Statistica Sinica **9**(1999), 297-324

A STOCHASTIC SYSTEM FOR MODELING LABOR FORCE SERIES OF SMALL AREAS

Jing-Shiang Hwang and Arthur P. Dempster

Academia Sinica, Taipei and Harvard University

Abstract: Time series models of sampling error, true unobserved rates, and covariates can be used to pool data across time and space to reduce variance in a subnational estimator. We present such models along with associated hierarchical Bayesian analyses. Specifically, we present a joint time series model for a 51 U.S. state labor force series in a Bayesian framework. Data are input in the form of optimal composite estimates from a sampling error model. The basic time series model is constructed from fractional Gaussian noise processes. Covariation of the true series across states is modeled by having a common national component modified by individual state components. Markov chain Monte Carlo methods are applied to develop samplers for a high-dimensional system of 105 parameters. The results indicate substantial gains in the efficient use of CPS data for U.S. state employment and unemployment rates series.

Key words and phrases: Bayesian inference, long memory process, Markov chain Monte Carlo, small area estimation, unemployment rates.

1. Introduction

Official statistical inferences on small area characteristics are often based on survey data obtained from each small area, but sample sizes within those areas are often too small to provide reliable estimates. In recent years, fitting time series models to repeated survey data has been proposed to pool data across time to reduce variance in the survey estimator, see, e.g., Bell and Hillmer (1990), Binder and Dick (1989), Pfeffermann (1992), and Tiller (1992). Pfeffermann and Burck (1990) and Pfeffermann and Bleuer (1993) proposed state-space models to improve small area estimation from both time series amd cross-sectional information. The model structure typically consists of two distinct models, with the main model describing the evolution of true time series over time, and the survey error model representing the sampling error pattern. For such state-space models, the Kalman filter is often used to compute likelihoods and estimate parameters, and the signals are extracted based on the estimated parameter values. However, the approach needs a specific model (e.g., ARIMA) for the sampling errors over time.

Rao and Yu (1994) also proposed a combined cross-sectional and time series model involving autocorrelated random effects and sampling errors with an arbitrary covariance matrix over time. Their model has the same form as in Pfeffermann and Burck (1990), which is an extention of the model due to Fay and Herriot (1979), but they assumed a known block-diagonal covariance matrix for the sampling errors. Basically, these models are special cases of a general mixed linear model involving fixed and random effects. Therefore, the popular best linear unbiased predictor (BLUP) methods can be used to estimate the small area mean which can be expressed as a linear combination of fixed effects and realized values of random effects. Estimation procedures for the uncertainty in the variance components are also available in the literature, see e.g., Rao and Yu (1994), Ghosh and Rao (1994), Prasad and Rao (1990), and Kackar and Harville (1984).

The Bayesian approach for inference on the small area means has been considered in various studies since the posterior computation was feasible. The purpose of this paper is also to illustrate applied Bayesian statistics in action. However, our emphasis is not so much on Bayes as it is on scientific model construction. The essence of our modeling of the labor force data is on getting the science right. That is, on attempting to recognize major aspects of sample design, temporal variation, and spatial variation.

We propose a stochastic system of three phases in a Bayesian framework that are cumulative in the sense that Phase 2 builds on Phase 1, and Phase 3 will build on Phases 1 and 2. The three phases relate to three different Gaussian linear models developed to represent different phenomena in the data, each having its own posterior sampler for sampling parameters. Phase 1 relates to sampling error of survey data. Phase 2 brings in time series models and develops estimation methods that balance sampling error and time series properties of the true series using familiar signal estimation methods joined with resampling schemes for parameter estimation. Phase 3 adds auxiliary variables to the already complicated Phase 2.

To describe the system and proposed techniques, we consider the problems of estimating employment rates (EMP) and unemployment rates (UNEMP) of 51 "small areas" consisting of the 50 United States plus the District of Columbia. The major data source for analyzing the subpopulation characteristics is the Current Population Survey (CPS), the monthly national survey of some 56,000 households conducted and tabulated by the U.S. Census Bureau for the Bureau of Labor Statistics. For the 11 large states, about 2,700 - 5,000 households are interviewed each month, while two thirds of the 51 small areas have monthly sample sizes less than 1,000. The specific data described below cover the 48

months of calendar years 1986 through 1989, and were used in developing both models and associated computational methods. This particular selection of a short time series resulted from a decision to focus on a recent period when the CPS design was not undergoing discrete changes of sample sizes or sampling frames. Our results are only meant to be indicative of the gains available from efficient Bayesian combination of information sources. Contemporary application of our methods would require analysis of more recent data, and allowance for current survey design.

The CPS design tracks each selected household over a period of 16 months, observing each household initially 4 times at monthly intervals, then making no observations over the following 8 months, and finally observing again for 4 successive months of a second year. Under this 4-8-4 rotation scheme, 1/8 of the sample is interviewed each month for the first time, 1/8 for the second time, ..., and 1/8 for the eighth and last time. The standard approach to composite estimation under this 4-8-4 scheme is motivated by the heuristic that a month-to-month change can be more accurately estimated from households in which both of the successive months are observed than it can from an equal number of observations on different households in the successive months. A simple composite estimator can therefore be defined by determining appropriate relative weights for two distinct unbiased estimators, namely the average (e.g., UNEMP) from households not observed in the previous month, and the sum of the previous month's composite estimator and average month-to-month change from households observed in both months.

As elaborated in Phase 1, a more detailed disaggregation of the sampled units into subgroups makes possible a broader class of compositors (Hwang (1992), (1996), Dempster and Hwang (1993), (1994)). The Phase 1 models we fit to the data specify optimal choices in the broader class, leading to assessments of optimal posterior standard errors. We find that the optimal compositor yields efficiency gains of 5-10% for UNEMP and 25-45% for EMP, respectively. Such gains are worth pursuing, but composite estimation alone is not a panacea. Because composite estimation does not incorporate the ideas of borrowing strength across states and time series projection that are expected to be major sources of potential efficiency gains, there is reason to pursue further variance reduction.

These two major sources of potential efficiency gains are incorporated and implemented in Phase 2 to improve accuracy. Phase 2 of our analysis strategy builds on Phase 1 in the sense that we carry forward, mostly unchanged, the modeling and Bayesian variance components estimation procedures of Phase 1. The difference comes in the final step of combining optimal compositors with prior distributions for the targets UENMP and EMP. In place of arbitrary and

unrealistic uniform prior distributions assigned to the true estimates in Phase 1, we develop stochastic models of the time series of estimates, for use as prior distributions for combination with the optimal compositors.

Section 2 starts with a brief description of Phase 1, the sampling error model. The main part of Section 2 covers the detailed model constructions and choices of Phase 2. The conditional posterior means and covariance matrices of the 51 true time series in Phase 1 are treated as sufficient statistics for the series and are modeled as a sum of a national component and a state's specific component. Each of the components is further decomposed into nonseasonal and seasonal components. Section 3 presents the techniques, including a computational framework for computing conditional posteriors, approximate likelihoods and parameter samplers. Based on the sampled parameters, we compute the conditional posterior means and covariance matrices of any target of interest which is a combination of the components in the time series model. The approximate posteriors are then obtained by averaging these conditional posteriors via Markov chain Monte Carlo methods. The results shown in Section 4 indicate substantial gains in the efficient use of CPS data for both the EMP and UNEMP series. In addition, preliminary attempts at Phase 3 analyses that bring in nonCPS series as covariates indicate further worthwhile savings (Hwang (1992)). We are not reporting Phase 3 results here, pending model improvements.

2. Model Construction

2.1. Phase 1

The noise component of the Current Population Survey data represents error that arises from sampling only a portion of the total population. Its structure depends upon the CPS design and population characteristics. For the first phase, we focus on those design features that are likely to have a major impact on the variance and autocovariance structure of the sampling error.

An important feature of the CPS is the large overlap in sample units from month to month. The sample is divided into eight independent panels or rotation groups. Units are partially replaced each month according to a 4-8-4 rotating panel. When new households are introduced into the sample, they are included for four consecutive months, dropped out for eight months, returned for the same four calendar months of the next year, and then dropped from the survey.

