_{it})\hat{\theta}_{it\,(reg)} \tag{2.3}$$

where $w_{it} = \hat{\sigma}^2_{vt}/(\hat{\sigma}^2_{vt} + \sigma^2_{it})$, $\underset{\sim}{\hat{\beta}}$ is the weighted least squares estimator of $\underset{\sim}{\beta}$ under the combined model (2.3), and $\hat{\sigma}^2_{vt}$ is an estimator of $\sigma^2_{vt}$. A simple moment estimator of $\sigma^2_{vt}$ (Prasad and Rao, 1990) or a more complicated estimator, such as the maximum likelihood estimtor, may be used. Using a moment estimator of $\sigma^2_{vt}$, Prasad and Rao (1990) obtained an estimator of mean squre error of $\hat{\theta}_{it}(\hat{\sigma}^2_{vt}, \underset{\sim}{\tilde{y}}_t)$, accurate to a second order of approximation as the number of small areas, $m$, increases, by taking account of the uncertainty in the estimator of $\sigma^2_{vt}$.

Fay and Herriot (1979) used estimators of the form (2.4) to estimate per capital income for small areas (with population less than 500 or 1000) from the 1970 U.S. Census of Population and Housing. They presented empirical evidence that (2.4) leads to smaller average error than either the direct survey estimator or the synthetic estimator using the county average. Datta, Fay and Ghosh (1991) extended the Fay-Herriot model to multiple characteristics of interest, and derived empirical Bayes and hierarchical Bayes estimators of small area means.

## 3. CROSS-SECTIONAL AND TIME SERIES MODELS

The methods in Section 2 use only cross-sectional data at each point in time. As a result they do not exploit the information in the data at other time points. In the following section, we extend the Fay-Herriot approach for small area estimation to time series of cross-sectional survey estimators of small areas in conjunction with census data and time varying supplementary data such as administrative records.

Extensive econometric literature exists on modelling and estimating relationships that combine time series and cross-sectional data (e.g., see Judge et al., 1985, Chapter 13), but sampling errors are seldom taken into account. We now consider some of these models on $\theta_{it}$:

$$\text{(I)} \qquad \theta_{it} = \underset{\sim}{x}'_{it}\,\underset{\sim}{\beta} + v_i + \varepsilon_{it},$$

where $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of regression parameters, the $v_i$'s are fixed small area effects and the $\varepsilon_{it}$'s are independent normal variables with mean 0 and variance $\sigma^2$, abbreviated $\varepsilon_{it} \overset{iid}{\sim} (0, \sigma^2)$.

$$\text{(II)} \qquad \theta_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i + \varepsilon_{it},$$

where $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$ and $\{v_i\}$ and $\{\varepsilon_{it}\}$ are independent. Here the $v_i$'s are random small area effects.

$$\text{(III)} \qquad \theta_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i + u_t + \varepsilon_{it},$$

where $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$, $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$ and $\{v_i\}$, $\{u_t\}$, $\{\varepsilon_{it}\}$ are mutually independent. Here the $v_i$'s and the $u_t$'s are random small area effects and random time effects respectively.

$$\begin{aligned} \text{(IV)} \qquad & \theta_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i + u_{it} \\ \text{and} \qquad & u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1, \end{aligned} \tag{3.1}$$

where $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$ and $\{v_i\}$, $\{\varepsilon_{it}\}$ are independent (Anderson and Hsiao, 1981). Here the $v_i$'s are random small area effects and the $u_{it}$'s follow an AR(1) process. Model (3.1) may be rewritten as a distributed lag model:

$$\theta_{it} = \rho\theta_{i,t-1} + (\underset{\sim}{x}_{it} - \rho \underset{\sim}{x}_{i,t-1})' \underset{\sim}{\beta} + (1-\rho)v_i + \varepsilon_{it}. \tag{3.1a}$$

Model IV appears to be the most realistic among the four models since the alternative form (3.1a) relates the current mean, $\theta_{it}$, to the previous period mean, $\theta_{i,t-1}$, and the values of the auxiliary variables for the current and previous periods, $\underset{\sim}{x}_{it}$ and $\underset{\sim}{x}_{i,t-1}$ respectively. More complex models may be formulated by assuming an ARMA process for the $u_{it}$'s instead of the simple AR(1) process, but the resulting dividends are unlikely to be significant. Similarly, random slopes $\{\underset{\sim}{\beta}_{it}\}$ obeying an autoregressive process may be used in place of constant slopes $\underset{\sim}{\beta}$ (Pfeffermann and Burck, 1990; Singh et al., 1991).

The $\theta_{it}$'s are related to the direct survey estimators through

$$y_{it} = \theta_{it} + e_{it}, \quad i = 1, \ldots, m; \; t = 1, \ldots, T. \tag{3.2}$$

Following Fay and Herriot (1979), we assume the covariance matrix of sampling errors, $e_{it}$, to be block diagonal with known blocks $\underset{\sim}{\Sigma}_i$, where $\underset{\sim}{\Sigma}_i$ is a $T \times T$ matrix, and $E(e_{it}) = 0$. Recent research has focused on modelling sampling errors of aggregates. For example, Binder and Dick (1989) and Tiller (1989) proposed ARMA models.

Choudhry and Rao (1989) treated the composite error $w_{it} = e_{it} + u_{it}$ as an AR(1) process: $w_{it} = \rho w_{i,t-1} + \varepsilon_{it}$ with $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$, and then considered $\theta_{it}$ as $\theta_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i$. The combined model under the above assumption, may be written as

$$\begin{aligned} & y_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i + w_{it} \\ \text{and} \qquad & w_{it} = \rho w_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1 \end{aligned} \tag{3.3}$$

where $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, and $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$. Tiller (1989) used a similar approach in the context of labour force estimation from aggregate time series data generated from repeated surveys. Model (3.3) does not depend on the sampling error covariance matrix, but it is less realistic than the combined model using (3.1) and (3.2):

$$\begin{aligned} & y_{it} = \underset{\sim}{x}'_{it} \underset{\sim}{\beta} + v_i + u_{it} + e_{it} \\ \text{and} \qquad & u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1 \end{aligned} \tag{3.4}$$

where $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $\varepsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$ and the $e_{it}$'s are normally distributed with zero mean and known block diagonal covariance matrix $\underset{\sim}{\Sigma} =$ block diag$(\underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_m)$. Model (3.4) provides an extension of the Fay-Herriot model to time series and cross-sectional data.

Choudhry and Rao (1989) obtained a two-stage estimator of small area mean $\theta_{it}$ under (3.3), and evaluated its efficiency relative to two synthetic estimators and the direct estimator, $y_{it}$, using monthly survey estimates of unemployment for census divisions (small areas) from the Canadian Labour Force Survey in conjunction with monthly administrative counts from the Unemployment Insurance System and monthly survey estimates of population in labour force as auxiliary variables.

We focus on the extended Fay-Herriot model (3.4) and obtain a two-stage estimator of $\theta_{it}$ in Section 4.

