Survey Methodology, June 2003
Vol. 29, No. 1, pp. 25-32
Statistics Canada, Catalogue No. 12-001-XPB

Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: A Hierarchical Bayes Approach

Yong You, J.N.K. Rao and Jack Gambino [1]

Abstract

The Canadian Labour Force Survey (LFS) produces monthly direct estimates of the unemployment rate at national and provincial levels. The LFS also releases unemployment estimates for sub-provincial areas such as Census Metropolitan Areas (CMAs) and Census Agglomerations (CAs). However, for some sub-provincial areas, the direct estimates are not very reliable since the sample size in some areas is quite small. In this paper, a cross-sectional and time-series model is used to borrow strength across areas and time periods to produce model-based unemployment rate estimates for CMAs and CAs. This model is a generalization of a widely used cross-sectional model in small area estimation and includes a random walk or AR(1) model for the random time component. Monthly Employment Insurance (EI) beneficiary data at the CMA or CA level are used as auxiliary covariates in the model. A hierarchical Bayes (HB) approach is employed and the Gibbs sampler is used to generate samples from the joint posterior distribution. Rao-Blackwellized estimators are obtained for the posterior means and posterior variances of the CMA/CA-level unemployment rates. The HB method smooths the survey estimates and leads to substantial reduction in standard errors. Bayesian model fitting is also investigated based on posterior predictive distributions.

Key Words: Gibbs sampling; Hierarchical Bayes; Labour Force Survey; Small area estimation; Unemployment rate.

1. Introduction

The unemployment rate is generally viewed as a key indicator of economic performance. In Canada, although provincial and national estimates get the most media attention, subprovincial estimates of the unemployment rate are also very important. They are used by the Employment Insurance (EI) program to determine the rules used to administer the program. In addition, the unemployment rates for Census Metropolitan Areas (CMAs, *i.e.*, cities with population more than 100,000) and Census Agglomerations (CAs, *i.e.*, other urban centres) receive close scrutiny at local levels. However, many CAs do not have a large enough sample to produce adequate direct estimates. Our objective in this paper is to obtain model-based estimators that lead to improvement over the direct estimator which is based solely on the sample falling in a given CMA or CA in a given month. For convenience, since CMAs are also CAs, we will refer to both CMAs and CAs as CAs.

In Canada, unemployment rates are produced by the Labour Force Survey (LFS). The LFS is a monthly survey of 53,000 households selected using a stratified, multistage design. Each month, one-sixth of the sample is replaced. Thus five-sixths of the sample is common between two consecutive months. This sample overlap induces correlations which can be exploited to produce better estimates by any method which borrows strength across time. For a detailed description of the LFS design, see Gambino, Singh, Dufour, Kennedy and Lindeyer (1998).

Traditional small area estimators borrow strength either from similar small areas or from the same area across time, but not both. In recent years, several approaches to borrowing strength simultaneously across both space and time have been developed. Estimators based on the approach developed by Rao and Yu (1994), such as those in Ghosh, Nangia and Kim (1996), Datta, Lahiri, Maiti and Lu (1999) and in this paper, successfully exploit the two dimensions simultaneously to produce improved estimates with desirable properties for small areas. Datta *et al.* (1999) applied their model to long time series ($T = 48$ months) data across small areas from the U.S. Current Population Survey. In this paper, we apply a similar model to the Canadian LFS. Unlike Datta *et al.* (1999), we have used short time series data across small areas. Therefore, our model does not contain seasonal parameters. This reduces substantially the number of parameters that need to be estimated; details on modelling and analysis are given in section 2 and section 4. Despite this simplification, we obtain both an adequate model fit and large reductions in the coefficients of variation (CVs) of the small area estimators of the unemployment rate. The CV reduction is due in part to our approach to computing covariance matrices, which uses smoothed CVs and lag correlations to obtain smoothed estimates of the sampling covariance matrices of the direct LFS estimators.

In section 2, we present the model, which borrows strength across small areas and time periods. In section 3, the model is placed in a hierarchical Bayes (HB) framework. The use of Gibbs sampling to generate samples from the joint posterior distribution is described and the corresponding HB estimators are obtained. The HB method is applied to the LFS data in section 4 to produce

1. Yong You, Jack Gambino, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6; J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, K1S 5B6.

unemployment rates for CAs. Specifically, subsections 4.2 and 4.3 present model selection and model fit analysis. Subsection 4.4 presents model-based estimates for the small area (CA) unemployment rates and the CV comparisons. Finally some concluding remarks are given in section 5.

2. Cross-Sectional and Time Series Models

Let y_{it} denote the direct LFS estimate of θ_{it}, the true unemployment rate of the i^{th} CA (small area) at time *t*, for $i = 1, ..., m, \ t = 1, ..., T$, where m is the total number of CAs and T is the (current) time of interest. We assume that

$$y_{it} = \theta_{it} + e_{it}, \ \ i = 1, ..., m, \ t = 1, ..., T, \qquad (1)$$

where e_{it}'s are sampling errors. Let $y_i = (y_{i1}, ..., y_{iT})'$, $\theta_i = (\theta_{i1}, ..., \theta_{iT})'$, and $e_i = (e_{i1}, ..., e_{iT})'$. Then e_i is a vector of sampling errors for the i^{th} CA. In the LFS design, the CAs are treated as strata. Thus the sampling vectors e_i are uncorrelated between areas (CAs). Because of the LFS sample rotation pattern, there is substantial sample overlap over short time periods within each area. As a result, the correlation between e_{it} and $e_{is} (t \neq s)$ has to be taken into account. It is customary to assume that e_i follows a multivariate normal distribution with mean vector 0 and covariance matrix Σ_i, *i.e.*, $e_i \sim N(0, \Sigma_i)$. Using (1), we have

$$y_i \sim N(\theta_i, \Sigma_i), \ \ i = 1, ..., m. \qquad (2)$$

Thus y_i is design-unbiased for θ_i. The variance-covariance matrix Σ_i in model (2) is assumed to be known. The assumption of normality and known Σ_i in model (2) is the customary practice in model-based small area estimation (see, for example, Fay and Herriot 1979; Ghosh and Rao 1994; Datta *et al.* 1999; Rao 1999). In this paper, we follow the customary approach and treat Σ_i as known. Specification of Σ_i may not be easy in practice. We use a smoothed estimator of Σ_i in the model, and then treat it as the true Σ_i. More details on constructing a smoothed estimator of Σ_i in the context of the LFS are given in section 4. Pfeffermann, Feder and Signorelli (1998) proposed a simple method of estimating the autocorrelations of sampling errors for rotating-panel designs, such as the Canadian LFS. It would be useful to study the feasibility of this approach in our context.