Another feature is that the use of a rotation system requires the periodic selection of additional samples. A new term, *stream*, was defined as a set of households and their successive replacements at 16 month intervals. Each new "sample", in the technical CPS sense of new households that are phased in over an 8-month period, contributes to each of the 8 streams. Because the replacement

of households within a stream is typically by neighboring households, it is evident that 8 random stream effects can be anticipated to persist across the 48 month observation period selected for study.

Therefore, the empirical input to our study is a pair of $51 \times 48 \times 8$ data arrays, one giving estimates of EMP and the other estimates of UNEMP. The two arrays come from the same surveys, and could be analyzed jointly, but for ease of modeling and analysis we treat them separately in parallel analyses. The "51" dimension refers to the 50 U.S. states plus Washington DC, referred to in the sequel as "51 states". The "48" dimension refers to the 48 calendar months from Jan. 1986 through Dec. 1989, and the "8" refers to the stream dimension that provides replication internal to the survey design, and is the basis for compositing.

We developed a variance components model with 4 variances, representing variance among streams, variance among *samples* within a stream, variance between the first and the second year within a sample within a stream, and a residual variance. Each state has in principle its own 4-vector of variances. These vary in part due to different sample sizes. But even after standardizing the variances to a "per sample household" basis that nominally adjusts for sample size differences, there is apparent variation. Thus we felt it important to develop statistical methodology to represent and fit state-to-state variation in the per household 4-vectors of variances in the model of sampling error.

For each state, the EMP and UNEMP data input in the jth stream at time t, denoted by $Y_{t,j}$, is represented as the sum of month level μ_t, month-in-sample bias effect ν, plus three random components: S_j for stream, $V_{j,g}$ for sample g within stream j, $W_{j,g}$ for annual change within sample g within stream j, and residual $e_{t,j}$. That is,

$$Y_{t,j} = \mu_t + \nu + S_j + V_{j,g} + W_{j,g} + e_{t,j},$$

$$t = 1, \ldots, n,\ j = 1, \ldots, 8 \text{ and } \ g = 1, \ldots, g_j,$$

where $n = 48$ is the number of months and g_j, a function of t within each j, is the number of different samples in stream j. The four random components are assumed normally distributed as follows:

$$\begin{aligned} S_j &\sim \mathrm{N}(0, \sigma^2_{str}) \\ V_{j,g} &\sim \mathrm{N}(0, \sigma^2_{sam}) \\ W_{j,g} &\sim \mathrm{N}(0, \sigma^2_{lag}) \\ e_{t,j} &\sim \mathrm{N}(0, \sigma^2), \end{aligned}$$

and all these random quantities are assumed independent.

Because a 48×8 data array for a single state is too small to produce accurate point estimates of the 4 variances for that state, especially in the case of stream-to-stream variance where only 7 degrees of freedom are available among the 8 streams, Dempster and Hwang (1994) developed a technique for borrowing strength across states, specifically assigning to a normal distribution the logs of the 4 variances across states whose parameters ("hyperparameters") can be roughly estimated.

Dempster and Hwang (1994) also developed a sampler to sample the four variances from an approximate posterior distribution of the four variances. Given a sample of the four variances and data $Y_{t,j}$ of the 51 states, we calculated the conditional mean vector and covariance matrix of the $n = 48$ monthly levels for state i, denoted as $Z_i^{(1)} = (Z_{i1}^{(1)}, \ldots, Z_{in}^{(1)})$ and $\Sigma_i^{(1)}$, respectively and used these as input data for Phase 2 model construction in Section 2.2. The form of covariance matrix $\Sigma_i^{(1)}$ is a little complicated. But, the conditional mean level of the ith state for the tth month is a composite estimator which can be explicitly written as

$$Z_{it}^{(1)} = \sum_{j=1}^{8} c_j Y_{t,j} + \sum_{l=1}^{t-1} \sum_{j=1}^{8} b_{lj} Y_{t-l,j}, \quad t = 2, \ldots, n, \tag{1}$$

where

$$\sum_{j=1}^{8} c_j = 1, \qquad \sum_{j=1}^{8} b_{lj} = 0 \quad \text{for } l = 1, \ldots, t-1.$$

The noncomposite or raw estimator of the monthly level can be also represented as equation (1) by choosing $c_j = 1/8, j = 1, \ldots, 8$ and setting $b_{lj} = 0$. That is, simply average the original 8 stream estimates. The conditional variance of the noncomposite estimator is the sum of the four given variances divided by 8.

To obtain posterior means and variances of the composite and noncomposite estimators, we repeat drawing four variances from the sampler of approximate posterior distribution of the four variances a total of K times. The posterior variance of the noncomposite estimator is the average of these K conditional variances. The posterior mean of the composite estimator, which we call the Phase 1 estimate, is the average of the K simulated conditional posterior means. The posterior variance for the Phase 1 estimate is the sample variance of the K simulated means plus the average of the K simulated conditional posterior variances.

2.2. Phase 2

Let $Z_i = (Z_{i1}, \ldots, Z_{in})^T$ be the true series of length $n = 48$ for the ith state. We model these 51 series simultaneously so that each individual series consists

of a common national component and its deviation from the national level. The reason for this decomposition is that it is a simple and plausible way to represent correlation among the 51 time series. We might try to group states in different ways, to allow for regional effects, or for industry effects, and develop a more complicated factor model. In view of the limitation on data, and inexperience with the statistical methods, we preferred to start with a simple dependence structure, yet one capable of picking up an important source of dependence. The common component can be estimated using all 51 states whence only the state-specific component relies on data from a single state. The national component is further decomposed into three components, namely, a fixed effects component for a seasonal pattern that repeats each year, and two stochastic components for nonseasonal and seasonal trends. Similarly, we allow each state to have its own repeating seasonal pattern, and stochastic nonseasonal and seasonal components.

The joint model of the 51 true series is defined as

$$\begin{aligned} Z_i^{(1)} &= Z_i + \mathcal{E}_i \\ &= J\mu + \tau_i \mathcal{N} + \mathcal{S} + J\mathcal{P}_i + \mathcal{N}_i + \mathcal{S}_i + \mathcal{E}_i, \\ &\qquad \mathcal{E}_i \sim N(0, \Sigma_i^{(1)}),\ i=1,\ldots,51. \end{aligned} \tag{2}$$

Equation (2) is interpreted as a mechanism for combining two sources of information about the unknown Z_i. First, the $Z_i^{(1)}$ and $\Sigma_i^{(1)}$ are the point estimates and associated error covariance matrices of Z_i obtained from Phase 1 of the sampling error model, though in practice they are represented only approximately from a posterior sampler that samples the Z_i from its posterior given only the 48 months of sample data and a uniform prior on Z_i. Second, the decomposition of the Z_i into 6 terms defines a genuine prior distribution for the Z_i (i.e., prior to the sample data). Combining the prior and sample information leads to our Phase 2 posterior for Z_i. The component $\mu = (\mu_1, \ldots, \mu_{12})^T$ represents the 12 months' fixed effects of the national seasonal pattern. Matrix $J = (I_{12}, \ldots, I_{12})^T$ is a linear operator mapping the seasonal pattern to each year. The other components on the right-hand side of (2) are stochastic components or random effects representing, respectively, the national nonseasonal term with a factor τ_i for the ith state, the national seasonal term, the seasonal pattern, and lastly nonseasonal, seasonal, and sampling error terms for the ith state.

We also assume that the 51 states' deviation components $(\mathcal{P}_i, \mathcal{N}_i, \mathcal{S}_i)$ are independent of each other, both within and between states, in the joint model. The dependence of $Z_1, Z_2, \ldots, Z_{51}$ is induced by the national components. Furthermore, Gaussian models will be used throughout, since these are essentially the only ones that are practicable with current computing tools and they appear to cover most situations adequately.