## 4. TWO-STAGE ESTIMATOR

Arranging the data $\{y_{it}\}$ as $\underset{\sim}{y} = (y_{11}, \ldots, y_{1T}; \ldots; y_{m1}, \ldots, y_{mT})' = (\underset{\sim}{y}'_1, \ldots, \underset{\sim}{y}'_m)'$ the proposed model (3.4) may be written, in matrix form, as

$$\underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{Z} \underset{\sim}{v} + \underset{\sim}{u} + \underset{\sim}{e} \tag{4.1}$$

with

$$\underset{\sim}{X} = (\underset{\sim}{X}'_1, \ldots, \underset{\sim}{X}'_m)', \qquad \underset{\sim}{X}'_i = (\underset{\sim}{x}_{i1}, \ldots, \underset{\sim}{x}_{iT}),$$
$$\underset{\sim}{Z} = \underset{\sim}{I}_m \otimes \underset{\sim}{\mathbf{1}},$$
$$\underset{\sim}{v} = (v_1, \ldots, v_m)', \qquad \underset{\sim}{u} = (\underset{\sim}{u}'_1, \ldots, \underset{\sim}{u}'_m)',$$
$$\underset{\sim}{e} = (\underset{\sim}{e}'_1, \ldots, \underset{\sim}{e}'_m)',$$

where $\underset{\sim}{u}_i' = (u_{i1}, \ldots, u_{iT})$, $\underset{\sim}{e}_i' = (e_{i1}, \ldots, e_{iT})$, $\underset{\sim}{1}_t$ is a $t$-vector of 1's, $\underset{\sim}{I}_m$ is the identity matrix of order $m$, and $\otimes$ denotes the direct product.

Further,

$$\begin{aligned} &E(\underset{\sim}{v}) = 0, \quad Cov(\underset{\sim}{v}) = \sigma_v^2 \underset{\sim}{I}_m, \\ &E(\underset{\sim}{u}) = 0, \quad Cov(\underset{\sim}{u}) = \sigma_u^2 \underset{\sim}{I}_m \otimes \underset{\sim}{\Gamma} = \sigma_u^2 \underset{\sim}{R}, \\ &E(\underset{\sim}{e}) = 0, \\ &Cov(\underset{\sim}{e}) = \underset{\sim}{\Sigma} = \text{block diag}(\underset{\sim}{\Sigma}_1, \ldots, \underset{\sim}{\Sigma}_m), \end{aligned} \tag{4.2}$$

and $\underset{\sim}{v}$, $\underset{\sim}{u}$ and $\underset{\sim}{e}$ are mutually independent, where $\underset{\sim}{\Gamma}$ is a $T \times T$ matrix with elements $\rho^{|i-j|}/(1-\rho^2)$. It follows from (4.1) and (4.2) that

$$\begin{aligned} Cov(\underset{\sim}{y}) &= \underset{\sim}{V} = \underset{\sim}{\Sigma} + \sigma_u^2 \underset{\sim}{R} + \sigma_v^2 \underset{\sim}{Z}\,\underset{\sim}{Z}' \\ &= \text{block diag}_i(\underset{\sim}{\Sigma}_i + \sigma_u^2 \underset{\sim}{\Gamma} + \sigma_v^2 \underset{\sim}{J}_T), \end{aligned}$$

with $\underset{\sim}{J}_T = \underset{\sim}{1}_T\, \underset{\sim}{1}_T'$.

### 4.1. Best Linear Unbiased Predictor

The small area mean $\theta_{it} = \underset{\sim}{x}_{it}' \underset{\sim}{\beta} + v_i + u_{it}$ is a special case of the linear combination $\tau = \underset{\sim}{l}' \underset{\sim}{\beta} + \underset{\sim}{l}_1' \underset{\sim}{v} + \underset{\sim}{l}_2' u$ with $\underset{\sim}{l} = \underset{\sim}{x}_{it}$, $\underset{\sim}{l}_1$ is the $m$-vector with 1 in the $i$-th position and 0 elsewhere, and $\underset{\sim}{l}_2$ is the $(mT)$-vector with 1 in the $(it)$-th position and 0 elsewhere. Noting that (4.1) is a special case of the general mixed linear model, the best linear unbiased predictor (BLUP) of $\tau = \theta_{it}$ can be obtained from Henderson's (1975) general results.

Assuming first that $\sigma_u^2$, $\sigma_v^2$ and $\rho$ are known, the BLUP of $\tau$ is given by

$$\tilde{\tau} = \underset{\sim}{l}' \tilde{\underset{\sim}{\beta}} + (\sigma_v^2 \underset{\sim}{l}_1' \underset{\sim}{Z}' + \sigma_u^2 \underset{\sim}{l}_2' \underset{\sim}{R}) \underset{\sim}{V}^{-1} (\underset{\sim}{y} - \underset{\sim}{X} \tilde{\underset{\sim}{\beta}}), \tag{4.3}$$

where $\tilde{\underset{\sim}{\beta}} = (\underset{\sim}{X}' \underset{\sim}{V}^{-1} \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{V}^{-1} \underset{\sim}{y}$ is the generalized least squares estimator of $\underset{\sim}{\beta}$. Using the special structures of $\underset{\sim}{l}_1$, $\underset{\sim}{l}_2$, $\underset{\sim}{Z}$, $\underset{\sim}{R}$ and $\underset{\sim}{V}$, it is easily seen that (4.3) reduces to

$$\begin{aligned} \tilde{\theta}_{it} &= t(\sigma_u^2, \sigma_v^2, \rho, \underset{\sim}{y}) \\ &= \underset{\sim}{x}_{it}' \tilde{\underset{\sim}{\beta}} + (\sigma_v^2 \underset{\sim}{1}_T' + \sigma_u^2 \underset{\sim}{\gamma}_t')(\underset{\sim}{\Sigma}_i \\ &\quad + \sigma_u^2 \underset{\sim}{\Gamma} + \sigma_v^2 \underset{\sim}{J}_T)^{-1} (\underset{\sim}{y}_i - \underset{\sim}{X}_i \tilde{\underset{\sim}{\beta}}), \end{aligned} \tag{4.4}$$

where $\underset{\sim}{\gamma}_t$ is the $t$-th row of $\underset{\sim}{\Gamma}$. An alternative form of (4.4) is given by

$$\tilde{\theta}_{it} = y_{it} - \underset{\sim}{l}_1' [\underset{\sim}{\Sigma}_i^{-1} + (\sigma_u^2 \underset{\sim}{\Gamma} + \sigma_v^2 \underset{\sim}{J}_T)^{-1}]^{-1} \times (\sigma_u^2 \underset{\sim}{\Gamma} + \sigma_v^2 \underset{\sim}{J}_T)^{-1} (\underset{\sim}{y}_i - \underset{\sim}{X}_i \tilde{\underset{\sim}{\beta}}).$$

### 4.2 Two-stage estimator

In practice, the parameters, $\sigma_u^2$, $\sigma_v^2$ and $\rho$ are usually unknown. A two-stage estimator (empirical BLUP or EBLUP) is therefore obtained by replacing them with their consistent estimators in the expression for BLUP:

$$\hat{\theta}_{it} = t(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}, \underset{\sim}{y}),$$

where $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\rho}$ are consistent estimators of $\sigma_u^2$, $\sigma_v^2$ and $\rho$ respectively.