To borrow strength across small areas and time periods, we model the true unemployment rate θ_{it} by a linear regression model with random effects through auxiliary variables x_{it}. We assume that

$$\theta_{it} = x_{it}'\beta + v_i + u_{it}, \ \ i = 1, ..., m, \ t = 1, ..., T, \qquad (3)$$

where $x_{it} = (x_{it1}, ..., x_{itp})'$ is the vector of area level auxiliary data for the i^{th} CA at time *t*; β is a vector of regression parameters of length *p*; v_i is a random area effect with $v_i \sim \text{iid } N(0, \sigma_v^2)$; u_{it} is a random time component. For a given area *i*, Datta *et al.* (1999) assumed that u_{it} follows a random walk process over time period $t = 1, ..., T$, that is,

$$u_{it} = u_{i, t-1} + \varepsilon_{it}, \ \ i = 1, ..., m, \ t = 2, ..., T, \qquad (4)$$

where $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$. Then $\text{cov}(u_{it}, u_{is}) = \min(t, s)\sigma_\varepsilon^2$. Also $\{v_i\}$, $\{\varepsilon_{it}\}$ and $\{e_i\}$ are assumed to be mutually independent. The regression parameter β and the variance components σ_v^2 and σ_ε^2 are unknown in the model. Rao and Yu (1994) used a stationary autoregressive model, AR(1), for u_{it}, that is, $u_{it} = \rho u_{i, t-1} + \varepsilon_{it}$, and $|\rho| < 1$. Datta *et al.* (1999) included month and year effects as seasonal effects for θ_{it} in (3) using a long time series ($T = 48$ months) in their analysis. In our modelling, we intend to study the effects of borrowing strength across areas and over time using short time series data instead of long time series data. In particular, based on the Canadian LFS design's six-month rotation cycle, we used only 6 months of data for smoothing; see section 4 for details. Thus the linking model (3) is simpler than Datta *et al.* (1999)'s model. This simplification is likely to reduce the instability in the smoothed covariance matrix Σ_i.

Arranging the data $\{y_{it}\}$ as a vector $y = (y_1', ..., y_m')'$ with $y_i = (y_{i1}, ..., y_{iT})'$, we can write models (2), (3) and (4) in matrix form as

$$y_i = X_i\beta + 1_T v_i + u_i + e_i, \ \ i = 1, ..., m, \qquad (5)$$

where $X_i' = (x_{i1}, ..., x_{iT})$, $u_i' = (u_{i1}, ..., u_{iT})$, and 1_T is a $T \times 1$ vector of 1's. Model (5) is a special case of a general linear mixed effects model. It also extends the well-known Fay-Herriot model (Fay and Herriot 1979) by borrowing strength across both areas and time.

For comparison, we also considered the Fay-Herriot model for the time points $t = 1, ..., T$ in our data analysis. The model at time point t is given as

$$y_{it} = \theta_{it} + e_{it}, \ \ i = 1, ..., m, \qquad (6)$$

and

$$\theta_{it} = x_{it}'\beta_t + v_{it}, \ \ i = 1, ..., m, \qquad (7)$$

where the sampling errors $e_{it} \sim \text{ind } N(0, \sigma_{it}^2)$ and the area random effects $v_{it} \sim \text{iid } N(0, \sigma_{vt}^2)$ for each time point t and independent of $v_{it'}$, $t' \neq t$. The sampling variances σ_{it}^2 are assumed to be known (smoothed estimates) and σ_{vt}^2 is unknown. The Fay-Herriot model combines cross-sectional information at each t for estimating θ_{it}, but does not borrow strength over the past time periods.

We are interested in obtaining a model-based estimator of θ_{it}, in particular, for the current time $t = T$. Datta, Lahiri and Maiti (2002) and You (1999) obtained two-stage estimators for θ_{iT} and MSE approximations for the estimators through the empirical best linear unbiased prediction (EBLUP) approach. In this paper, we study both

AR(1) and random walk models on u_{it}'s, under a complete HB approach using the Gibbs sampling method.

3. Hierarchical Bayes Analysis

In this section, we apply the hierarchical Bayes approach to the cross-sectional and time series model given by (2), (3) and (4) and the Fay-Herriot model given by (6) and (7). Estimates of the posterior mean and posterior variance of the small area means, θ_{iT}, are obtained using the Gibbs sampling method.

3.1 The Hierarchical Bayes Model

We now present the cross-sectional and time series model in a hierarchical Bayes framework as follows:

- Conditional on the parameters $\theta_i = (\theta_{i1}, ..., \theta_{iT})'$, $[y_i \mid \theta_i] \sim \text{ind } N(\theta_i, \Sigma_i)$;
- Conditional on the parameters β, u_{it} and σ_v^2, $[\theta_{it} \mid \beta, u_{it}, \sigma_v^2] \sim \text{ind } N(x_{it}'\beta + \rho u_{it}, \sigma_v^2)$;
- Conditional on the parameters $u_{i,t-1}$ and σ_ε^2, $[u_{it} \mid u_{i,t-1}, \sigma_\varepsilon^2] \sim \text{ind } N(\rho u_{i,t-1}, \sigma_\varepsilon^2)$;

Marginally β, σ_v^2 and σ_ε^2 are mutually independent with priors given as $\beta \propto 1$, $\sigma_v^2 \sim IG(a_1, b_1)$, and $\sigma_\varepsilon^2 \sim IG(a_2, b_2)$, where *IG* denotes an inverted gamma distribution and a_1, b_1, a_2, b_2 are known positive constants and usually set to be very small to reflect our vague knowledge about σ_v^2 and σ_ε^2. For the random walk model, we take $\rho = 1$ and for the AR(1) model, $|\rho| < 1$ and ρ is assumed to be known.