The vector μ is used to model the initial levels of the national time series for the 12 calendar months. It is assumed to have a diffuse distribution *a priori*. Thus μ is assumed to be normally distributed with mean zero and covariance $\Sigma_\mu \to \infty$. But for each state the seasonal pattern may have its own characteristics. This stochastic phenomenon is expressed for each state as

$$\mathcal{P}_i \sim N(\mathbf{0}, \sigma_P^2 I_{12}).$$

In modeling the nonseasonal and seasonal terms, we turn to the theory of stationary time series and in particular to a relatively new class of Gaussian models based on the so-called fractional Gaussian noise (fGn) process (Carlin and Dempster (1989)). A unit variance fGn process has autocovariance function $\gamma(t; d)$ defined for all integer lags t with a shape parameter $0 < d < 1$, where

$$\gamma(t; d) = \frac{1}{2}[|t-1|^{1+d} - 2|t|^{1+d} + |t+1|^{1+d}]. \tag{3}$$

Let Σ_N^d be the Toeplitz matrix with first row $\gamma(0; d), \ldots, \gamma(n-1; d)$, the autocovariances of a unit variance fGn process with shape parameter d, length n. Recall that $\Sigma = (\sigma_{jk})$ is Toeplitz whenever $\sigma_{jk} = \sigma_{|j-k|}$, i.e., the value of any entry is a function only of its distance to the main diagonal. Thus, it is easy to see that the first row (or column) defines a Toeplitz matrix. Here the subscript N is introduced in order to distinguish this covariance matrix from one later used in the modeling of seasonal components. Thus the nonseasonal random component is

$$\mathcal{N} \sim N(\mathbf{0}, \Sigma_N^d), \tag{4}$$

$$\mathcal{N}_i \sim N(\mathbf{0}, \lambda_i^2 \Sigma_N^d), \tag{5}$$

where λ_i is a scale parameter.

Next we turn to modeling the stationary seasonal processes. Here we use a seasonal analogue of fGn, where the autocovariance function is

$$\gamma_S(t; d) = \begin{cases} \gamma(t/12; d), & t = 0 \bmod 12 \\ 0, & \text{otherwise.} \end{cases}$$

This is a very simple modification of the basic nonseasonal fGn process and could be elaborated in other ways; see Carlin (1987). Let Σ_S^d be the Toeplitz matrix with first row $\gamma_S(0; d), \ldots, \gamma_S(n-1; d)$; then

$$\mathcal{S} \sim N(\mathbf{0}, \sigma_N^2 \Sigma_S^d), \tag{6}$$

$$\mathcal{S}_i \sim N(\mathbf{0}, \sigma_S^2 \Sigma_S^d), \tag{7}$$

where σ_N and σ_S are two scale parameters for nation and state, respectively.

The reason for choosing the fGn process as the building block of the time series models has been supported by some empirical results. The true time series of each state is modeled as a sum of a national component and the state's deviation from the national level. Each of these two components has a nonseasonal stochastic term which is set to be a scale parameter times a unit variance fGn process.

An important characteristic of fGn process is its approximately linear spectral density with slope $-d$ on the log–log scale. In our model the spectral densities of the national component and each state's deviation component are straight lines with peaks at the Fourier frequencies in log–log scale. This property provides a basis for checking whether this model will fit the data.

The data for the time series model are 51 series of optimal composite estimates from the sampling error model. That is $\{Z_i^{(1)}\}_{i=1}^{51}$. Let $\bar{Z}$, the average of these 51 series, be an estimate of the national component. The deviation series $\{Z_i^{(1)} - \bar{Z}\}_{i=1}^{51}$ are used to estimate the state's common component of deviation from the nation.

It is a known property of the periodogram that the expected periodogram of the averaged series, $\bar{Z}$, converges uniformly to the spectral density of the national component in the model at Fourier frequencies in $(0, 0.5]$. So we can compute the expected periodograms and plot them with the spectral density of a unit variance fGn process with proper shape parameter value. If the model, equation (4) is reasonable, then we should get two roughly parallel lines in the log scale.

The expected periodogram (expected under sampling error) of the average of the 51 series, $\bar{Z}$, is estimated by the sum of the 51 individual periodograms divided by 51 squared. Similarly for the state's deviation component, we may average the 51 periodograms of the state's deviation series to compare with the spectral density of a unit variance fGn process with proper shape parameter value. If we obtain two parallel lines, we are comfortable with the selected model in equation (5). The expected periodograms for nation and state drawn in solid curves in Figure 1 are computed based on a set of optimal composite estimates $Z_1^{(1)}, \ldots, Z_{51}^{(1)}$ from Phase 1. The dotted line is the spectral density of a unit variance fGn process with $d = 0.99$.

The pictures provide strong support for our choice of a time series model with long memory for the nation and state components. Note that in choosing the shape parameter value d we compare the posteriors of log likelihood of several d values and we find that $d = 0.99$ is a better choice for both UNEMP and EMP (Hwang (1992)).

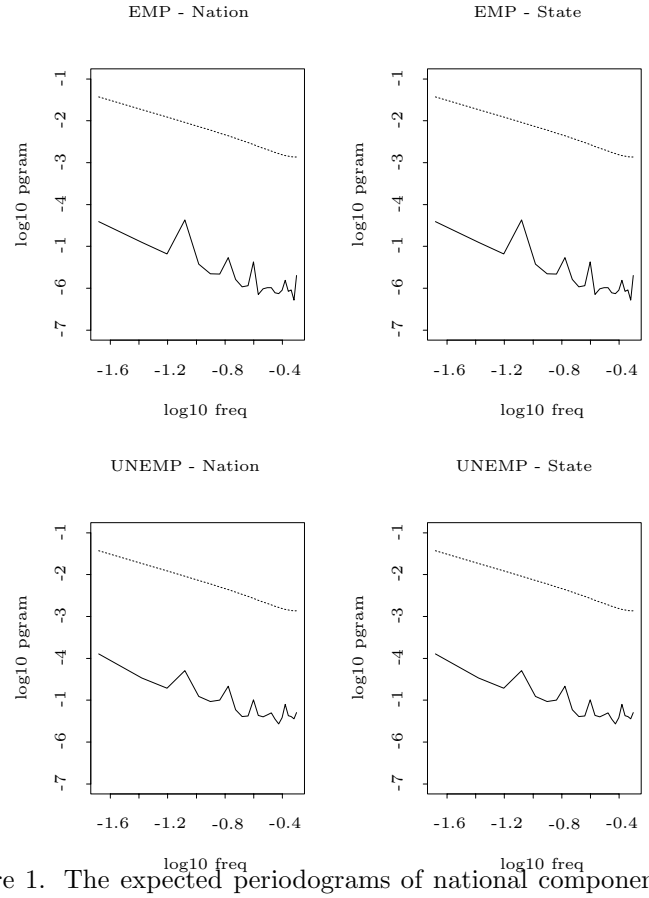

Figure 1. The expected periodograms of national components and the averaged periodograms of state's deviation from national level and spectral density of unit variance fGn process with shape parameter $d = 0.99$.

Another model choice issue is whether to introduce state-specific variation in one or both of nonseasonal and seasonal compnents. There are four possibilities that we denote schematically by 1N + 1S, 1N + 51S, 51N + 1S, and 51N + 51S. For example, 51N + 1S means that we introduce state-to-state variation among the nonseasonal components of the time series, but we impose a common nonseasonal term across states, and similarly for the other three cases. To compare different statistical models for the same data, we plot the posterior distribution of the likelihood of each in Figure 2. This is an alternative to standard methods of comparing models via penalized maximized likelihoods, where the posterior variation spread downward from the maximum of the likelihood plays the role of the penalty term in AIC. Motivation and theory concerning the use of posterior likelihood can be found in Aitkin (1991), (1997), and Dempster (1997). The posterior distributions of likelihood for the four models listed above are natural byproducts of Bayesian Gibbs samplers. We selected the 51N + 51S model since it clearly dominates the others in fit.

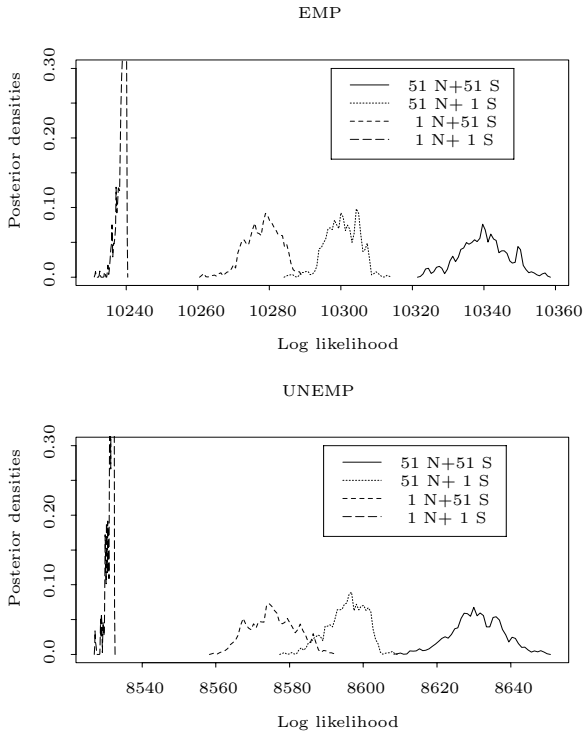

Figure 2. Posterior distributions of the log likelihood for the four models.