Pantula and Pollock (1985) estimated $\sigma_u^2$, $\sigma_v^2$ and $\rho$ in the nested error regression model (3.3) with autocorrelated errors $w_{it}$ by extending the method of fitting constants for the special case of independent errors $w_{it}$ (Fuller and Battese, 1973). We now extend the method of Pantula and Pollock (1985) to the more general model (3.4) with autocorrelated errors $w_{it}$ and sampling errors $e_{it}$ with known block diagonal covariance matrix $\underset{\sim}{\Sigma}$.

**Estimator of $\rho$**

Let $a_{it} = v_i + u_{it} + e_{it}$. Then we have

$$\begin{aligned} &E\left[m^{-1}(T-2)^{-1} \sum_{i=1}^{m} \sum_{t=1}^{T-2} a_{it}(a_{it} - a_{i,t+1})\right] \\ &= \sigma_u^2(1-\rho) + m^{-1}(T-2)^{-1} \sum_{i=1}^{m} \sum_{t=1}^{T-2} (\sigma_{tt}^{(i)} - \sigma_{t,t+1}^{(i)}) \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} &E\left[m^{-1}(T-2)^{-1} \sum_{i=1}^{m} \sum_{t=1}^{T-2} a_{it}(a_{i,t+1} - a_{i,t+2})\right] \\ &= \sigma_u^2 \rho(1-\rho) \\ &\quad + m^{-1}(T-2)^{-1} \sum_{i=1}^{m} \sum_{t=1}^{T-2} (\sigma_{t,t+1}^{(i)} - \sigma_{t,t+2}^{(i)}), \end{aligned} \tag{4.6}$$

where $\sigma_{t,t+s}^{(i)} = cov(e_{it}, e_{i,t+s})$ and $\sigma_{tt}^{(i)} = var(e_{it})$. It follows from (4.5) and (4.6) that a moment estimator of $\rho$, assuming known errors $a_{it}$, is given by

$$\rho^* = \frac{\sum_{i=1}^{m} \sum_{t=1}^{T-2} [a_{it}(a_{i,t+1} - a_{i,t+2}) - (\sigma_{t,t+1}^{(i)} - \sigma_{t,t+2}^{(i)})]}{\sum_{i=1}^{m} \sum_{t=1}^{T-2} [a_{it}(a_{i,t} - a_{i,t+1}) - (\sigma_{t,t}^{(i)} - \sigma_{t,t+1}^{(i)})]}. \tag{4.7}$$

Now replace the $a_{ij}$'s in (4.7) by the ordinary least squares residuals $\hat{a}_{ij} = y_{it} - \underset{\sim}{x}_{it}' (\underset{\sim}{X}' \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{y}$ to get an estimator $\tilde{\rho}$. Since $|\tilde{\rho}|$ may be greater than or equal to 1, we need to truncate $\tilde{\rho}$:

$$\hat{\rho} = sign(\tilde{\rho})\ \max(1-\delta, |\tilde{\rho}|). \tag{4.8}$$

where $sign(x) = x/|x|$ when $x \neq 0$, or $= 0$ otherwise, and $\delta > 0$ is arbitrarily small. It can be shown that $\hat{\rho}$ is asymptotically equal to $\rho^*$.

**Estimators of $\sigma_u^2$ and $\sigma_v^2$**

We first obtain unbiased estimators of $\sigma_u^2$ and $\sigma_v^2$ assuming that $\rho$ is known. For this purpose, we transform the model (4.1) such that the covariance matrix of the transformed errors is independent of $\sigma_v^2$. Let

$$f_t = \begin{cases} (1-\rho^2)^{1/2}, & t = 1 \\ 1-\rho, & 2 \leq t \leq T, \end{cases}$$

and

$$\underset{\sim}{D} = c^{-1}\underset{\sim}{f}\,\underset{\sim}{f}\,',$$

where $\underset{\sim}{f} = (f_1, \ldots, f_T)'$ and

$$c = \underset{\sim}{f}\,'\underset{\sim}{f} = (1-\rho)[T-(T-2)\rho].$$

Also, we note that $\underset{\sim}{\Gamma} = \underset{\sim}{P}^{-1}(\underset{\sim}{P}^{-1})'$, where $\underset{\sim}{P}$ is a $T \times T$ matrix with first diagonal element $(1-\rho^2)^{1/2}$ and remaining diagonal elements 1, and $-\rho$ for the elements $(t+1, t)$ and 0 for the remaining elements, $t = 1, \ldots, T-1$.

Now transform $\underset{\sim}{y}_i$ to $\underset{\sim}{z}_i = \underset{\sim}{P}\underset{\sim}{y}_i$ so that

$$\underset{\sim}{z}_i = \underset{\sim}{P}\underset{\sim}{X}_i\underset{\sim}{\beta} + \underset{\sim}{f}\,v_i + \underset{\sim}{P}(\underset{\sim}{u}_i + \underset{\sim}{e}_i) \tag{4.9}$$

and

$$\underset{\sim}{z}_i^{(1)} = (\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{z}_i = \underset{\sim}{H}_i^{(1)}\underset{\sim}{\beta} + \underset{\sim}{e}_i^* \tag{4.10}$$

noting that $(\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{f} = 0$, where $\underset{\sim}{H}_i^{(1)} = (\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{P}\underset{\sim}{X}_i$ and $\underset{\sim}{e}_i^* = (\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{P}(\underset{\sim}{u}_i + \underset{\sim}{e}_i)$.

Since $Cov(\underset{\sim}{e}_i^*) = (\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{P}(\sigma_u^2\underset{\sim}{\Gamma} + \underset{\sim}{\Sigma}_i)\underset{\sim}{P}'(\underset{\sim}{I}_T - \underset{\sim}{D}) = (\underset{\sim}{I}_T - \underset{\sim}{D})(\sigma_u^2\underset{\sim}{I}_T + \underset{\sim}{P}\underset{\sim}{\Sigma}_i\underset{\sim}{P}')(\underset{\sim}{I}_T - \underset{\sim}{D})'$ is independent of $\sigma_v^2$, we can estimate $\sigma_u^2$ through model (4.10). Let $\underset{\sim}{z}^{(1)} = ((\underset{\sim}{z}_1^{(1)})', \cdots, (\underset{\sim}{z}_m^{(1)})')'$, and $\underset{\sim}{H}^{(1)} = ((\underset{\sim}{H}_1^{(1)})', \cdots, (\underset{\sim}{H}_m^{(1)})')'$. Denote $\underset{\sim}{P}_{\text{H}} = \underset{\sim}{H}^{(1)}(\underset{\sim}{H}^{(1)\prime}\underset{\sim}{H}^{(1)})^{+}\underset{\sim}{H}^{(1)\prime})$ and $\underset{\sim}{\Sigma}_{(u)} =$ block $\text{diag}_i(\underset{\sim}{P}\underset{\sim}{\Sigma}_i\underset{\sim}{P}')$. An unbiased estimator $\tilde{\sigma}_u^2$ is then given by

$$\tilde{\sigma}_u^2(\rho) = \big[\underset{\sim}{\hat{e}}'\,\underset{\sim}{\hat{e}} - tr\big\{(\text{block diag}_i(\underset{\sim}{I}_T - \underset{\sim}{D}) - \underset{\sim}{P}_{\text{H}}) \cdot \underset{\sim}{\Sigma}_{(u)}\big\}\big]\big[m(T-1) - rank(H^{(1)})\big]^{-1}, \tag{4.11}$$

