Nonlinear mixed effects cross-sectional and time series models for unemployment rate estimation

Yong You, Edward Chen and Jack Gambino, Statistics Canada
Yong You, HSMD, RHC-16, Statistics Canada, Ottawa, ON, Canada, K1A 0T6

Key Words: Gibbs sampling, Labour Force Survey, Small area, Unemployment rate, Unmatched model.

1. Introduction

Traditional small area estimators borrow strength either from similar small areas or from the same area across time, but not both. In the past ten 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), Datta et al. (1999) and You et al. (2000), successfully exploit the two dimensions simultaneously to produce improved estimates with desirable properties for small areas. Datta et al. (1999) applied their model to data from the U.S. Current Population Survey, while You et al. (2000) applied a similar model to the Canadian Labour Force Survey (LFS). Unlike Datta et al. (1999), the model proposed by You et al. (2000) does not contain seasonal parameters. This reduces substantially the number of parameters that need to be estimated. Despite this simplification, You et. al. (2000) obtained both an adequate model fit and large reductions in the coefficients of variation (CVs) of the small area estimators of the unemployment rate. However, a major limitation of the models proposed by Datta et al. (1999) and You et al. (2000) is that the linking model for the parameter of interest, the true unemployment rate, is a linear model with normal random effects. The linear linking model may lead to negative estimates for some small areas. To overcome this limitation, in this paper we propose a nonlinear linking model for the parameters of interest; see Section 2.

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 Urban Centers (UCs) including 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 UCs do not have a large enough sample to produce adequate direct estimates. Our objective in this paper is to obtain an estimator that is an improvement over the direct estimator which is based solely on the sample falling in a given UC in a given month. In Canada, unemployment rates are produced by the Labour Force Survey. 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 et al. (1998).

In Section 2, we present a cross-sectional and time series model, which borrows strength across small areas and time periods to produce unemployment rates for UCs. In Section 3, the model is placed in a hierarchical Bayes framework, and the use of Gibbs sampling to generate samples from the joint posterior distribution is described. Specifically, estimates for the UC (small area) unemployment rate and its posterior variance are obtained. In Sections 4 and 5, we apply our method to LFS data and check the adequacy of the model. We offer some concluding remarks in Section 6.

2. Cross-sectional and Time Series Models

Let y_{it} denote the direct LFS estimate of θ_{it}, the true unemployment rate of the ith UC at time t, $i = 1,...,m,$, $t = 1,...,T$, where m is the total number of UCs and T is the (current) time of interest. Following You et al. (2000), we assume that

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

where e_{it} 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 ith UC. Because of the LFS sample rotation pattern, there is substantial sample overlap over short time periods. As a result, the correlation between e_{it} and $e_{is} (t \neq s)$ has to be taken into account. We 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

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

Thus y_i is design-unbiased for θ_i. Specification of the covariance matrix Σ_i may not be easy in practice. Usually a smoothed estimator of Σ_i is used in the model, and then Σ_i is treated as known. More details on constructing a smoothed estimator of Σ_i in the context of the LFS are given in Section 4. To borrow strength across small areas and time periods, You et al. (2000) modelled the true unemployment rate θ_{it} by a linear regression model with random effects through auxiliary variables x_{it}, that is,

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

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

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

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 }, { ε_{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 for u_{it}, whereas Datta et al. (1999) included month and year effects as seasonal effects for θ_{it} in (3).

However, the linear linking model (3) for the true unemployment rate θ_{it} has some limitations. Since θ_{it} is the true unemployment rate, it is a positive number between 0 and 1, and it is close to 0. The linear linking model with normal random effects may lead to negative estimates for θ_{it}. To avoid this problem, we propose the following log-linear linking model for θ_{it}, that is,

$$\log(\theta_{it}) = x'_{it}\beta + v_i + u_{it}, \quad i = 1,...,m, t = 1,...,T. \qquad (5)$$

Note that the log-linear linking model (5) is a nonlinear model and the sampling model (1) is a linear model. The proposed linking model (5) is an extension of the unmatched sampling and linking models of You and Rao (2002) to cross-sectional and time series data.

We are interested in obtaining a model-based estimator of θ_i, in particular, for the current time unemployment rate θ_{iT}. Due to the complex model, following You et al. (2000), we consider a complete HB approach to inference using the Gibbs sampling method.