We have specified model (2) with a parameter vector $\boldsymbol{\theta}$ consisting of $\{\tau_i^2, \lambda_i^2\}_{i=1}^{51}$, σ_P^2, σ_N^2, and σ_S^2 for the true time series $Z = (Z_1, \ldots, Z_{51})$. For our Bayesian analysis, first we have to at least develop a sampler for sampling $\boldsymbol{\theta}$ from its posterior and compute the conditional estimate of any target given $\boldsymbol{\theta}$ and $\{Z_i^{(1)}, \Sigma_i^{(1)}\}_{i=1}^{51}$. The target here may be any linear function of the components in the joint model. For example, the target may be the whole series of a state; then it equals the sum of the 6 components. The approximate posterior mean and covariance matrix of the target are obtained through the following procedures. First, a large number of conditional posteriors of the target are computed. Each of the conditional posteriors is based on a new set of $Z_i^{(1)}$ and $\Sigma_i^{(1)}$ from Phase 1 and a sample $\boldsymbol{\theta}$ from the approximate posterior of the parameters in Phase 2. Second, we average these conditional posteriors across the posterior of parameters in Phase 1 and the posterior of $\boldsymbol{\theta}$ in Phase 2 by approximate Monte Carlo method in the usual way, i.e., the approximate posterior mean is the average of the simulated conditional posterior means, and the approximate posterior covariance matrix is the sum of the average of the conditional posterior covariance matrices and the sample covariance matrix of the conditional posterior means.

Sampling of $\boldsymbol{\theta}$, and the computation of conditional posterior means and variances, requires the likelihood function of the parameter vector $\boldsymbol{\theta}$. The likelihood of $\boldsymbol{\theta}$ is very complicated, so there is no easy way to derive an exact posterior distribution of $\boldsymbol{\theta}$ for sampling use. We overcome this obstacle by computing numerical likelihood values via a conceptual framework. The computations of the posterior of $\boldsymbol{\theta}$ and posterior of target are also based on the same framework. These computational details are illustrated in the next section.

3. Computations

The computations of the conditional distribution and likelihood of components in Gaussian linear models are simplified in a sweeping framework. The formulas derived in Dempster (1982) are applied in the process of repeated computation. The computational framework starts with a big array containing the mean and covariance matrix of the initial joint distribution of observed data and unknown components. The formulas produce the conditional distribution of the components, and also the log likelihood, by sweeping (SWP) on the observed data. The operators SWP and RSW (for reverse sweep) are defined and their properties are described in Dempster (1969), (1982).

3.1. Framework

To obtain a likelihood or the conditional posterior of any target we have to sweep on the observable $Z_1^{(1)}$, $Z_2^{(1)}$, ..., $Z_{51}^{(1)}$. Sweeping on $Z_i^{(1)}$, denoted as SWP$[Z_i^{(1)}]$, encounters $\mathrm{Cov}(Z_i^{(1)}, Z_j^{(1)})$ for all $j \neq i$, the sum of Σ_μ, $\tau_i\tau_j\Sigma_N^d$, and $\sigma_N^2\Sigma_S^d$, and approaches infinity. These operations can be avoided by initializing the big covariance array by implementing SWP$[\mu]$, SWP$[\mathcal{N}]$ and SWP$[\mathcal{S}]$ to remove temporarily the national components, making the blocks corresponding to $\mathrm{Cov}(Z_i^{(1)}, Z_j^{(1)})$ zero for all $i \neq j$. The remaining steps involved in the complete algorithm after initialization can now be represented as SWP$[Z_1^{(1)}]$, ..., SWP$[Z_{51}^{(1)}]$ and then RSW$[\mu]$, RSW$[\mathcal{N}]$ and RSW$[\mathcal{S}]$ to sweep back these components.

A schematic illustration of the computational array for the initialized big array with the three national components swept out is shown in Figure 3, where row (and column) blocks are labeled according to a set of abbreviations: E is expected values of estimates; T_i's are targets; $Z_i^{(1)}$'s are the observations. We work in terms of the upper triangular half of the array, since the lower half is redundant by symmetry. Sub-arrays are named in an obvious way by suitably combining the abbreviations: for instance, the block that corresponds to the variance of $Z_i^{(1)}$ is labeled $Z_i^{(1)}Z_i^{(1)}$. Initial values are described as follows. First, the sum of squares EE is set to zero. Next, the arrays $EZ_i^{(1)}$ contain initially the appropriate observed values which are the outputs of the sampling error

model. The vector $E\mu$ is initialized at zero, in accordance with the rule that all blocks indexed by the fixed effects are zero except those crossing with observable series or targets, because of the limiting infinite variance that has been swept out. Arrays $E\mathcal{N}$ and $E\mathcal{S}$ also begin with value zero throughout, since that is the prior mean of the generator. Finally, each ET_i contains a prior mean of zero.

	E	$Z_1^{(1)}$	$\cdots$	$Z_{51}^{(1)}$	μ	$\mathcal{N}$	$\mathcal{S}$	T_1	$\cdots$	T_{51}
E	0	$Z_1^{(1)}$	$\cdots$	$Z_{51}^{(1)}$	0	0	0	0	$\cdots$	0
$Z_1^{(1)}$		Ω_1	0	0	J	$\tau_1 I$	I	Λ_1	0	0
$\vdots$			$\ddots$	0	$\vdots$	$\vdots$	0	0	$\ddots$	0
$Z_{51}^{(1)}$				Ω_{51}	J	$\tau_{51} I$	I	0	0	Λ_{51}
μ					$-\Sigma_\mu^{-1} = 0$	0	0	J^T	$\cdots$	J^T
$\mathcal{N}$						$-\Sigma_N^{d^{-1}}$	0	$\tau_1 I$	$\cdots$	$\tau_{51} I$
$\mathcal{S}$							$-\sigma_N^{-2}\Sigma_S^{d^{-1}}$	I	$\cdots$	I
T_1								Λ_1	0	0
$\vdots$									$\ddots$	0
T_{51}										Λ_{51}

Figure 3. Initial values for the computational array $\mathcal{T}_I$ with national components swept out.

Note that the national components $\mu, \mathcal{N}$ and $\mathcal{S}$ have been swept out in Figure 3 so that the covariance blocks for any two Z's and two T's are zero matrices. The target here is the whole series, i.e.,

$$T_i = J\mu + \tau_i\mathcal{N} + \mathcal{S} + J\mathcal{P}_i + \mathcal{N}_i + \mathcal{S}_i.$$

The covariance matrix of $Z^{(1)} = (Z_1^{(1)}, \ldots, Z_{51}^{(1)})$ consists of 51 diagonal blocks which are

$$\Omega_i = \sigma_P^2 JJ^T + \lambda_i^2\Sigma_N^d + \sigma_S^2\Sigma_S^d + \Sigma_i^{(1)}, \ \ i = 1, \ldots, 51.$$

The blocks $Z_i^{(1)}\mu$, $Z_i^{(1)}\mathcal{N}$ and $Z_i^{(1)}\mathcal{S}$ are reduced to J, $\tau_i I$ and I, respectively, for $i = 1, \ldots, 51$. The covariance of $Z_i^{(1)}$ and T_i is

$$\Lambda_i = \sigma_P^2 JJ^T + \lambda_i^2\Sigma_N^d + \sigma_S^2\Sigma_S^d.$$

Blocks $\mu\mu$, $\mathcal{N}\mathcal{N}$ and $\mathcal{S}\mathcal{S}$ contain the negative inverse of covariance matrices of μ, $\mathcal{N}$ and $\mathcal{S}$, where Σ_μ^{-1} is set to zero.

Initialization of the computational array is completed by considering the covariances required for the T_iT_i blocks. These are determined in the familiar

way, from the stochastic model structure of specific target quantities. Thus they are $\Lambda_i, i = 1, \ldots, 51$. For most inferential purposes, the diagonal elements of $T_i T_i$ are all that are needed, providing finally the posterior variances of the targets.