where $\underset{\sim}{\hat{e}}'\,\underset{\sim}{\hat{e}}$ is the residual sum of squares obtained by regressing $\underset{\sim}{z}^{(1)}$ on $\underset{\sim}{H}^{(1)}$ using ordinary least squares. The unbiasedness of $\tilde{\sigma}_u^2(\rho)$ follows from the fact

$$\begin{aligned} E(\underset{\sim}{\hat{e}}'\,\underset{\sim}{\hat{e}}) &= E\big[\underset{\sim}{e}^{*\prime}(\underset{\sim}{I}_{mT} - \underset{\sim}{P}_{\text{H}})\underset{\sim}{e}^{*\prime}\big] \\ &= tr\big[(\underset{\sim}{I}_{mT} - \underset{\sim}{P}_{\text{H}})\text{block diag}_i(cov(\underset{\sim}{e}_i^*))\big] \\ &= \sigma_u^2\big[mT - m - rank(\underset{\sim}{H}^{(1)})\big] \\ &\quad + tr\big[\{\text{block diag}_i(\underset{\sim}{I}_T - \underset{\sim}{D}) - \underset{\sim}{P}_{\text{H}}\}\underset{\sim}{\Sigma}_{(u)}\big]. \end{aligned}$$

Turning to the estimation of $\sigma_v^2$, transform (4.9) to

$$\begin{aligned} c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{z}_i &= c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{P}\underset{\sim}{X}_i\underset{\sim}{\beta} \\ &\quad + c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{f}\,v_i + c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{P}(\underset{\sim}{u}_i + \underset{\sim}{e}_i) \end{aligned}$$

with error variance $c\sigma_v^2 + \sigma_u^2 + c^{-1}\underset{\sim}{f}\,'\underset{\sim}{P}\underset{\sim}{\Sigma}_i\underset{\sim}{P}'\underset{\sim}{f}$. Let $\underset{\sim}{\hat{u}}'\,\underset{\sim}{\hat{u}}$ be the residual sum of squares by regressing $c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{z}_i$ on $c^{-1/2}\underset{\sim}{f}\,'\underset{\sim}{P}\underset{\sim}{X}_i$ using ordinary least squares. Denote $\underset{\sim}{P}_{\text{F}} = \underset{\sim}{F}(\underset{\sim}{F}'\underset{\sim}{F})^{+}\underset{\sim}{F}'$ and $\underset{\sim}{\Sigma}_{(v)} = \text{diag}_i(c^{-1}\underset{\sim}{f}\,'\underset{\sim}{P}\underset{\sim}{\Sigma}_i\underset{\sim}{P}'\underset{\sim}{f})$. An unbiased estimator $\tilde{\sigma}_v^2$ is then obtained as

$$\begin{aligned} \tilde{\sigma}_v^2(\rho) &= c^{-1}[m - rank(\underset{\sim}{F})]^{-1}\big[\underset{\sim}{\hat{u}}'\,\underset{\sim}{\hat{u}} \\ &\quad - tr\{(\underset{\sim}{I}_m - \underset{\sim}{P}_{\text{F}})\underset{\sim}{\Sigma}_{(v)}\}\big] - c^{-1}\tilde{\sigma}_u^2(\rho), \end{aligned} \tag{4.12}$$

where $\underset{\sim}{F} = (\underset{\sim}{X}_1'\underset{\sim}{P}'\underset{\sim}{f}, \ldots, \underset{\sim}{X}_m'\underset{\sim}{P}'\underset{\sim}{f})'$. The unbiasedness of $\tilde{\sigma}_v^2$ follows from the fact

$$\begin{aligned} &E(\underset{\sim}{\hat{u}}'\,\underset{\sim}{\hat{u}}) = \\ &(c\sigma_v^2 + \sigma_u^2)(m - rank(\underset{\sim}{F})) + tr\big[(\underset{\sim}{I}_m - \underset{\sim}{P}_{\text{F}})\underset{\sim}{\Sigma}_{(v)}\big]. \end{aligned}$$

Note that $\underset{\sim}{P}_{\text{F}}$ and $\underset{\sim}{P}_{\text{H}}$ are invariant to the choice of generalized inverses $(\underset{\sim}{F}'\underset{\sim}{F})^{-}$ and $(\underset{\sim}{H}^{(1)\prime}\underset{\sim}{H}^{(1)})^{-}$ respectively, we have therefore chosen the Moore-Penrose inverses $(\underset{\sim}{F}'\underset{\sim}{F})^{+}$ and $(\underset{\sim}{H}^{(1)\prime}\underset{\sim}{H}^{(1)})^{+}$ in $\underset{\sim}{P}_{\text{F}}$ and $\underset{\sim}{P}_{\text{H}}$ respectively.

Two-step estimators of $\sigma_v^2$ and $\sigma_u^2$ are now obtained as

$$\hat{\sigma}_v^2 = \max\{0, \tilde{\sigma}_v^2(\hat{\rho})\}. \qquad \hat{\sigma}_u^2 = \max\{0, \tilde{\sigma}_u^2(\hat{\rho})\},$$

where $\hat{\rho}$ is given by (4.8). Note that $\tilde{\sigma}_v^2(\hat{\rho})$ and $\tilde{\sigma}_u^2(\hat{\rho})$ are no longer unbiased estimators, but the asymptotic consistency of $\hat{\sigma}_v^2$, and $\hat{\sigma}_u^2$ and $\hat{\rho}$ can be established, as $m \to \infty$.

A two-stage estimator of the small area mean $\theta_{it}$ is now obtained from (4.4) by substituting $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\rho}$ for $\sigma_u^2$, $\hat{\sigma}_v^2$ and $\rho$ respectively:

$$\hat{\theta}_{it} = t(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}, \underset{\sim}{y}). \tag{4.13}$$

The two-stage estimator, $\hat{\theta}_{it}$, remains unbiased, noting that $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$, and $\hat{\rho}$ are even functions of $\underset{\sim}{y}$ and translation invariant (i.e., $\hat{\sigma}_u^2(\underset{\sim}{y}) = \hat{\sigma}_u^2(-\underset{\sim}{y})$, $\hat{\sigma}_u^2(\underset{\sim}{y} - \underset{\sim}{X}\,\underset{\sim}{a}) = \hat{\sigma}_u^2(\underset{\sim}{y})$, for all $\underset{\sim}{y}$ and $\underset{\sim}{a}$, and similarly for $\hat{\sigma}_v^2$ and $\hat{\rho}$; see Kackar and Harville (1984) ). It is not necessary to assume normality of the errors in the model (3.4); only symmetric distributions are needed. However, we need normality to derive an estimator of MSE of $\hat{\theta}_{it}$, corrected to a second-order approximation (see Section 5).