3. Hierarchical Bayes Analysis

In this section, we apply the hierarchical Bayes approach to the nonlinear cross-sectional and time series models given by (2), (4) and (5). Estimates of the posterior mean and posterior covariance matrix of the small area means, θ_i, 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 $\theta_i = (\theta_{i1},...,\theta_{iT})'$, $[y_i \mid \theta_i] \sim ind\ N(\theta_i, \Sigma_i)$;
- Conditional on the parameters β, u_{it} and σ_v^2, $[\log(\theta_{it}) \mid \beta, u_{it}, \sigma_v^2] \sim ind\ N(x'_{it}\beta + u_{it}, \sigma_v^2)$;
- Conditional on the parameters $u_{i,t-1}$ and σ_ε^2, $[u_{it} \mid u_{i,t-1}, \sigma_\varepsilon^2] \sim ind\ N(u_{i,t-1}, \sigma_\varepsilon^2)$;

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 inverse 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.

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 the Gibbs sampling method (Gelfand and Smith, 1990) to obtain the posterior mean and the posterior variance of θ_{iT}.

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.

For the hierarchical Bayes model 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 full conditional distributions are given in the Appendix. The distributions of β, σ_v^2 and σ_ε^2, u_{it} are standard normal or inverse gamma distributions that can be easily sampled. However, the conditional distribution of θ_i does not have a closed form. We use the Metropolis-Hastings algorithm within the Gibbs sampler (Chib and Greenberg, 1995) to update θ_i. From the Appendix, we note that

$$\theta_i \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u \propto h(\theta_i) f(\theta_i) ,$$

where

$$h(\theta_i) = \exp\{-\frac{1}{2}(y_i - \theta_i)'\Sigma_i^{-1}(y_i - \theta_i)\}$$

and

$$f(\theta_i) = \exp\{-\frac{1}{2\sigma_v^2}\sum_{t=1}^{T}(\log(\theta_{it}) - x_{it}'\beta - u_{it})^2\}(\prod_{t=1}^{T}\frac{1}{\theta_{it}}).$$

To update θ_i, we proceed as follows:

(1) For $t = 1,...,T$, draw
$\theta_{it}^{(k+1)} \sim \log N(x_{it}'\beta^{(k+1)} + u_{it}^{(k+1)}, \sigma_v^{2(k+1)})$,
then we have $\theta_i^{(k+1)} = (\theta_{i1}^{(k+1)},...,\theta_{im}^{(k+1)})'$.

(2) Compute the rejection probability
$$\alpha(\theta_i^{(k)}, \theta_i^{(k+1)}) = \min\{\frac{h(\theta_i^{(k+1)})}{h(\theta_i^{(k)})}, 1\}.$$

(3) Generate $\lambda \sim Uniform(0,1)$, if $\lambda < \alpha(\theta_i^{(k)}, \theta_i^{(k+1)})$, then accept $\theta_i^{(k+1)}$; otherwise reject $\theta_i^{(k+1)}$ and set $\theta_i^{(k+1)} = \theta_i^{(k)}$.

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. The convergence monitoring is discussed in Section 4. Estimates of the posterior mean $E(\theta_{iT} \mid y)$ and the posterior variance $V(\theta_{iT} \mid y)$ are obtained based on the samples generated from the Gibbs sampler.

4. Application to the LFS

We used the 1999 LFS unemployment estimates, y_{it}, in our HB analysis. There are 64 UCs across Canada in the LFS. Employment Insurance (EI) beneficiary rates are used as auxiliary data, x_{it}, in the model. Since the EI beneficiary data are available for only 62 UCs, we included only those *m*=62 UCs in the model. Within each UC, we considered six consecutive monthly estimates y_{it} from January 1999 to June 1999, so that *T*=6. 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.

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 UC over time and the average lag correlation coefficients over time and all UCs. By using these smoothed CVs and lag correlation coefficients, we obtained a smoothed estimate of Σ_i. You et al. (2000) found that using the smoothed estimate of Σ_i in the model can significantly improve CV reduction and model fit in terms of posterior predictive p-values.