One interesting target is the seasonally adjusted series. The true seasonally adjusted series for the ith state can be generated by $\tau_i \mathcal{N} + \mathcal{N}_i$ plus a fixed overall mean $\bar{\mu}$. To make any inference on the deseasonalized series, we may just add this target $T_i^* = \bar{\mu} + \tau_i \mathcal{N} + \mathcal{N}_i$ to the big array.

If the prior covariance matrix of the $Z_i^{(1)}$ is of Toeplitz form, then the Levinson algorithm (Carlin (1987)) can be introduced to reduce the computational burden from $O(n^3)$ to $O(n^2)$. In our case, the covariance Ω_i is not a Toeplitz matrix because the covariance matrix of sampling error term does not satisfy the condition. The operations of the remaining stages are therefore carried out using standard SWP.

Applying matrix operations on the above formulas we can obtain the final values of the determinant, which is a by-product of the operations, the sum of squares term, and the conditional posterior means and variances of the targets from the final array. For the short series we are dealing with, computing time to run the matrix operations several times is not a problem. But it is a computational burden to compute the exact likelihood values for a large number of $\boldsymbol{\theta}$'s. We need therefore to develop a valid and time saving approximation to the likelihood.

3.2. Frequency domain representations

To obtain an approximation of the likelihood function, we transform the model to the frequency domain, i.e., Fourier transform each term in model (2). We actually apply a Fourier transform matrix as an operator on the time domain generators, $\mathcal{N}$, $\mathcal{S}$, $\mathcal{P}_i$, $\mathcal{N}_i$, and $\mathcal{S}_i$ and sampling error component $\mathcal{E}_i$. The operator is derived and the prior covariance matrix (given $\boldsymbol{\theta}$) of the Fourier transformed components are shown to be close to diagonal in this subsection.

We start with frequency domain theory from the Cramér representation of a stationary time series as a stochastic integral. The account given here is equivalent but we prefer to work with finite representations using circular processes, since the results follow from straightforward linear algebra, with no need to invoke the general theory of stochastic integration. It is also consistent with our overall theme of keeping theoretical developments as closely as possible in tune with their computational implementations.

To keep the mathematics completely elementary, suppose the observed series of length n is taken from a circularly invariant process of length m. The usual stationary process theory is the case where $m \to \infty$. Asymptotics as $n \to \infty$ can

be thought about in several ways: either fixing the ratios $k = m/n$ and letting m and n both get large at the same rate, or letting k get large too. For the immediate purpose, it matters little.

Formulas are simpler if the time origin is centered at the middle of the data. So instead of writing the basic process as y_t for $t = 0, 1, \ldots, n-1$, we will use y_t for $t = -\frac{n-1}{2}, -\frac{n-1}{2} + 1, \ldots, \frac{n-1}{2}$, with points t at half-integer points if n is even and whole integer points if n is odd. Also, the expressions for finite Fourier transform are a little different for even and odd n and m. So we will assume in the following that both n and m are even for consistency.

We denote the white noise frequency domain generators of the circular process by

$$U_0, U_{\frac{1}{m}}, V_{\frac{1}{m}}, \ldots, U_{\frac{1}{2}-\frac{1}{m}}, V_{\frac{1}{2}-\frac{1}{m}}, V_{\frac{1}{2}},$$

where U_ω multiplies a cosine and V_ω multiplies a sine for $\omega = 0, \frac{1}{m}, \ldots, \frac{1}{2}$.

Specifically the Cramér representation is

$$y_t = (\frac{f_0}{m})^{\frac{1}{2}} + (\frac{2}{m})^{\frac{1}{2}} \sum_{j=1}^{\frac{m}{2}-1} (f_{\frac{j}{m}})^{\frac{1}{2}} (U_{\frac{j}{m}} \cos 2\pi \frac{j}{m} t + V_{\frac{j}{m}} \sin 2\pi \frac{j}{m} t) + (\frac{f_{\frac{1}{2}}}{m})^{\frac{1}{2}} V_{\frac{1}{2}} \sin \pi t,$$

where $t = -\frac{m-1}{2} + l$, $l = 0, \ldots, m-1$, and f_ω is the spectral density at frequency ω.

The matrix terms denote the time series by column vector

$$Y = [y_{-\frac{m-1}{2}}, y_{-\frac{m-1}{2}+1}, \ldots, y_{\frac{m-1}{2}}]^T,$$

the frequency domain generators by column vector

$$W = [U_0, U_{\frac{1}{m}}, V_{\frac{1}{m}}, \ldots, V_{\frac{1}{2}}]^T,$$

and the $m \times m$ diagonal matrix with diagonal $[(f_0)^{\frac{1}{2}}, (f_{\frac{1}{m}})^{\frac{1}{2}}, (f_{\frac{1}{m}})^{\frac{1}{2}}, \ldots, (f_{\frac{1}{2}})^{\frac{1}{2}}]$ by $(f)^{\frac{1}{2}}$.

Thus the matrix form of Cramér representation is

$$Y = H_m (f)^{\frac{1}{2}} W,$$

where H_m is the orthonormal finite Fourier transform matrix

$$H_m = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \mathbf{s}_1 \\ \vdots \\ \mathbf{c}_{\frac{m}{2}-1} \\ \mathbf{s}_{\frac{m}{2}-1} \\ \mathbf{s}_{\frac{m}{2}} \end{bmatrix}, \tag{8}$$

where the pairs of rows for $j = 1, \ldots, \frac{m}{2} - 1$ are

$$\mathbf{c}_j = (\frac{2}{m})^{\frac{1}{2}}(\cos 2\pi \frac{j}{m}\frac{1-m}{2}, \cdots, \cos 2\pi \frac{j}{m}t, \cdots, \cos 2\pi \frac{j}{m}\frac{m-1}{2})$$
$$\mathbf{s}_j = (\frac{2}{m})^{\frac{1}{2}}(\sin 2\pi \frac{j}{m}\frac{1-m}{2}, \cdots, \sin 2\pi \frac{j}{m}t, \cdots, \sin 2\pi \frac{j}{m}\frac{m-1}{2}),$$

and $(\frac{2}{m})^{\frac{1}{2}}$ is replaced by $(\frac{1}{m})^{\frac{1}{2}}$ for the $\mathbf{c}_0$ and $\mathbf{s}_{\frac{m}{2}}$.

In sampling terms, Y denotes the population while the sample is the piece out of the middle at times $t = -\frac{n-1}{2}, -\frac{n-1}{2} + 1, \ldots, \frac{n-1}{2}$. Denoting this piece by Y^n and the corresponding middle set of columns of H_m by H_m^n, we have the Cramér representation of the data

$$Y^n = H_m^n(f)^{\frac{1}{2}}W.$$

Finally, for discussing asymptotics, we need notation for sample finite Fourier transforms, specifically

$$F = H_n Y^n.$$

Now we recall the original idea to treat F as the data. The key then becomes to understand the representation of F in terms of the frequency domain generators W, or K in the following

$$\begin{aligned} F &= H_n H_m^{nT}(f)^{\frac{1}{2}}W \\ &= K(f)^{\frac{1}{2}}W. \end{aligned} \tag{9}$$

Explicit formulas for the elements of K are very simple, and easily derived using the submatrix representation of H_m, together with same schematic representation with n in place of m, and using the following formulas

$$\sum_{t=-\frac{m-1}{2}}^{\frac{m-1}{2}} \cos 2\pi\omega' t \cos 2\pi\omega'' t = \frac{1}{2}[D_n(\omega' - \omega'') + D_n(\omega' + \omega'')]$$
$$\sum_{t=-\frac{m-1}{2}}^{\frac{m-1}{2}} \sin 2\pi\omega' t \sin 2\pi\omega'' t = \frac{1}{2}[D_n(\omega' - \omega'') - D_n(\omega' + \omega'')]$$
$$\sum_{t=-\frac{m-1}{2}}^{\frac{m-1}{2}} \cos 2\pi\omega' t \sin 2\pi\omega'' = 0,$$

where

$$D_n(\omega) = \frac{\sin \pi\omega n}{\sin \pi\omega}$$

is the *Dirichlet* kernel.