## 5. ESTIMATOR OF MSE

### 5.1. Second order approximation to MSE

We first derive a second order approximation to MSE of the general EBLUP, $\hat{\tau} = t(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}, \underset{\sim}{y})$, where $\tilde{\tau} = t(\sigma_u^2, \sigma_v^2, \rho, \underset{\sim}{y})$ is the BLUP of the linear combination $\tau = \underset{\sim}{l}'\underset{\sim}{\beta} + \underset{\sim}{l}_1'\underset{\sim}{v} + \underset{\sim}{l}_2'\underset{\sim}{u}$ and it is given by (4.3).

Following Kackar and Harville (1984), we have

$$MSE[\hat{\tau}] = MSE[\tilde{\tau}] + E[\hat{\tau} - \tilde{\tau}]^2, \tag{5.1}$$

under normality of errors, where $MSE(\hat{\tau}) = E(\hat{\tau} - \tau)^2$ and $MSE(\tilde{\tau}) = E(\tilde{\tau} - \tau)^2$. Further, Henderson (1975) has given an exact expression for $MSE[\tilde{\tau}]$ as

$$\begin{aligned} MSE(\tilde{\tau}) &= \underset{\sim}{l}'(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{l} - (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' \\ &\quad + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}\underset{\sim}{A}(\sigma_v^2\underset{\sim}{Z}\,\underset{\sim}{l}_1 + \sigma_u^2\underset{\sim}{R}\,\underset{\sim}{l}_2) \\ &\quad - 2\,\underset{\sim}{l}'(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}^{-1}(\sigma_v^2\underset{\sim}{Z}\,\underset{\sim}{l}_1 + \sigma_u^2\underset{\sim}{R}\,\underset{\sim}{l}_2) \\ &\quad + \sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{l}_1 + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R}\,\underset{\sim}{l}_2, \end{aligned} \tag{5.2}$$

where $\underset{\sim}{A} = \underset{\sim}{I}_{mT} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{V}^{-1}$. If $\tau = \theta_{it}$, then $MSE(\tilde{\theta}_{it})$ is obtained from (5.2) by letting $\underset{\sim}{l} = \underset{\sim}{x}_{it}$, $\underset{\sim}{l}_1 = m$-vector with 1 in the $i$-th position and 0 elsewhere, and $\underset{\sim}{l}_2 =$ (mT)-vector with 1 in the $(it)$-th position and 0 elsewhere.

Following Kackar and Harville (1984) and Prasad and Rao (1990), we propose a Taylor series approximation to $E(\hat{\tau} - \tilde{\tau})^2$, by writing

$$\begin{aligned} E(\hat{\tau} - \tilde{\tau})^2 &= E[\hat{\tau} - t(\tilde{\sigma}_u^2(\rho), \tilde{\sigma}_v^2(\rho), \rho, \underset{\sim}{y})] \\ &\quad + \{t(\tilde{\sigma}_u^2(\rho), \tilde{\sigma}_v^2(\rho), \rho, \underset{\sim}{y}) - t(\sigma_u^2, \sigma_v^2, \rho, \underset{\sim}{y})\}]^2, \end{aligned}$$

We have

$$\begin{aligned} E(\hat{\tau} - \tilde{\tau})^2 &\approx E\big[(dt(\tilde{\sigma}_u^2(\rho), \tilde{\sigma}_v^2(\rho), \rho)/d\rho)(\tilde{\rho} - \rho) \\ &\quad + \big(\partial t(\sigma_u^2, \sigma_v^2, \rho)/\partial\sigma_u^2\big)(\tilde{\sigma}_u^2(\rho) - \sigma_u^2) \\ &\quad + \big(\partial t(\sigma_u^2, \sigma_v^2, \rho)/\partial\sigma_v^2\big)(\tilde{\sigma}_v^2(\rho) - \sigma_v^2)\big]^2, \end{aligned} \tag{5.3}$$

where the dependence on $\underset{\sim}{y}$ is suppressed for simplicity. Note that the truncated estimators $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\rho}$ are replaced by their untruncated counterparts $\tilde{\sigma}_u^2$, $\tilde{\sigma}_v^2$ and $\tilde{\rho}$ in (5.3). This amounts to ignoring terms of lower order, $o(m^{-1})$, for large $m$.

The derivatives involved in (5.3) have been evaluated in an unpublished report. Using these expression, we obtain

$$E(\hat{\tau} - \tilde{\tau})^2 \approx tr(\underset{\sim}{\Delta}'\underset{\sim}{V}\underset{\sim}{\Delta}\underset{\sim}{\Sigma}_{\rho uv}) \tag{5.4}$$

where $\underset{\sim}{\Sigma}_{\rho uv}$ is the $3 \times 3$ covariance matrix of $\tilde{\rho} - \rho$, $\tilde{\sigma}_u^2(\rho) - \sigma_u^2$, and $\tilde{\sigma}_v^2(\rho) - \sigma_v^2$, $\underset{\sim}{\Delta} = (\underset{\sim}{b}_1, \partial\underset{\sim}{b}/\partial\sigma_u^2, \partial\underset{\sim}{b}/\partial\sigma_v^2)'$ with $\underset{\sim}{b} = (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}$,

$$\partial\underset{\sim}{b}/\partial\sigma_u^2 = \big[\underset{\sim}{l}_2' + (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}\big]\underset{\sim}{R}\underset{\sim}{V}^{-1},$$

$$\partial\underset{\sim}{b}/\partial\sigma_v^2 = \big[\underset{\sim}{l}_1' + (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}\underset{\sim}{Z}\big]\underset{\sim}{Z}'\underset{\sim}{V}^{-1}$$

and

$$\begin{aligned} \underset{\sim}{b}_1 &= \left(\frac{d\tilde{\sigma}_v^2(\rho)}{d\rho}\,\underset{\sim}{l}_1'\underset{\sim}{Z}' + \frac{d\tilde{\sigma}_u^2(\rho)}{d\rho}\,\underset{\sim}{l}_2'\underset{\sim}{R} + \sigma_u^2\,\underset{\sim}{l}_2'\frac{d\underset{\sim}{R}}{d\rho}\right)\underset{\sim}{V}^{-1} \\ &\quad - (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}\left(\frac{d\tilde{\sigma}_u^2(\rho)}{d\rho}\,\underset{\sim}{R} + \sigma_u^2\frac{d\underset{\sim}{R}}{d\rho} \right. \\ &\quad \left. + \frac{d\tilde{\sigma}_v^2(\rho)}{d\rho}\,\underset{\sim}{Z}\,\underset{\sim}{Z}'\right)\underset{\sim}{V}^{-1}. \end{aligned}$$