To implement the Gibbs sampling, we considered *L*=10 parallel chains, each of length *2d*=3000. For each chain, the first *d*=1500 “burn-in” iterations were deleted. To reduce the auto-correlation in the chain, we took every 5th iteration of the remaining iterations, leading to 300 iterations for each chain. 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=1,…, K (K=300), l=1,…, L*. In the first step, the overall mean $\bar{\theta}_{iT} = \sum_{l=1}^{L}\sum_{k=1}^{K}\theta_{iT}^{(lk)}/(LK)$ and the within sequence mean $\bar{\theta}_{iT}^{(l)} = \sum_{k=1}^{K}\theta_{iT}^{(lk)}/K$ for *l=1,…, L* are computed. Then compute B_{iT}/K, the variance between the *L* sequence means as $B_{iT}/K = \sum_{l=1}^{L}(\bar{\theta}_{iT} - \bar{\theta}_{iT}^{(l)})^2/(L-1)$. In the second step, calculate W_{iT}, the average of the *L* within sequence variances, s_{iTl}^2, each based on *K-1* degrees of freedom; that is, $W_{iT} = \sum_{l=1}^{L}s_{iTl}^2/L$. In the third step, calculate

$$s_{iT}^2 = (K-1)W_{iT}/K + B_{iT}/K$$

and

$$V_{iT} = s_{iT}^2 + B_{iT}/(LK).$$

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 $\hat{R}_{iT}$.

Figure 1 displays the LFS direct estimates and the HB estimates of the June 1999 unemployment rates for the 62 UCs across Canada. The 62 UCs appear in the order of population size with the smallest UC (Dawson Creek, BC) on the left and the largest UC (Toronto, Ontario) on the right. The HB approach leads to moderate smoothing of the direct LFS estimates. For the UCs with large population sizes and therefore large sample sizes, the direct estimates and the model-based estimates are very close to each other as expected; for smaller UCs, the direct and HB estimates differ substantially for some regions and the model-based estimates tend to shrink to the mean value.

Figure 1. Comparison of Estimates

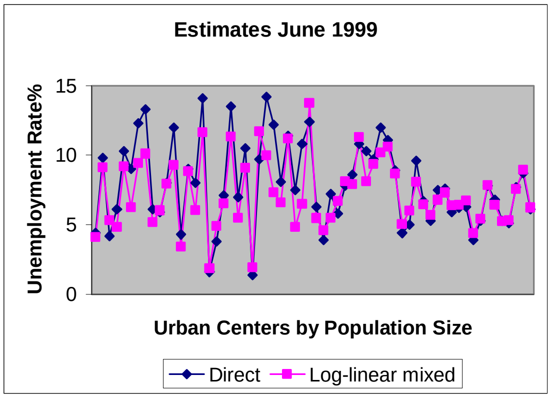
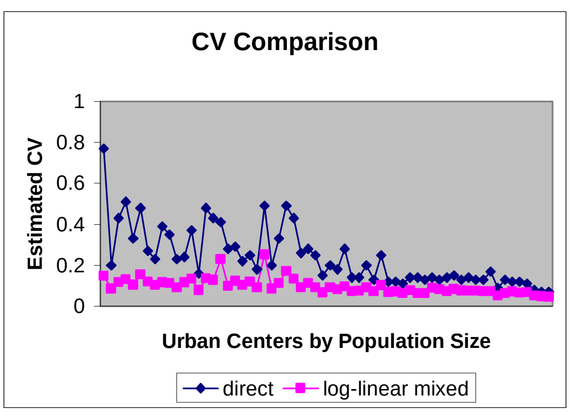


Figure 2 displays the coefficients of variation (CV) of the estimates. The CV of the model-based estimate is taken as the ratio of the square root of the posterior variance and the posterior mean. It is clear from Figure 2 that the model-based approach leads to substantial CV reduction over the direct estimates. The efficiency gain of the model-based estimates is obvious, particularly for the UCs with smaller population sizes.

Figure 2. Comparison of CVs



5. Model Checking

To check the overall fit of the proposed model, we used the method of posterior predictive *p* values (Meng, 1994; Gelman et al., 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. 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 $f(y|y_{obs})$, where y_{obs} represents the observed data. The posterior predictive *p* value is defined as

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

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 $T(y_{obs},\theta)$ 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.

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 $T(y_{obs},\theta^*)$. Then the *p* value is estimated by the proportion of times $T(y^*,\theta^*)$ exceeds $T(y_{obs},\theta^*)$.

In the present context, the discrepancy measure used for overall fit is given by

$$T(y,\theta) = \sum_{i=1}^{m}(y_i-\theta_i)'\Sigma_i^{-1}(y_i-\theta_i).$$

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 chains. We obtained an estimated *p* value equal to 0.553. Thus we have no indication of lack of overall model fit.