The matrix K has n rows and m columns. The n rows correspond to a cosine at $\omega' = 0$, a cosine and a sine at $\omega' = \frac{1}{n}, \ldots$, a cosine and a sine at $\omega' = \frac{j}{n}, \ldots$, and finally a sine at $\omega' = \frac{1}{2}$. Similarly, the m columns correspond to a cosine at $\omega'' = 0$, a cosine and a sine at $\omega'' = \frac{1}{m}, \ldots$, a cosine and a sine at $\omega'' = \frac{l}{m}, \ldots$, and finally a sine at $\omega'' = \frac{1}{2}$. Thus, the 2×2 piece of K with $\omega' = \frac{j}{n}, \omega'' = \frac{l}{m}$, is

$$(\frac{1}{nmg})^{\frac{1}{2}} \begin{bmatrix} D_n(\frac{l}{m} - \frac{j}{n}) + D_n(\frac{l}{m} + \frac{j}{n}) & 0 \\ 0 & D_n(\frac{l}{m} - \frac{j}{n}) - D_n(\frac{l}{m} + \frac{j}{n}) \end{bmatrix}, \tag{10}$$

where $q = 1$ for $0 < i < n/2$, and $q = 2$ for $i = 0, n/2$.

The covariance matrix of F, from (9), is

$$\Sigma_F = KfK^T \equiv \begin{pmatrix} \sigma_{ij}^c \\ \sigma_{ij}^s \end{pmatrix}_{n \times n}. \tag{11}$$

Here the elements σ_{ij}^c are the covariances corresponding to frequencies i/n and j/n for cosine part. Similarly σ_{ij}^s is defined for sine part. Both σ_{ij}^c and σ_{ij}^s converge to

$$\delta_{i-j} K_n * f(\frac{i}{n}) + M_f \cdot O(\frac{\log n}{n}) \text{ as } m \to \infty,$$

where $K_n(\omega) = \frac{1}{n} D_n^2(\omega)$ is the *Fejer*'s kernel, M_f is a constant and also can be viewed as the derivative of f at i/n. See the proof in the appendix, or Ramos (1988). Furthermore, well-known theory (see Hannan (1970)) shows that the convolution

$$K_n * f(\omega) \to f(\omega) \text{ as } n \to \infty.$$

It is clear that the off-diagonal elements of Σ_F will depend on the spectral density of the process. For the white noise process, Σ_F is a diagonal matrix since its spectral density is a constant. For the fGn process we are concerned with here, where spectral density is steep near zero frequency and flat and low in the high frequency range, Σ_F can be well approximated by keeping the submatrix of Σ_F corresponding to the first few low frequencies and letting the other entries be zero except the diagonal elements.

Therefore, the frequency domain representation of model (2) is written as

$$Z_i^{(1)F} = J^F \mu + \tau_i \mathcal{N}^F + \mathcal{S}^F + J^F \mathcal{P}_i + \mathcal{N}_i^F + \mathcal{S}_i^F + \mathcal{E}_i^F, \tag{12}$$

where $Y^F = H_n Y$ for each component Y. Models (2) and (12) are equivalent under the linear transformation. The components of the right-hand side of (12)

are still independent of each other and normally distributed. Thus we have

$$\begin{aligned}
\mathcal{N}^F &\sim N(\mathbf{0}, H_n \Sigma_N^d H_n^T) \\
\mathcal{S}^F &\sim N(\mathbf{0}, \sigma_N^2 H_n \Sigma_S^d H_n^T) \\
J^F \mathcal{P}_i &\sim N(\mathbf{0}, \sigma_P^2 J^F J^{FT}) \\
\mathcal{N}_i^F &\sim N(\mathbf{0}, \lambda_i^2 H_n \Sigma_N^d H_n^T) \\
\mathcal{S}_i^F &\sim N(\mathbf{0}, \sigma_S^2 H_n \Sigma_S^d H_n^T) \\
\mathcal{E}_i^F &\sim N(\mathbf{0}, H_n \Sigma_i^{(1)} H_n^T).
\end{aligned}$$

Note that the covariance matrix of $J^F \mathcal{P}_i$ is diagonal. We have noted that the covariance matrix $H_n \Sigma_N^d H_n^T$ of an fGn process is close to diagonal except the upper left submatrix corresponding to the first small Fourier frequencies. The empirical results show the sampling error process $\mathcal{E}_i^F$ also has the same property.

Now we recall the time domain big initial array in Figure 3. Suppose the first row of this array is $\mathbf{m}$, a vector of observations and unobserved effects. The rest is the covariance matrix V. Under this linear transform we obtain a similar initial frequency domain big array. The first row of the new big array is $H_n \mathbf{m}$ and the covariance matrix is $H_n V H_n^T$. The algorithm for computing the log likelihood of $\boldsymbol{\theta}$ is then the same as in the time domain formulation. In practice we run an approximate frequency domain version of the algorithm to obtain approximate likelihood of $\boldsymbol{\theta}$ by replacing the covariance matrix in the big array with its approximation.

The main reason for seeking an approximation of the likelihood here is to save computing time. Most of the computing time consumed in the algorithm comes from the matrix operations on the covariance matrices in these models. Since the off-diagonal of the covariance matrix of a long memory process in the frequency domain representation is ignorable for the high frequency part, we first assume each of the above covariance matrices Σ is replaced in the following form:

$$\begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \operatorname{diag}(\Sigma_{22}) \end{bmatrix} \approx \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \Sigma.$$

Then the matrix operations on Σ are reduced to operations on a small submatrix Σ_{11}, and scale operations on the diagonal of Σ_{22}. If we reduce the dimension of Σ_{11} to 1×1, i.e., Σ is replaced by $\operatorname{diag}(\Sigma)$, then the whole computational structure is reconstructed on a scale operation basis which of course provides a huge relief of computational burden. Though the approximation may be not well satisfied when the dimension of Σ_{11} is small, it is helpful to start with such

simple approximations to get a rough insight into the likelihood function. We then can keep improving the approximation by increasing the dimension of Σ_{11} as long as the computer capacity allows.

3.3. Bayesian computations

The sampler for drawing the 105-dim parameter $\boldsymbol{\theta}$ is created using a Metropolis algorithm within the conditioning of the Gibbs sampler. We describe our sampler using similar notations as in Smith and Roberts (1993), who give a comprehensive review of recent uses of Markov chain Monte Carlo (MCMC) methods for exploring and summarizing posterior distributions in Bayesian statistics. Let the parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{54})$, where $\theta_i = (\tau_i^2, \lambda_i^2)$, for $i = 1, \ldots, 51$, and $\theta_{52} = \sigma_N^2$, $\theta_{53} = \sigma_S^2$, $\theta_{54} = \sigma_P^2$. The computation of an approximate likelihood of $\boldsymbol{\theta}$ has been discussed in previous subsections. The approximate posterior is obtained by choosing a prior for $\boldsymbol{\theta}$. For the nonseasonal components $\theta_1, \ldots, \theta_{51}$, we assume these log θ_i are bivariate normally distributed. The common prior means, variances and correlation are roughly estimated from the approximate maximum likelihood estimates of these 51 pairs of parameters. For the three seasonal parameters, we assume the priors are proportional to the inverse of the parameters. Let $\pi(\boldsymbol{\theta}) = \pi(\theta_1, \ldots, \theta_{54})$ denote the approximate posterior, and let $\pi(\theta_i|\theta_{-i})$ denote the induced full conditional for each of the components θ_i, given values of the other components $\theta_{-i} = (\theta_j; j \neq i)$, $i = 1, \ldots, 54$. Note that π is only proportional to the true posterior; we need never compute the normalization of the posterior.

The Gibbs sampler algorithm proceeds as follows. First, pick arbitrary starting values $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_{54}^{(0)})$. Then successively make random drawings $\theta_i^{(1)}$ from the full conditional distributions $\pi(\theta_i|\theta_j^{(1)}, 1 \leq j < i \text{ and } \theta_k^{(0)}, 54 \geq k > i)$, $i = 1, \ldots, 54$.