Combining (5.1) and (5.4), we get a second order approximation to $MSE(\hat{\tau})$ as

$$MSE[\hat{\tau}] \approx g_1(\sigma_u^2, \sigma_v^2, \rho) + g_2(\sigma_u^2, \sigma_v^2, \rho) + g_3(\sigma_u^2, \sigma_v^2, \rho), \tag{5.5}$$

where

$$\begin{aligned} g_1(\sigma_u^2, \sigma_v^2, \rho) &= [\underset{\sim}{l} - \underset{\sim}{X}'\underset{\sim}{V}^{-1}(\sigma_v^2\underset{\sim}{Z}\,\underset{\sim}{l}_1 + \sigma_u^2\underset{\sim}{R}\,\underset{\sim}{l}_2)]' \times \\ &\quad (\underset{\sim}{X}'\underset{\sim}{V}^{-1}\underset{\sim}{X})^{-1}[\underset{\sim}{l} - \underset{\sim}{X}'\underset{\sim}{V}^{-1}(\sigma_v^2\underset{\sim}{Z}\,\underset{\sim}{l}_1 + \sigma_u^2\underset{\sim}{R}\,\underset{\sim}{l}_2)], \\ g_2(\sigma_u^2, \sigma_v^2, \rho) &= \sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{l}_1 + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R}\,\underset{\sim}{l}_2 - (\sigma_v^2\,\underset{\sim}{l}_1'\underset{\sim}{Z}' \\ &\quad + \sigma_u^2\,\underset{\sim}{l}_2'\underset{\sim}{R})\underset{\sim}{V}^{-1}(\sigma_v^2\underset{\sim}{Z}\,\underset{\sim}{l}_1 + \sigma_u^2\underset{\sim}{R}\,\underset{\sim}{l}_2), \\ g_3(\sigma_u^2, \sigma_v^2, \rho) &= tr(\underset{\sim}{\Delta}'\underset{\sim}{V}\underset{\sim}{\Delta}\underset{\sim}{\Sigma}_{\rho uv}). \end{aligned}$$

The neglected terms in (5.5) are of lower order, $o(m^{-1})$, for large $m$.

We now obtain expressions for the elements of the covariance matrix $\underset{\sim}{\Sigma}_{\rho u v}$ using the following well-known lemma on the covariance of two quadratic forms of normally distributed variables.

**Lemma.** *If $\underset{\sim}{y}$ is distributed as a multivariate normal vector with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Omega}$, then $Cov(\underset{\sim}{y}'\underset{\sim}{G}_1\underset{\sim}{y}, \underset{\sim}{y}'\underset{\sim}{G}_2\underset{\sim}{y}) = 2tr(\underset{\sim}{G}_1\underset{\sim}{\Omega}\underset{\sim}{G}_2\underset{\sim}{\Omega})$, where $\underset{\sim}{G}_1$ and $\underset{\sim}{G}_2$ are two symmetric matrices.*

It follows from (4.11) and (4.12) that

$$\tilde{\sigma}_u^2(\rho) = [(m-1)\underset{\sim}{T} - rank(\underset{\sim}{H}^{(1)})]^{-1}\underset{\sim}{a}'\underset{\sim}{C}_1\underset{\sim}{a} + C_u, \tag{5.6}$$

$$\begin{aligned}\tilde{\sigma}_v^2(\rho) = {} & c^{-1}[m - rank(\underset{\sim}{F})]^{-1}\underset{\sim}{a}'\underset{\sim}{C}_2\underset{\sim}{a} \\ & - c^{-1}[(m-1)\underset{\sim}{T} \\ & - rank(\underset{\sim}{H}^{(1)})]\underset{\sim}{a}'\underset{\sim}{C}_1\underset{\sim}{a} + C_v,\end{aligned} \tag{5.7}$$

where $\underset{\sim}{a} = \underset{\sim}{Z}\underset{\sim}{v} + \underset{\sim}{u} + \underset{\sim}{e}$ is normal with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{V}$,

$$\underset{\sim}{C}_1 = \underset{\sim}{C}'[\underset{\sim}{I} - \underset{\sim}{C}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{C}'\underset{\sim}{C}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{C}']\underset{\sim}{C}$$
$$\text{with } \underset{\sim}{C} = \text{diag}_i[(\underset{\sim}{I}_T - \underset{\sim}{D})\underset{\sim}{P}],$$
$$\underset{\sim}{C}_2 = \underset{\sim}{C}^{*\prime}[\underset{\sim}{I} - \underset{\sim}{C}^{*\prime}\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{C}^{*\prime}\underset{\sim}{C}^{*}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{C}^{*\prime}]\underset{\sim}{C}^{*}$$
$$\text{with } \underset{\sim}{C}^{*} = \text{diag}_i(c_i^{-1/2}\underset{\sim}{f}'\underset{\sim}{P}),$$

and $C_u$ and $C_v$ are two constants. We can now evaluate $Var[\tilde{\sigma}_u^2(\rho)]$, $Var[\tilde{\sigma}_v^2(\rho)]$ and $cov[\tilde{\sigma}_u^2(\rho),\ \tilde{\sigma}_v^2(\rho)]$ using (5.6) and (5.7) in the lemma.

Turning to the evaluation of remaining elements of $\underset{\sim}{\Sigma}_{\rho u v}$, we first used the usual ratio approximation for $\tilde{\rho} - \rho \approx \rho^* - \rho$ to get

$$\begin{aligned}\tilde{\rho} - \rho \approx {} & [m(T-2)(1-\rho)]^{-1} \\ & \times \underset{\sim}{a}'[(\underset{\sim}{\bar{G}}_2 - \underset{\sim}{\bar{G}}_3) - \rho(\underset{\sim}{\bar{G}}_1 - \underset{\sim}{\bar{G}}_2)]\underset{\sim}{a} + \text{constant},\end{aligned} \tag{5.8}$$

where $\underset{\sim}{\bar{G}}_i = \underset{\sim}{G}_i \otimes \underset{\sim}{I}_m$, $\underset{\sim}{G}_1$ is a $T \times T$ matrix with 1 for elements $(t,t)$ , $t = 1, \ldots, T-2$, 0 elsewhere, $\underset{\sim}{G}_2$ is a $T \times T$ matrix with 1/2 for elements $(t, t-1)$ and $(t-1, t)$, $t = 2, \ldots, T-1$, 0 elsewhere, and $\underset{\sim}{G}_3$ is a $T \times T$ matrix with 1/2 for elements $(t, t-2)$ and $(t-2, t)$, $t = 3, \ldots, T$, 0 elsewhere. Using (5.6), (5.7) and (5.8) in the Lemma, we can readily evaluated $var(\tilde{\rho})$, $cov(\tilde{\rho}, \tilde{\sigma}_v^2(\rho))$, and $cov(\tilde{\rho}, \tilde{\sigma}_u^2(\rho))$.

### 5.2. Second order approximation to estimator of MSE

Following Prasad and Rao (1990), it can be shown that

$$E[g_1(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_1(\sigma_u^2, \sigma_v^2, \rho) + g_3(\sigma_u^2, \sigma_v^2, \rho), \tag{5.9}$$

$$E[g_2(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_2(\sigma_u^2, \sigma_v^2, \hat{\rho}), \tag{5.10}$$

$$E[g_3(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho})] \approx g_3(\sigma_u^2, \sigma_v^2, \hat{\rho}). \tag{5.11}$$

Hence, it follows from (5.5), (5.9), (5.10) and (5.11) that a second order approximation to the estimator of $MSE(\hat{\tau})$ is given by

$$mse(\hat{\tau}) \approx g_1(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) + g_2(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) + 2g_3(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}).$$

The evaluation of $g_3(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho})$ involves the calculation of the derivatives $d\tilde{\sigma}_v^2(\rho)/d\rho$ and $d\tilde{\sigma}_u^2(\rho)/d\rho$ at the point $\rho = \hat{\rho}$. Analytical evaluation of these derivatives is difficult, but they can be easily evaluated numerically using the formulae for $\tilde{\sigma}_v^2(\rho)$ and $\tilde{\sigma}_u^2(\rho)$.