We are interested in estimating θ_i, and in particular, the current unemployment rate θ_{iT}. In the HB analysis, θ_{iT} is estimated by its posterior mean $E(\theta_{iT} \mid y)$ and the uncertainty associated with the estimator is measured by the posterior variance $V(\theta_{iT} \mid y)$. We use Gibbs sampling (Gelfand and Smith 1990; Gelman and Rubin 1992) to obtain the posterior mean and the posterior variance of θ_{iT}.

Similarly, the Fay-Herriot model (6)–(7) can be expressed as:

- Conditional on the parameters θ_{it}, $[y_{it} \mid \theta_{it}] \sim \text{ind } N(\theta_{it}, \sigma_{it}^2)$;
- Conditional on the parameters β_t, and σ_v^2, $[\theta_{it} \mid \beta_t, \sigma_{vt}^2] \sim \text{ind } N(x_{it}'\beta_t, \sigma_{vt}^2)$;

Marginally β_t and σ_{vt}^2 are mutually independent with priors given as $\beta_t \propto 1$, $\sigma_{v_t}^2 \sim IG(a_t, b_t)$.

3.2 Gibbs Sampling Method

The Gibbs sampling method is an iterative Markov chain Monte Carlo sampling method to simulate samples from a joint distribution of random variables by sampling from low dimensional densities and to make inferences about the joint and marginal distributions (Gelfand and Smith 1990). The most prominent application is for inference within a Bayesian framework. In Bayesian inference one is interested in the posterior distribution of the parameters. Assume that $y_i \mid \theta$ has conditional density $f(y_i \mid \theta)$ for $i = 1, ..., n$ and that the prior information about $\theta = (\theta_1, ..., \theta_k)'$ is summarized by a prior density $\pi(\theta)$. Let $\pi(\theta \mid y)$ denote the posterior density of θ given the data $y = (y_1, ..., y_n)'$. It may be difficult to sample from $\pi(\theta \mid y)$ directly in practice due to the high dimensional integration with respect to θ. However, one can use the Gibbs sampler to construct a Markov chain $\{\theta^{(g)} = (\theta_1^{(g)}, ..., \theta_k^{(g)})'\}$ with $\pi(\theta \mid y)$ as the limiting distribution. For illustration, let $\theta = (\theta_1, \theta_2)'$. Starting with an initial set of values $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})'$, we generate $\theta^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)})'$ by sampling $\theta_1^{(g)}$ from $\pi(\theta_1 \mid \theta_2^{(g-1)}, y)$ and $\theta_2^{(g)}$ from $\pi(\theta_2 \mid \theta_1^{(g)}, y)$. Under certain regularity conditions, $\theta^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)})'$ converges in distribution to $\pi(\theta \mid y)$ as $g \to \infty$. Marginal inferences about $\pi(\theta_i \mid y)$ can be based on the marginal samples $\{\theta_i^{(g+k)}; k = 1, 2, ...\}$ for large g.

For the hierarchical Bayes models in section 3.1, to implement the Gibbs sampler we need to generate samples from the full conditional distributions of the parameters β, σ_v^2 and σ_ε^2, u_{it} and θ_i. These conditional distributions are given in Appendix A.1. All the full conditional distributions in the Appendix are standard normal or inverted gamma distributions that can be easily sampled.

3.3 Posterior Estimation

To implement Gibbs sampling, we follow the recommendation of Gelman and Rubin (1992) and independently run $L (L > 2)$ parallel chains, each of length *2d*. The first *d* iterations of each chain are deleted. After *d* iterations, all the subsequent iterates are retained for calculating the posterior means and posterior variances, as well as for monitoring the convergence of the Gibbs sampler. The convergence monitoring is discussed in section 4.

We use the Rao-Blackwellization approach to obtain estimators for the posterior mean and the posterior variance of interest. The Rao-Blackwellization can substantially reduce the simulation errors compared to naive estimates based on the simulated samples (Gelfand and Smith 1991; You and Rao 2000). For the cross-sectional and time series model, the Rao-Blackwellized estimates of $E(\theta_i \mid y)$ and $V(\theta_i \mid y)$ are obtained as

$$\hat{E}(\theta_i \mid y) = \sum_{l=1}^{L} \sum_{k=d+1}^{2d} \left[(\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1}) \times (\Sigma_i^{-1} y_i + \sigma_v^{-2(lk)} (X_i \beta^{(lk)} + u_i^{(lk)})) \right] \Big/ (Ld)$$

and

$$\hat{V}(\theta_i \mid y) = \sum_{l=1}^{L}\sum_{k=d+1}^{2d}(\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1})/(Ld)$$

$$+ \sum_{l=1}^{L}\sum_{k=d+1}^{2d}$$

$$\left[(\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1})^{-1} \times (\Sigma_i^{-1} y_{it} + \sigma_v^{-2(lk)}(X_i \beta^{(lk)} + u_i^{(lk)}))\right]$$

$$\times \left[\begin{array}{c}(\Sigma_i^{-1} y_i + \sigma_v^{-2(lk)}(X_i \beta^{(lk)} + u_i^{(lk)}))' \\ \times (\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1})^{-1}\end{array}\right] \Big/ (Ld)$$

$$- \left[\begin{array}{c}\sum_{l=1}^{L}\sum_{k=d+1}^{2d}(\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1})^{-1} \\ \times (\Sigma_i^{-1} y_i + \sigma_v^{-2(lk)}(X_i \beta^{(lk)} + u_i^{(lk)}))\end{array}\right]$$