This completes a transition from $\boldsymbol{\theta}^{(0)}$ to $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \ldots, \theta_{54}^{(1)})$. Iteration of this cycle of random variate generation produces a sequence $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(t)}, \ldots$ which is a realization of a Markov chain, with transition probability from $\boldsymbol{\theta}^{(t)}$ to $\boldsymbol{\theta}^{(t+1)}$ given by

$$K_G(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)}) = \prod_{i=1}^{54} \pi(\theta_i^{(t+1)}|\theta_j^{(t+1)}, 1 \leq j < i \text{ and } \theta_k^{(t)}, 54 \geq k > i).$$

Within each cycle of the Gibbs sampler, the drawing $\theta_i^{(t+1)}$ is obtained from the Metropolis algorithm (Metropolis *et al.* (1953)). A point θ^* is generated uniformly from a small neighborhood of $\theta_i^{(t)}$. We actually accept $\theta_i^{(t+1)} = \theta^*$

with acceptance probability $\alpha(\theta_i^{(t)}, \theta^*)$; otherwise, we reject the value and set $\theta_i^{(t+1)} = \theta_i^{(t)}$. The acceptance probability is set as

$$\alpha(\theta_i^{(t)}, \theta^*) = \min\{\frac{\pi(\theta^*|\theta_j^{(t+1)}, 1 \le j < i \text{ and } \theta_k^{(t)}, 54 \ge k > i)}{\pi(\theta_i^{(t)}|\theta_j^{(t+1)}, 1 \le j < i \text{ and } \theta_k^{(t)}, 54 \ge k > i)}, 1\}.$$

Under suitable regularity conditions, $\boldsymbol{\theta}^{(t)}$ converges to a sample from the posterior $\pi(\boldsymbol{\theta})$. Although there is a reassuring theoretical literature concerning the convergence of MCMC methods (see Smith and Roberts (1993)), results do not easily translate into clear guidelines for the practitioner. We propose monitoring the posterior of the log likelihood of $\boldsymbol{\theta}$ for the output analysis (convergence diagnostics). Figure 4 shows that the posteriors of the log likelihood of $\boldsymbol{\theta}^{(t)}$ move within some range after several iterations.

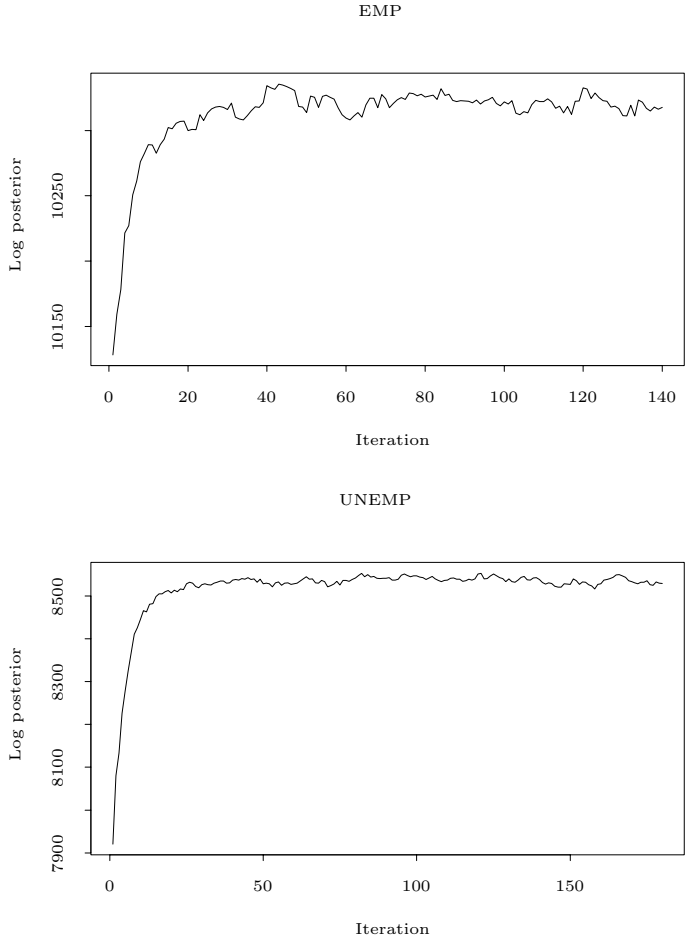

Figure 4. Convergence diagnostics of Gibbs sampler for EMP and UNEMP

This indicates that we may start collecting samples from each chain very soon. Since successive $\boldsymbol{\theta}^{(t)}$ are correlated, we may collect the samples by picking one from every several realizations in a single chain so that the autocovariances of the selected sample series are small.

The Gibbs sampling procedures described above are indeed an easily implemented technique that allows one to draw random sample from this complicated posterior and make the Bayesian inference feasible. However, Hobert and Casella (1996) showed that improper priors in Gibbs sampling may have ill-behaved effect on the posterior. So we have to check the posterior simulated by the Gibbs sampler, although only three of the 105 parameters use improper priors in the model.

4. Results and Discussion

For each of EMP and UNEMP, we analyzed a $51 \times 48 \times 8$ data array of weighted estimates from BLS sources, where the 51 and 48 dimensions refer, respectively, to 51 states and 48 months. The internal replication dimension 8 refers to what we call *streams* that are implicit in the 4-8-4 rotation system that controls selection and replacement of households. In order to stabilize the variance of the rate estimates and fit the data in normal models we take arc sin square root of each entry of the arrays. We implement on this scale until the final step at which point the results are transformed back to the original scale.

We compare the time series model estimates of the series with the optimal composite estimates in Phase 1 and the noncomposite estimates which are the 48 row averages of each state's 48×8 sample estimates described in Section 2.1. Figure 5 plots the posterior mean of the whole series (solid line) and the deseasonal series (dashed line), and the noncomposite estimates (dotted line) in the small population state of Nebraska. As expected, the model-based estimates of the series move much smoother than the noncomposite estimates. The deseasonal series generated without the seasonal components show stable trends of the series. Note that the posterior means of the UNEMP series are below the noncomposite estimates for the first 30 months or so and then they are mostly above, while this does not occur with the EMP series. One possible reason for the results is that Nebraska's specific characteristics of unemployment are far from those of other states and our Phase 2 inference had moved its unemployment rates toward the national trend with more weight because of the state's high variation. It is also very likely that both large variation and bias may occur in the noncomposite estimates for a state with extremely small sample size.

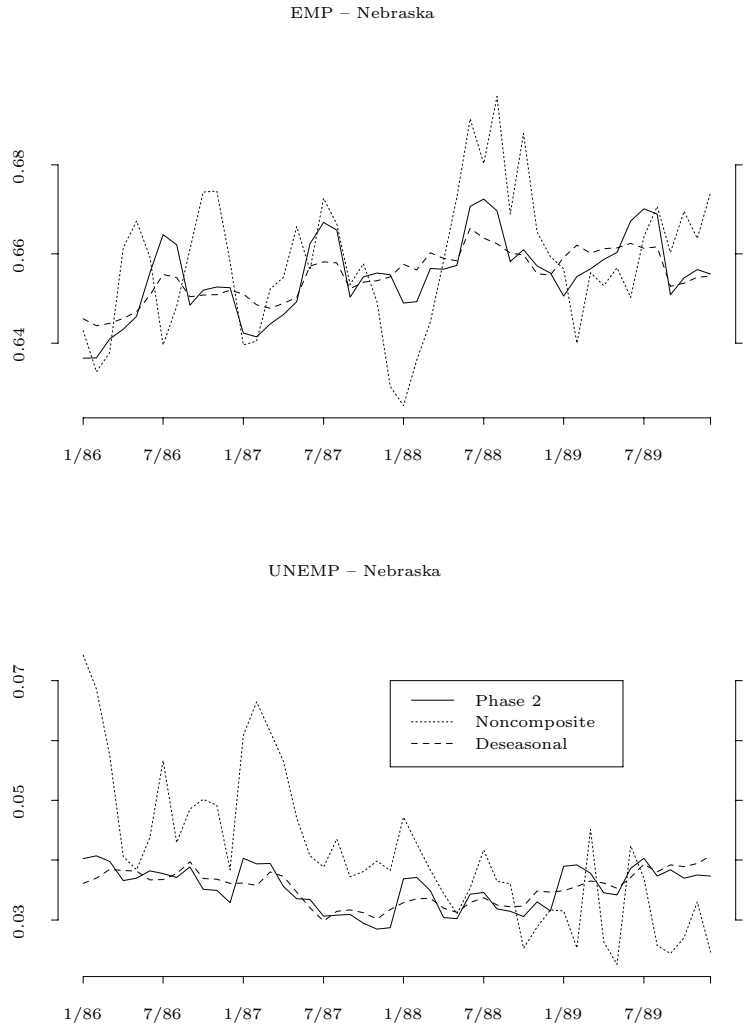

Figure 5. The posterior means of the whole series, deseasonal series, and the noncomposite estimates in Nebraska.