Datta et al. (1999) used a model similar to ours to obtain HB estimates of unemployment rates for the states of the U.S. They included month and year effects in the time component of the model since they considered a long time series (48 months). They stated that if the month and year effects were deleted, the corresponding *p* value in their application is equal to 0.758, compared to the *p* value of 0.614 for their proposed model with month and year effects. Our study has shown that when data for a short time period (12 months or less) are used, the simpler model without seasonal effects can provide adequate overall fit. The CVs obtained from the simpler model are likely to be comparable to those obtained from a more complex model involving seasonal effects and long time series data. Also, the determination of Σ_i becomes more difficult as the number of time periods, *T*, included in the model increases.

6. Concluding Remarks

In this paper we have presented a hierarchical Bayes nonlinear mixed effects cross-sectional and time series model to obtain efficient model-based estimates of unemployment rates for UCs across Canada using LFS data. The model borrows strength across areas and over time periods simultaneously. Our analysis has shown that the proposed model fits the data quite well and the hierarchical Bayes model-based estimates improve the direct survey estimates significantly in terms of CV reduction, especially for UCs with small population. The proposed log-linear linking model overcomes the limitations of Datta et al. (1999) and You et al. (2000).

In our approach, we have treated the variance-covariance matrices Σ_i as known even though they are estimated. We plan to study the sensitivity of the estimates and their CVs to this assumption by trying different methods for smoothing variances and covariances and, possibly, by putting a prior distribution on the variance-covariance matrix. Finally, we plan to produce estimates for additional months to study the behaviour of the time series produced by our approach.

Appendix

We list the full conditional distributions for the Gibbs sampler in Section 3 as follows:

$$\beta \mid Y, \sigma_v^2, \sigma_\varepsilon^2, u, \theta \sim N((X'X)^{-1}X'(\log(\theta)-u), \sigma_v^2(X'X)^{-1});$$

$$\sigma_v^2 \mid Y, \beta, \sigma_\varepsilon^2, u, \theta \sim IG(a_1 + mT/2, b_1 + \textstyle\sum_{i=1}^m \sum_{t=1}^T (\log(\theta_{it}) - x_{it}'\beta - u_{it})^2)/2);$$

$$\sigma_\varepsilon^2 \mid Y, \beta, \sigma_v^2, u, \theta \sim IG(a_2 + m(T-1)/2, b_2 + \textstyle\sum_{i=1}^m \sum_{t=2}^T (u_{it} - u_{i,t-1})^2)/2);$$

For $i = 1,\ldots,m$,

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

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

$$u_{it} \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,t-1}, u_{i,t+1}, \theta \sim N\left(\left(\frac{1}{\sigma_v^2}+\frac{2}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\log(\theta_{it})-x_{it}'\beta}{\sigma_v^2}+\frac{u_{i,t-1}+u_{i,t+1}}{\sigma_\varepsilon^2}\right), \left(\frac{1}{\sigma_v^2}+\frac{2}{\sigma_\varepsilon^2}\right)^{-1}\right);$$

For $i = 1,\ldots,m$,

$$u_{iT} \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u_{i,T-1}, \theta \sim N\left(\left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\left(\frac{\log(\theta_{iT})-x_{iT}'\beta}{\sigma_v^2}+\frac{u_{i,T-1}}{\sigma_\varepsilon^2}\right), \left(\frac{1}{\sigma_v^2}+\frac{1}{\sigma_\varepsilon^2}\right)^{-1}\right);$$

For $i = 1,\ldots,m$,

$$\theta_i \mid Y, \beta, \sigma_v^2, \sigma_\varepsilon^2, u \propto \exp\{-\frac{1}{2}(y_i-\theta_i)'\Sigma_i^{-1}(y_i-\theta_i)\}\times \exp\{-\frac{1}{2\sigma_v^2}\sum_{\mathrm{t}=1}^{\mathrm{T}}(\log(\theta_{\mathrm{it}})-x_{it}'\beta-u_{it})^2\}\left(\prod_{\mathrm{t}=1}^{\mathrm{T}}\frac{1}{\theta_{\mathrm{it}}}\right).$$

REFERENCES

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 327-335.

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.

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.

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

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

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

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. *Canadian Journal of Statistics*, 22, 511-528.

You, Y., Rao, J.N.K. and Gambino, J. (2000) Hierarchical Bayes estimation of unemployment rates for sub-provincial regions using cross-sectional and time series data. *American Statistical Association, Proceedings of the Section on Government Statistics and Section on Social Statistics*,160-165.

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.