$$\times \left[\begin{array}{c}\sum_{l=1}^{L}\sum_{k=d+1}^{2d}(\sigma_v^{-2(lk)} I_T + \Sigma_i^{-1})^{-1} \\ \times (\Sigma_i^{-1} y_i + \sigma_v^{-2(lk)}(X_i \beta^{(lk)} + u_i^{(lk)}))\end{array}\right]' \Big/ (Ld)^2,$$

where $\{\beta^{(lk)}, \sigma_v^{2(lk)}, u_i^{(lk)}; k = d+1, ..., 2d, l = 1, ..., L\}$ are the samples generated from the Gibbs sampler and I_T is the identity matrix of order T. Thus by using Gibbs sampling, we can estimate the current time small area mean θ_{iT} and the small area means θ_{it} for the past time periods $t = 1, ..., T-1$ simultaneously for each area. The posterior covariance matrix estimate $\hat{V}(\theta_i \mid y)$ also provides an estimate of the posterior covariance of θ_{it} and θ_{is} for $t \neq s = 1, ..., T$.

Under the Fay-Herriot model, letting $y_T = (y_{1T}, ..., y_{mT})'$ denote the current time cross-sectional data and using the conditional distributions given in Appendix A.2, we can similarly obtain the Rao-Blackwellized estimators of $E(\theta_{iT} \mid y_T)$ and $V(\theta_{iT} \mid y_T)$:

$$\hat{E}(\theta_{iT} \mid y_T) = \sum_{l=1}^{L}\sum_{k=d+1}^{2d}$$

$$[(1 - r_{iT}^{(lk)})\, y_{iT} + r_{iT}^{(lk)}\, x_{iT}'\, \beta_T^{(lk)}]/(Ld)$$

and

$$\hat{V}(\theta_{iT} \mid y_T) = \sum_{l=1}^{L}\sum_{k=d+1}^{2d}[\sigma_{iT}^2 (1 - r_{iT}^{(lk)})]/(Ld)$$

$$+ \sum_{l=1}^{L}\sum_{k=d+1}^{2d}[(1 - r_{iT}^{(lk)})\, y_{iT} + r_{iT}^{(lk)}\, x_{iT}'\, \beta_T^{(lk)}]^2/(Ld)$$

$$- \left\{\sum_{l=1}^{L}\sum_{k=d+1}^{2d}[(1 - r_{iT}^{(lk)}) y_{iT} + r_{iT}^{(lk)}\, x_{iT}'\, \beta_T{}^{(lk)}]\right\}^2 \Big/ (Ld)^2,$$

where $r_{iT}^{(lk)} = \sigma_{iT}^2/(\sigma_{iT}^2 + \sigma_v^{2(lk)})$. Note that $E(\theta_{iT} \mid y_T)$ and $V(\theta_{iT} \mid y_T)$ use only the cross-sectional data at $t = T$. As a result, $E(\theta_{iT} \mid y_T)$ will be less efficient than the HB estimator $E(\theta_{iT} \mid y_T)$ based on all the data; see section 4.

4. Application to the LFS

4.1 Data Description and Implementation

We used the 1999 LFS unemployment estimates, y_{it}, in our HB analysis. There are 64 CAs across Canada. Employment Insurance (EI) beneficiary rates were used as auxiliary data, x_{it}, in the model. But the EI beneficiary data were available for only 62 CAs. So we included only those $m = 62$ CAs in the model. Within each CA, we considered six consecutive monthly estimates y_{it} from January 1999 to June 1999, so that $T = 6$ and the parameter of interest θ_{iT} is the true unemployment rate for area *i* in June, 1999. The reason that we only used six months of data is that the LFS sample rotation is based on a six-month cycle. Each month, one sixth of the LFS sample is replaced. Thus after six months, the correlation between estimates is very weak. The one-month lag correlation coefficient is about 0.48, and the lag correlation coefficients decrease as the lag increases. Figure 1 shows the estimated (smoothed) lag correlation coefficients for the LFS unemployment rate estimates. It is clear that after 6 months the lag correlation coefficients are all below 0.1.

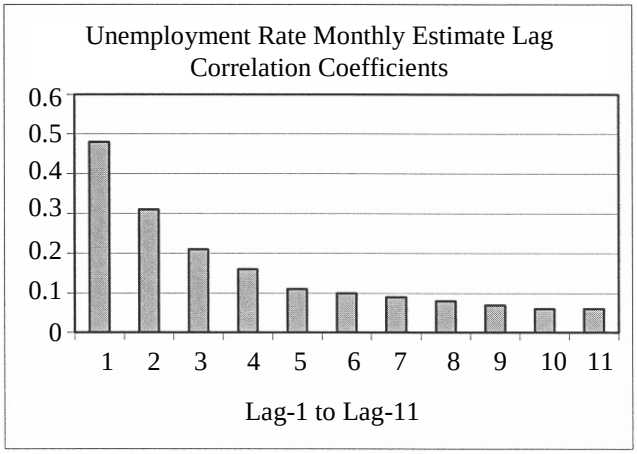

Figure 1. LFS unemployment rate lag correlation coefficients.

To obtain a smoothed estimate of the sampling covariance matrix Σ_i used in the model, we first computed the average coefficient of variation (CV) for each CA over time (12 months in this study), denoted as $\overline{CV}_i$, and the average lag correlation coefficients over time and all CAs. By using these smoothed CVs and lag correlation coefficients, we obtained a smoothed estimate $\hat{\Sigma}_i$ with diagonal elements $\hat{\sigma}_{itt} = (\overline{CV}_i)^2 y_{it}^2$ and off-diagonal elements equal to $\hat{\sigma}_{its} = \bar{\rho}_{|t-s|}(\hat{\sigma}_{itt}\hat{\sigma}_{iss})^{1/2}$ and treated $\hat{\Sigma}_i$ as the true Σ_i, where $\bar{\rho}_{|t-s|}$ is the average lag correlation coefficient of lag $|t-s|$. Our study found that using the smoothed estimate of Σ_i in the model can significantly improve the estimates in terms of CV reduction.