The jackknife method can also be used to estimate $\underset{\sim}{\Sigma}_{\rho u v}$ noting that the data sets across areas are assumed to be independent. Denote the estimators of $\sigma_u^2$, $\sigma_v^2$ and $\rho$ as $\hat{\sigma}_{u(-k)}^2$, $\hat{\sigma}_{v(-k)}^2$ and $\hat{\rho}_{(-k)}$ when the data set from the $k$-th area is deleted ($k = 1, \ldots, m$). Then, a jackknife estimator of $\underset{\sim}{\Sigma}_{\rho u v}$ is given by

$$\begin{aligned}\underset{\sim}{\hat{\Sigma}}_{\rho u v} = {} & \frac{k-1}{k}\sum_{k=1}^{m}[ \\ & (\hat{\rho}_{(-k)} - \hat{\rho}),\ (\hat{\sigma}_{u(-k)}^2 - \sigma_u^2),\ (\hat{\sigma}_{v(-k)}^2 - \sigma_v^2)]' \times \\ & [(\hat{\rho}_{(-k)} - \hat{\rho}), (\hat{\sigma}_{u(-k)}^2 - \sigma_u^2), (\hat{\sigma}_{v(-k)}^2 - \sigma_v^2)].\end{aligned}$$

## 6. HIERARCHICAL BAYES ANALYSIS

The hierarchical Bayes approach accounts for the uncertainty of prior parameters by assigning diffuse priors to such parameters. As a result, the posterior joint distribution of the $\theta_{it}$ given the data $\underset{\sim}{y}$ can be obtained, and the posterior mean of $\theta_{it}$ is employed as an estimator of $\theta_{it}$ with associated measure of uncertainty given by the posterior variance of $\theta_{it}$.

We use the following four-stage hierarchical model:

Stage 1. Conditional on $\theta_{it}$'s, $v_i$'s, $\underset{\sim}{\beta}$ , $r_1$ $(= 1/\sigma_u^2)$, $r_2$ $(= 1/\sigma_v^2)$ and $\rho$, $\underset{\sim}{y}_i \overset{ind}{\sim} N(\underset{\sim}{\theta}_i,\ \underset{\sim}{\Sigma}_i)$ where $\underset{\sim}{\theta}_i = (\theta_{i1}, \ldots, \theta_{iT})'$.

Stage 2. Conditional on $v_i$'s, $\underset{\sim}{\beta}$, $r_1$, $r_2$ and $\rho$, $\underset{\sim}{\theta}_i \overset{ind}{\sim} N(\underset{\sim}{X}_i\underset{\sim}{\beta} + v_i\underset{\sim}{1},\ r_1^{-1}\underset{\sim}{\Gamma})$.

Stage 3. Conditional on $\underset{\sim}{\beta}$, $r_1$, $r_2$ and $\rho$, $v_i \sim N(0, r_2^{-1}\underset{\sim}{I})$,

Stage 4. $\underset{\sim}{\beta}$, $r_1$, $r_2$ and $\rho$ are independently distributed with $\underset{\sim}{\beta} \sim uniform(R^p)$, $r_1 \sim Gamma(a_1/2, b_1/2)$, $r_2 \sim Gamma(a_2/2, b_2/2)$ and some suitable diffuse prior distribution on $\rho$, say $f(\rho)$.

They are almost two approaches to obtaining the posterior distribution of $\theta_{it}$ given $\underset{\sim}{y}$ and the posterior expectation and posterior variance of $\theta_{it}$. In the first approach, the posterior mean and the posterior covariance matrix of $\underset{\sim}{\theta} = (\underset{\sim}{\theta}'_1, \ldots, \underset{\sim}{\theta}'_m)'$ given $\underset{\sim}{y}$, $r_1$, $r_2$ and $\rho$ and the posterior distribution of $r_1$, $r_2$ and $\rho$ given $\underset{\sim}{y}$ are first obtained analytically, using the diffuse prior distributions assumed at stage 4 of the hierarchical model. Denoting these quantities by $\underset{\sim}{M}(\underset{\sim}{y}, r_1, r_2, \rho)$, $\underset{\sim}{V}(\underset{\sim}{y}, r_1, r_2, \rho)$ and $f(r_1, r_2, \rho \mid \underset{\sim}{y})$, we can compute the posterior mean and the posterior covariance matrix of $\underset{\sim}{\theta}$ as

$$E(\underset{\sim}{\theta} \mid \underset{\sim}{y}) = \int\!\!\int\!\!\int \underset{\sim}{M}(\underset{\sim}{y}, r_1, r_2, \rho) f(r_1, r_2, \rho \mid \underset{\sim}{y}) dr_1 dr_2 d\rho$$

and

$$\begin{aligned} Cov(\underset{\sim}{\theta} \mid \underset{\sim}{y}) = & \int\!\!\int\!\!\int \underset{\sim}{V}(\underset{\sim}{y}, r_1, r_2, \rho) f(r_1, r_2, \rho \mid \underset{\sim}{y}) dr_1 dr_2 d\rho \\ & + \int\!\!\int\!\!\int [\underset{\sim}{M}(\underset{\sim}{y}, r_1, r_2, \rho) - E(\underset{\sim}{\theta} \mid \underset{\sim}{y})][\underset{\sim}{M}(\underset{\sim}{y}, r_1, r_2, \rho) \\ & \qquad - E(\underset{\sim}{\theta} \mid \underset{\sim}{y})]' f(r_1, r_2, \rho \mid \underset{\sim}{y}) dr_1 dr_2 d\rho. \end{aligned}$$

However, closed-form expression for $E(\underset{\sim}{\theta} \mid \underset{\sim}{y})$ and $Cov(\underset{\sim}{\theta} \mid \underset{\sim}{y})$ cannot be obtained, and numerical integration becomes necessary. In the present case, we need to solve three-dimensional integrals numerically.

The second approach uses the currently popular Gibbs sampling. The desired posterior mean and posterior covariance matrix of $\underset{\sim}{\theta}$ given $\underset{\sim}{y}$ can be obtained through an iterative Monte-Carlo procedure by sampling from the full conditional distributions $f(\underset{\sim}{\theta} \mid \underset{\sim}{\beta}, \underset{\sim}{v}, r_1, r_2, \rho, \underset{\sim}{y})$, $f(\underset{\sim}{v} \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}, r_1, r_2, \rho, \underset{\sim}{y})$, $f(\underset{\sim}{\beta} \mid \underset{\sim}{\theta}, \underset{\sim}{v}, r_1, r_2, \rho, \underset{\sim}{y})$, $f(r_1 \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}, \underset{\sim}{v}, r_2, \rho, \underset{\sim}{y})$, $f(r_2 \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}, \underset{\sim}{v}, r_1, \rho, \underset{\sim}{y})$ and $f(\rho \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}, \underset{\sim}{v}, r_1, r_2, \underset{\sim}{y})$ alone which determine the joint distribution of $\underset{\sim}{\theta}, \underset{\sim}{v}, \underset{\sim}{\beta}, r_1, r_2, \rho$ conditional on $\underset{\sim}{y}$ (for example, see Gelfand and Smith, 1989). The desired full conditional distributions are obtained as follows:

(1) $\underset{\sim}{\theta} \mid \underset{\sim}{v}, \underset{\sim}{\beta}, r_1, r_2, \rho, \underset{\sim}{y} \sim N[(\underset{\sim}{\Sigma}^{-1} + r_1 \underset{\sim}{R}^{-1})^{-1} \times (\underset{\sim}{\Sigma}^{-1}\underset{\sim}{y} + r_1 \underset{\sim}{R}^{-1}(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{v})),\ (\underset{\sim}{\Sigma}^{-1} + r_1 \underset{\sim}{R}^{-1})^{-1}]$ (independent of $r_2$),

(2) $\underset{\sim}{v} \mid \underset{\sim}{\theta}, \underset{\sim}{\beta}, r_1, r_2, \rho, \underset{\sim}{y} \sim N[(r_1 \underset{\sim}{Z}' \underset{\sim}{R}^{-1} \underset{\sim}{Z} + r_2 \underset{\sim}{I})^{-1} r_1 \underset{\sim}{Z}' \underset{\sim}{R}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta}), (r_1 \underset{\sim}{Z}' \underset{\sim}{R}^{-1} \underset{\sim}{Z} + r_2^{-1} \underset{\sim}{I})^{-1}]$ (independent of $y$),

(3) $\underset{\sim}{\beta} \mid \underset{\sim}{\theta}, \underset{\sim}{v}, r_1, r_2, \rho, \underset{\sim}{y} \sim N[(\underset{\sim}{X}' \underset{\sim}{R}^{-1} \underset{\sim}{X})^{-1} \times \underset{\sim}{X}' \underset{\sim}{R}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{Z}\underset{\sim}{v}),\ (\underset{\sim}{X}' \underset{\sim}{R}^{-1} \underset{\sim}{X})^{-1}]$ (independent of $y$ and $r_2$),

(4) $r_1 \mid \underset{\sim}{\theta}, \underset{\sim}{v}, \underset{\sim}{\beta}, r_2, \rho, \underset{\sim}{y} \sim Gamma[\frac{1}{2}\{a_1 + (\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})' \underset{\sim}{R}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})\}, \frac{1}{2}(mT + b_1)]$ (independent of $y$ and $r_2$),

(5) $r_2 \mid \underset{\sim}{\theta}, \underset{\sim}{v}, \underset{\sim}{\beta}, r_1, \rho, \underset{\sim}{y} \sim Gamma[\frac{1}{2}(a_2 + \underset{\sim}{v}'\underset{\sim}{v}), \frac{1}{2}(m + b_2)]$ (independent of $y$, $\theta$, $r_1$),

(6) $\rho \mid \underset{\sim}{\theta}, \underset{\sim}{v}, \underset{\sim}{\beta}, r_1, r_2, \underset{\sim}{y} = [c(\underset{\sim}{\beta}, \underset{\sim}{\theta}, \underset{\sim}{v}, r_1)]^{-1} \times |\underset{\sim}{R}|^{-1/2} \exp[-\frac{r_1}{2}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})' \underset{\sim}{R}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})] f(\rho)$ (independent of $y$, $r_2$)

where

$$c(\underset{\sim}{\beta}, \underset{\sim}{\theta}, \underset{\sim}{v}, r_1) = \int |R|^{-1/2} \exp\{-\frac{r_1}{2}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})' \underset{\sim}{R}^{-1}(\underset{\sim}{\theta} - \underset{\sim}{X}\underset{\sim}{\beta} - \underset{\sim}{Z}\underset{\sim}{v})\} f(\rho) d\rho.$$

Random variates can be readily generated from the conditional distributions (1) to (5) since only normal and gamma distributions are involved. The conditional distribution (6) does not have a closed form, but can be generated using rejection sampling without evaluating the integral $c(\underset{\sim}{\beta}, \underset{\sim}{\theta}, \underset{\sim}{v}, r_1)$ (for example, see Zeger and Karim, 1991).

## REFERENCES


Anderson, T. W. and Hsiao, C. (1981). Formulation and estimation of dynamic models using panel data. *J. Econometrics*, **18**, 67–82.

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **80**, 28–36.

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

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

Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: application to small area estimation. *Ann. Statist.*, **19**, 1748–70.

Datta, G. S., Fay, R. E. and Ghosh, M. (1991). Hierarchical and empirical Bayes analysis in small area estimation. *Proceedings of the Annual Research Conference, U.S. Bureau of the Census*, 63–78.

Drew, J. D., Singh, M. P. and Choudhry, G. H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Survey Methology*, **8**, 17–47.

Ericksen, E. P. (1974). A regression method for estimating population of local areas. *J. Amer. Statist. Assoc.*, **69**, 867–75.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269–77.

Fuller, W. A. and Battese, G. E. (1973). Transformations for estimation of linear models with nested error structures. *J. Amer. Statist. Assoc.*, **68**, 626–32.

Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.*, **82**, 1153–62.

Gelfand, A. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimates. *Proc. Amer. Statist. Assoc., Soc. Statist. Sec.*, 33-6.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrica*, **31**, 423–47.

Jones, R. G. (1980). Best linear unbiased estimators for repeated surveys. *J. R. Statist. Soc. Ser. B*, **42**, 221-6.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T. (1985). *The Theory and Practice of Econometrics* (2nd Ed.), New York: Wiley.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.*, **79**, 853–62.

Pantula, S. G. and Pollock, K. H. (1985). Nested analysis of variance with autocorrelated errors. *Biometrica*, **41**, 909–20.

Pfeffermann, D. and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methology*, **16**, 217–37.

Prasad, N. G. N. and Rao, J. N. K. (1986). On the estimation of mean squared error of small area predictors. In Tech. Report **97**, Carleton University, Laboratory for Research in Statistics and Probability, Ottawa.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163–71.

Purcell, N. J. and Kish, L. (1980). Postcensal estimates for local areas (or domains). *Int. Statist. Rev.*, **48**,3–18.

Rao, J. N. K. (1986). Synthetic estimators, SPREE and best model-based predictors of small area means. In Tech. Report **97**, Carleton University, Laboratory for Research in Statistics and Probability, Ottawa.

Särndal, C. E. and Hidiroglou, M. A. (1989). Small domain estimation: a conditional analysis. *J. Amer. Statist. Assoc.*, **84**, 266–75.

Scott, A. J. and Smith, T. M. P. (1974). Analysis of repeated surveys using time series methods. *J. Amer. Statist. Assoc.*, **69**, 674–8.

Singh, A. C., Mantel, H. and Thomas, B. W. (1991). Time series generalizations of Fay-Herriot estimator for small areas. *Proc. Amer. Statist. Assoc., Sec. Survey Res. Methods.*

Tilles, R. B. (1989). A Kalman filter approach to labor force estimation using survey data. *Proc. Sec. Survey Res. Methods, Amer. Statist. Assoc.*, 11–25.

Zeger, S. L. and Karim (1991). Generalized linear models with random effects: a Gibbs sampling approach. *J. Amer. Statist. Assoc.*, **86**, 79–86.