To implement the Gibbs sampling, we considered $L = 10$ parallel runs, each of length $2d = 2{,}000$. The first $d = 1{,}000$ "burn-in" iterations were deleted. To monitor the convergence of the Gibbs sampler, for the parameters of interest $\theta_{iT}\,(i = 1, ..., m)$, we followed the method of

Gelman and Rubin (1992) involving the following steps: For each θ_{iT}, let $\theta_{iT}^{(lk)}$ denote the k^{th} simulated value in the l^{th} chain, $k = d+1, ..., 2d$, $l = 1, ..., L$. In the first step, compute the overall mean

$$\overline{\theta}_{iT} = \sum_{l=1}^{L} \sum_{k=d+1}^{2d} \theta_{iT}^{(lk)} / (Ld)$$

and the within sequence mean

$$\overline{\theta}_{iT}^{(l)} = \sum_{k=d+1}^{2d} \theta_{iT}^{(lk)} / d, \; l = 1, ..., L.$$

Then compute B_{iT}/d, the variance between the L sequence means as $B_{iT}/d = \sum_{l=1}^{L} (\overline{\theta}_{iT} - \overline{\theta}_{iT}^{(l)})^2 / (L-1)$. In the second step, calculate W_{iT}, the average of the L within sequence variances, s_{iTl}^2, each based on $(d-1)$ degrees of freedom; that is, $W_{iT} = \sum_{l=1}^{L} s_{iTl}^2 / L$. In the third step, calculate $s_{iT}^2 = (d-1)W_{iT}/d + B_{iT}/d$ and $V_{iT} = s_{iT}^2 + B_{iT}/(Ld)$. In the last step, find the potential scale reduction factors $\hat{R}_{iT} = V_{iT}/W_{iT}$ $(i = 1, ..., m)$. If these potential scale reduction factors are near 1 for all of the scalar estimands θ_{iT} of interest, then this suggests that the desired convergence is achieved by the Gibbs sampler. In our study, the Gibbs sampler converged very well in terms of the values of $\hat{R}_{iT}$.

4.2 Model Selection

In this section, we compare the proposed model with the Rao and Yu (1994) AR(1) time component model. A number of methods for model comparison in a Bayesian framework have been developed, and several are implemented in the well-known BUGS program (see Spiegelhalter, Thomas, Best and Gilks 1996). In practice, when there is more than one model of interest, Bayesian model selection or model choice can be made on the basis of a Bayes factor, which is difficulty to calculate directly. Alternative strategies for model selection involve the predictive likelihood and predictive log-likelihood. In particular, Dempster (1974) suggested examining the posterior distribution of the log-likelihood of the observed data. The quantities of the posterior distribution of the log-likelihood may be obtained from the predictive posterior distribution of the deviance, $-2\log f(y|\theta)$. The posterior deviance is straightforward to estimate using the Gibbs sampling output since it is the expectation of $-2\log f(y|\theta)$ over the posterior $\pi(\theta|y)$. For non-hierarchical models, the minimum feasible value of $-2\log f(y|\theta)$ is the traditional deviance statistic. For hierarchical models, the minimum of the deviance is likely to be very poorly estimated by the sample minimum, and the mean is a more reasonable measure (Karim and Zeger 1992; Gilks, Wang, Yvonnet and Coursagt 1993). For the AR(1) time component model, we considered two choices of ρ: $\rho = 0.75$ and $\rho = 0.5$. We calculated the log-likelihood at each iteration of the Gibbs sampler. Then we obtained the mean of the predictive posterior deviance: 1,311.5 for the proposed model, 1,372.8 for the AR(1) with $\rho = 0.5$ and 1,358.3 for the AR(1) with $\rho = 0.75$. Thus, the deviance measure suggests that the random walk model on u_{it}'s provides a slightly better fit to the data than the AR(1) model.

For model comparison, we also computed the divergence measure of Laud and Ibrahim (1995) based on the posterior predictive distribution. Let θ^* represent a draw from the posterior distribution of θ given y, and let y^* represent a draw from $f(y|\theta^*)$. Then, marginally y^* is a sample from the posterior predictive distribution $f(y|y_{\text{obs}})$, where y_{obs} represents the observed data. The expected divergence measure of Laud and Ibrahim (1995) is given by $d(y^*, y_{\text{obs}}) = E(k^{-1} \| y^* - y_{\text{obs}} \|^2 | y_{\text{obs}})$, where k is the dimension of y_{obs}. Between two models, we prefer a model that yields a smaller value of this measure. As in Datta, Day and Maiti (1998) and Datta *et al.* (1999), we approximated the divergence measure $d(y^*, y_{\text{obs}})$ by using the simulated samples from the posterior predictive distribution. Using the Gibbs sampling output, we obtained a divergence measure of 13.36 for the proposed model, 14.62 for the AR(1) with $\rho = 0.5$ and 14.52 for the AR(1) with $\rho = 0.75$. Thus the divergence measure also suggests a slightly better fit of the random walk model compared to the AR(1) model.

It should be mentioned that the posterior deviance and the divergence measure are intended for comparing two or more alternative models. After selecting a model, we need to check if the selected model fits the data, which we turn to next.

4.3 Test of Model Fit

To check the overall fit of the proposed model, we used the method of posterior predictive p values (Meng 1994; Gelman, Carlin, Stern and Rubin 1995). In this approach, simulated values of a suitable discrepancy measure are generated from the posterior predictive distribution and then compared to the corresponding measure for the observed data. More precisely, let $T(y, \theta)$ be a discrepancy measure depending on the data y and the parameter θ. The posterior predictive p value is defined as

$$p = \text{prob}(T(y^*, \theta) > T(y_{\text{obs}}, \theta) | y_{\text{obs}}),$$

where y^* is a sample from the posterior predictive distribution $f(y|y_{\text{obs}})$. Note that the probability is with respect to the posterior distribution of θ given the observed data. This is a natural extension of the usual p value in a Bayesian context. If a model fits the observed data, then the two values of the discrepancy measure are similar. In other words, if the given model adequately fits the observed data, then $T(y_{\text{obs}}, \theta)$, should be near the central part of the histogram of the $T(y^*, \theta)$ values if y^* is generated repeatedly from the posterior predictive distribution. Consequently, the posterior predictive p value is expected to be near 0.5 if the model adequately fits the data. Extreme p values (near 0 or 1) suggest poor fit. The p value is self-contained in the sense that it is computed without regard to an alternative model.

Computing the p value is relatively easy using the simulated values of θ^* from the Gibbs sampler. For each

simulated value θ^*, we can simulate y^* from the model and compute $T(y^*, \theta^*)$ and $T(y_{\text{obs}}, \theta^*)$. Then the p value is approximated by the proportion of times $T(y^*, \theta^*)$ exceeds $T(y_{\text{obs}}, \theta^*)$. For the cross-sectional and time series model, the discrepancy measure used for overall fit is given by $d(y, \theta) = \sum_{i=1}^{m}(y_i - \theta_i)' \Sigma_i^{-1}(y_i - \theta_i)$. Datta *et al.* (1999) used the same discrepancy measure. We computed the p value by combining the simulated θ^* and y^* from all 10 parallel runs. We obtained a p value equal to 0.615. Thus we have no indication of lack of overall model fit for the random walk time series and cross-sectional model.

For the Fay-Herriot model that uses only the current cross-sectional data, an approximate discrepancy measure is given by

$$d_{\text{FH}}(y_T, \theta_T) = \sum_{i=1}^{m}(y_{iT} - \theta_{iT})^2 / \sigma_{iT}^2,$$

where $\theta_T = (\theta_{1T}, ..., \theta_{mT})'$. In this case, the estimated p value is about 0.587, indicating a good fit of the Fay-Herriot model for the current cross-sectional data only. However, the associated HB estimates are substantially less efficient compared to the HB estimates based on the proposed cross-sectional and time series model that borrows strength across regions and over time simultaneously; see Figures 3 and 4.

A limitation of the posterior predictive p value is that it makes "double use" of the observed data, y_{obs}, first to generate samples from $f(y \mid y_{\text{obs}})$ and then to compute the p value. This double use of the data can induce unnatural behaviour, as demonstrated by Bayarri and Berger (2000). To avoid double use of the data, Bayarri and Berger (2000) proposed two alternative p–measures, named the partial posterior predictive p value and the conditional predictive p value. These measures, however, seem to be more difficult to implement than the posterior predictive p value, especially for a complex model like the time series and cross-sectional small area model.

4.4 Estimation

We now obtain the posterior estimates of the unemployment rates under the random walk time series and cross-sectional model given by (3) and (4). We used the Rao-Blackwellized estimators, given in section 3.3, to obtain estimates for the posterior mean and the posterior variance of θ_{iT}. We denote these estimates by HB1. To study the impact of using a smoothed estimate of the sampling covariance matrix Σ_i, we also used the direct survey estimate of Σ_i in the model. We denote the estimates obtained in this case by HB2. To study the effect of borrowing strength over time, we also obtained the HB estimates of θ_{iT} under the Fay-Herriot model based only on the current cross-sectional data, denoted by HB3. Figure 2 displays the LFS direct estimates and the three HB estimates of the June 1999 unemployment rates for the 62 CAs across Canada. The 62 CAs appear in the order of population size with the smallest CA (Dawson Creek, BC, population is 10,107) on the left and the largest CA (Toronto, Ont., population is 3,746,123) on the right. For the point estimates, the Fay-Herriot model (HB3) tends to shrink the estimates towards the average of the unemployment rates, which leads to estimates that are too smooth in general. HB2 has more variation and tends to have more extreme values than HB1, since HB2 uses the direct estimates of Σ_i subject to sampling errors. HB1 leads to moderate smoothing of the direct LFS estimates. For the CAs with large population sizes and therefore large sample sizes, the direct estimates and the HB estimates are very close to each other; for smaller CAs, the direct and HB estimates differ substantially for some regions.

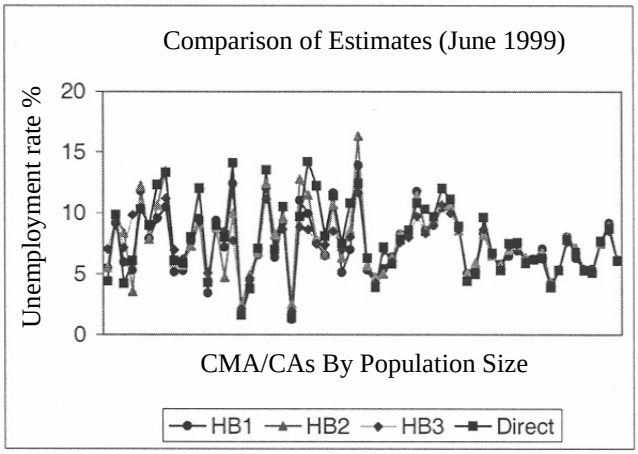

Figure 2. Comparison of direct and HB estimates.

Figure 3 displays the coefficients of variation (CV) of the estimates. The CV of the HB estimate is taken as the ratio of the square root of the posterior variance and the posterior mean. It is clear from Figure 3 that the direct estimate has the largest CV and HB1 has the smallest CV. HB1 has smaller CV than HB2 for all CAs, and HB2 has smaller CV than HB3 for all CAs except two relatively small CAs. The efficiency gain of the HB estimates is obvious, particularly for the CAs with smaller population sizes.

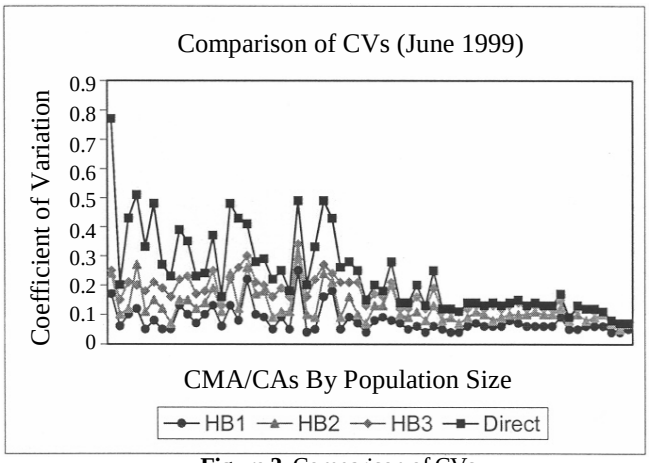

Figure 3. Comparison of CVs.

Figure 4 shows the percent CV reduction over the direct survey estimates for HB1, HB2 and HB3. The percent CV reduction is defined as the difference of the LFS CV and the HB CV relative to the LFS CV and is expressed as a percentage. It is clear that HB1 achieves the largest CV

reduction and that HB3 has the smallest reduction. The average percent reduction in CVs over the direct LFS estimates for the Fay-Herriot model (HB3) is 21%, for HB2 is 40%, and for HB1 is 62%. Also the CV reduction for smaller CAs is more significant than for larger CAs. As population size increases, the CV reduction tends to decrease.

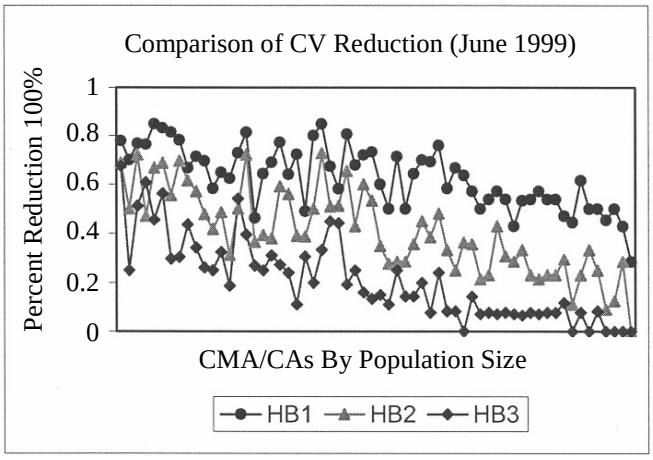

Figure 4. Comparison of CV reduction.

In summary, we conclude the following: (1) The model-based HB estimates improve the direct LFS estimates. In particular, the cross-sectional and random walk time series model (HB1) improves the LFS estimates considerably in terms of CV reduction. (2) The cross-sectional and random walk time series model is more effective than the Fay-Herriot model. (3) Use of smoothed estimate of the sampling variance-covariance matrix Σ_i is very effective.

5. Concluding Remarks

In this paper we have presented a hierarchical Bayes cross-sectional and time series model to obtain model-based estimates of unemployment rates for CAs across Canada using LFS data. The model borrows strength across areas and over time periods simultaneously. Our analysis has shown that this model with a random walk process on the random time series components fits the data quite well. The hierarchical Bayes estimates, based on this model, improve the direct survey estimates significantly in terms of CV, especially for CAs with small population. However, these CVs are based on the assumption that the sampling variance covariance matrices Σ_i in the model are known. As a result, the uncertainty associated with the estimation of Σ_i is ignored.

We also considered the well-known Fay-Herriot model that combines cross-sectional information only, using the data at a specific time point, for example, at the current time of interest *T*. We found that the CVs under the Fay-Herriot model lie between the CVs for the direct and the model-based approach presented here. The cross-sectional and time series model is uniformly superior to the Fay-Herriot model in terms of CV reduction. This is expected since our model extends the Fay-Herriot model by borrowing strength over time as well as across space.

In our application to the LFS, we used simple smoothed estimates of the sampling variance-covariance matrices Σ_i and then treated them as the true Σ_i. We plan to study the sensitivity of the HB estimates of small area parameters θ_{iT} and the associated CVs to different methods of smoothing the Σ_i. In particular, it may be more realistic to use smoothed estimates of the form $\tilde{\sigma}_{itt} = (\overline{\mathrm{CV}_i})^2 \theta_{it}^2$ and $\tilde{\sigma}_{its} = \overline{\rho}_{|t-s|} (\tilde{\sigma}_{itt} \tilde{\sigma}_{iss})^{1/2}$ instead of the simple smoothed estimates we have used. However, it is more difficult to implement the HB method in this case since $\tilde{\sigma}_{itt}$ and $\tilde{\sigma}_{its}$ depend on the unknown parameters θ_{it}.

In this paper, we used a linear mixed linking model (3) for the parameters θ_{it}, which matches with the sampling model (1). Recently, You and Rao (2002) developed unmatched sampling and linking models for cross-sectional data, where the linking model is a non-linear mixed model, unlike the sampling model (1). You, Chen and Gambino (2002) extended this method to cross-sectional and time series data, using a log-linear linking model for θ_{it}.

Acknowledgements

The authors would like to thank two referees and the Editor for their helpful comments and suggestions. This work was partially supported by a research grant from the Natural Sciences and Engineering Research Council of Canada to J.N.K. Rao.

Appendix

A.1. Let $X = (X_1', \ldots, X_m')$, $\theta = (\theta_1', \ldots, \theta_m')$, $u = (u_1', \ldots, u_m')'$, with $\theta_i' = (\theta_{i1}, \ldots, \theta_{iT})$, $u_i' = (u_{i1}, \ldots, u_{iT})$. In the following, we list the full conditional distributions for the cross-sectional and time series model. For the proposed model (random walk time component), $\rho = 1$; for the alternative AR(1) time component model, $|\rho| < 1$.

- $\beta \mid y, \sigma_v^2, \sigma_\varepsilon^2, u, \theta \sim N((X'X)^{-1}(\theta - u), \sigma_v^2 (X'X)^{-1})$;
- $\sigma_v^2 \mid y, \beta, \sigma_\varepsilon^2, u, \theta \sim IG(a_1 + mT/2, b_1 + \sum_{i=1}^m \sum_{t=1}^T (\theta_{it} - x_{it}'\beta - u_{it})^2/2)$;
- $\sigma_\varepsilon^2 \mid y, \beta, \sigma_v^2, u, \theta \sim IG(a_1 + m(T-1)/2, b_2 + \sum_{i=1}^m \sum_{t=2}^T (u_{it} - \rho u_{i,t-1})^2/2)$;
- For $i = 1, \ldots, m$,

$$u_{i1} \mid y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i2}, \theta \sim N\left(\left(\frac{1}{\sigma_v^2} + \frac{\rho^2}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\theta_{i1} - x_{i1}'\beta}{\sigma_v^2} + \frac{\rho u_{i2}}{\sigma_\varepsilon^2}\right), \left(\frac{1}{\sigma_v^2} + \frac{\rho^2}{\sigma_\varepsilon^2}\right)^{-1}\right);$$

- For $i = 1, \ldots, m$, and $2 \le t \le T-1$,

$$u_{i1} \mid y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,\,t-1}, u_{i,\,t+1}, \theta$$

$$\sim N\left(\begin{array}{c}\left(\frac{1}{\sigma_v^2}+\frac{1+\rho^2}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\theta_{i1}-x_{i1}'\beta}{\sigma_v^2}+\frac{\rho u_{i,\,t-1}+\rho u_{i,\,t+1}}{\sigma_\varepsilon^2}\right), \\ \left(\frac{1}{\sigma_v^2}+\frac{1+\rho^2}{\sigma_\varepsilon^2}\right)^{-1}\end{array}\right);$$

– For $i = 1, \ldots, m$,

$$u_{i1} \mid y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,\,T-1}, \theta$$

$$\sim N\left(\begin{array}{c}\left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\theta_{iT}-x_{iT}'\beta}{\sigma_v^2}+\frac{\rho u_{i,\,T-1}}{\sigma_\varepsilon^2}\right), \\ \left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\end{array}\right);$$

– For $i = 1, \ldots, m$,

$$\theta_i \mid y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u \sim N((\sigma_v^2 I_T + \Sigma_i^{-1})^{-1} \times (\Sigma_i^{-1} y_i + \sigma_v^{-2}(X_i\beta + u_i)), (\sigma_\varepsilon^{-2} I_T + \Sigma_i^{-1})^{-1}).$$

A.2. Let $y_t = (y_{1t}, \ldots, y_{mt})'$, $X_t' = (x_{1t}, \ldots, x_{mt})$, $\theta_t' = (\theta_{1t}, \ldots, \theta_{mt})'$, $t = 1, \ldots, T$, we list the full conditional distributions for the Fay-Herriot model at time point t as follows:

– $\beta_t \mid y_t, \sigma_{vt}^2, \theta_t \sim N((X_t' X_t)^{-1} X_t' \theta_t, \sigma_{vt}^2 (X_t' X_t)^{-1})$;

$$\sigma_{vt}^2 \mid y_t, \beta_t, \sigma_\varepsilon^2, u, \theta \sim IG(a_1 + m/2, b_1 + \sum_{i=1}^{m} (\theta_{it} - x_{it}' \beta_t)^2 / 2);$$

– For $i = 1, \ldots, m$,

– $\theta_{it} \mid y_t, \beta_t, \sigma_{vt}^2 \sim N((1 - r_{it}) y_{it} + r_{it}\, x_{i1}' \beta_t, \sigma_{it}^2 (1 - r_{it}))$, where $r_{it} = \sigma_{it}^2 / (\sigma_{it}^2 + \sigma_{vt}^2)$.

References

Bayarri, M.J., and Berger, J.O. (2000). *P* values for composite null models. *Journal of the American Statistical Association*, 95, 1127-1142.

Datta, G.S., Day, B. and Maiti, T. (1998). Multivariate Bayesian small area estimation: An application to survey and satellite data. *Sankhyā*, 60, 344-362.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the U.S. *Journal of the American Statistical Association*, 94, 1074-1082.

Datta, G.S., Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102, 83-97.

Dempster, A.P. (1974). The direct use of likelihood for significance testing (with discussion). In *Proceedings of Conference on Foundational Questions in Statistical Inference* (Eds. O. Barndorff-Nielsen, P. Blaeslid and G. Schou). Dept. of Theoretical Statistics, University of Aarhus, Denmark. 335-354.

Fay, R.E., and Herriot, R.A. (1979). Estimates of Income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. and Lindeyer, J. (1998). *Methodology of the Canadian Labour Force Survey*, Statistics Canada, Catalogue No. 71-526.

Gelfand, A.E., and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Gelfand, A.E., and Smith, A.F.M. (1991). Gibbs sampling for marginal posterior expections. *Communications In Statistics–Theory and Methods*, 20, 1747-1766.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-472.

Ghosh, M., Nangia, N. and Kim, D.H. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, 91, 1423-1431.

Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal (with discussion). *Statistical Science*, 9, 55-93.

Gilks, W.R., Wang, C.C., Yvonnet, B. and Coursagt, P. (1993). Random-effects models for longitudinal data using Gibbs sampling. *Biometrics*, 49, 441-453.

Karim, M.R., and Zeger, S.L. (1992). Generalized linear models with random effects: Salamander mating revisited. *Biometrics*, 48, 631-644.

Laud, P., and Ibrahim, J. (1995). Predictive model selection. *Journal of Royal Statistical Society, Series B*, 57, 247-262.

Meng, X.L. (1994). Posterior predictive *p* value. *The Annals of Statistics*, 22, 1142-1160.

Pfeffermann, D., Feder, M. and Signorelli, D. (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics*, 16, 339-348.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-186.

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

Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1996). BUGS 0.6: Bayesian inference Using Gibbs Sampling Manual. Available at http://www.mrc-bsu.cam.ac.uk/bugs.

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

You. Y., Chen, E. and Gambino, J. (2002). Nonlinear mixed effects cross-sectional and time series models for unemployment rate estimation. *2002 Proceedings of the American Statistical Association, Section on Government Statistics* [CD-ROM], Alexandria, VA: American Statistical Association.

You, Y., and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, 26, 173-181.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30, 3-15.