The variance reductions of the time series models show large efficiency gains for most of the 51 states. For demonstration, we compute the variances of the three EMP and UNEMP estimates in December 1989. Figure 6 shows the gains of UNEMP at Phase 1 are small. The largest gains appear at Phase 2 from the time series model. The standard deviations of the time series model estimates are reduced about 50% over the noncomposite estimate.

Note that Louisiana has small gain in UNEMP from the time series model. This indicates the model does not fit this state's UNEMP series too well. To figure

out what happens on this series, we checked the 51 noncomposite estimates of UNEMP series and found the sample variance of Louisiana to be much larger than for the other 50 states, reflecting this state's nonseasonal component in the model. That is, the posterior distribution of this state's nonseasonal component is much greater than the others. So the variance reduction is limited in Louisiana's UNEMP estimation.

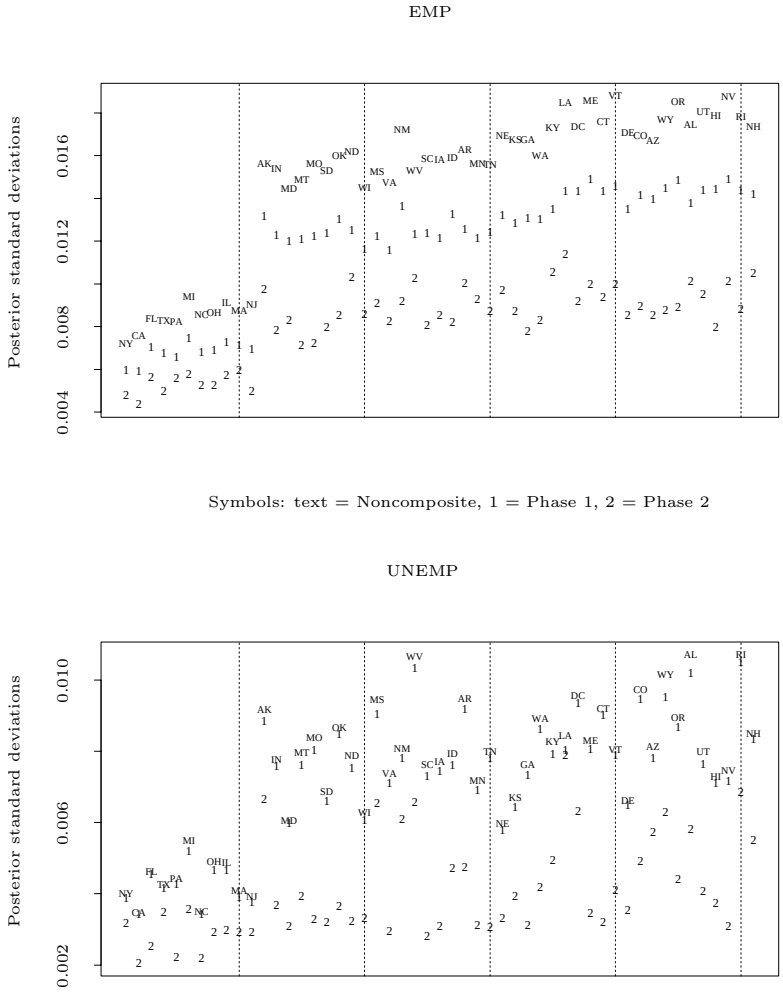

Figure 6. The posterior standard deviations of three estimates of Dec. 1989.

Although computing times and costs continue to improve rapidly, it remains important to face these issues when considering Monte Carlo simulation of Bayesian posteriors. Especially for the problem of small area estimates of labor force which need to be carried out each month, the use of approaches with heavy

computations has to be carefully evaluated. From the experience of analyzing the BLS data, we believe that the bureau can afford the needed computation of the proposed approach. When we have computed the posterior distributions of the parameters for the first time, we may obtain the results for the next month with less effort because of closeness of posteriors of parameters in two consecutive months. This will make the proposed approach more possible for practical application.

The choice of the fGn process as the building block of the time series models is strongly supported by a comparison of the periodograms of the data series with the corresponding spectral density of the fractional Gaussian model. This is a model construction technique, although it is not a usual procedure. Instead of frequency domain analysis, readers may be more familiar with an *a priori* ARMA time domain assumption. Model identification and diagnostic tools are then applied iteratively to fix a final model and make inferences. Among the alternative methods for comparison, Rao and Yu (1994) had a similar study on small area estimation for a model involving autocorrelated random effects and sampling errors. Their sampling errors are assumed to have an arbitrary covariance matrix over time, and they compute empirical BLUP estimates of small area means and associated standard errors for each period. Applying such alternative model and estimation approaches directly to the same data set may produce similar results to ours. But the comparison of results from different models on the same practical database requires more careful study, especially when Bayesian and classical contrasts are involved. Furthermore, it is hard to conclude that one approach is better than the other based on the results of modeling one data set.

Although no external comparison has been done, we have demonstrated how the fully Bayesian hierarchical approach proposed in this paper has reduced the standard errors of the noncomposite estimators of labor force in the 51 states. For easy implementation and computation, we have restricted the model to have the shape parameters d's assigned a single value chosen simply by comparing the posteriors of log likelihood of several d values. If the computing facilities were affordable for including the shape parameters in the model, the shape parameters would be more realistically estimated.

The techniques and empirical results presented in this paper support the recent trend at BLS to study and adopt time series models and related smoothing procedures for estimation at subnational levels. We recommend that research on Bayesian modeling and inference be expanded in order that consensus may be reached on methods that make close to optimal use of expensive data, and at the same time accurately assess standard errors of estimation. These methods should be used not only for the smaller states where estimation problems are

more acute, but also for the 11 "direct use" states where gains are smaller but still important. Note that, due to budgetary reductions, the CPS sample size was decreased to 50,000 households at the beginning of 1996, and, therefore, direct-use estimation of the 11 large states was discontinued.

Acknowledgement

The authors wish to thank an anonymous referee and an Associate Editor for their helpful comments, suggestions and for supplying references.

Appendix. Covariance of the finite Fourier transform

In this appendix we show some asymptotic properties of the covariances of finite Fourier transform of a real time series. It is similar to a discussion in Ramos (1988) for complex time series. First, we recall the definition of Lipschitz condition and a lemma in Ramos (1988).

Definition 1. A function f satisfies Lipschitz condition whenever there exists a smallest constant M_f such that

$$sup_{\omega,|h|<\epsilon}|f(\omega+h)-f(\omega)| \leq M_f \cdot \epsilon$$

for all $\epsilon > 0$. The definition is used to describe the "smoothness" of a function f. The constant M_f may be viewed roughly as the largest slope of f in the support of f. For the spectral density function f of a fractional Gaussian noise process with positive shape parameter d, the function is not too smooth for frequency near zero and therefore M_f tends to be large.

Lemma 1. *If D_n is the Dirichlet kernel, then*

$$\int_0^1 f(\omega)D_n(\omega-t)D_n(\omega-s)d\omega = D_n(s-t)f(s) + M_f \cdot O(\log n).$$

Recall the covariances σ_{ij}^c and σ_{ij}^s defined in (11). The structures of σ_{ij}^c and σ_{ij}^s are very similar. Here we derive the case for the cosine part. It is obtained directly from (10) and (11).

$$\begin{aligned}
\sigma_{ij}^c &= \frac{1}{nmq_iq_j}\Big\{\frac{1}{2}f(0)D_n(\frac{i}{n})D_n(\frac{j}{n}) \\
&\quad + \sum_{l=1}^{\frac{m}{2}-1} f(\frac{l}{m})[D_n(\frac{l}{m}-\frac{i}{n})+D_n(\frac{l}{m}+\frac{i}{n})][D_n(\frac{l}{m}-\frac{j}{n})+D_n(\frac{l}{m}+\frac{j}{n})] \\
&\quad + \frac{1}{2}f(\frac{1}{2})[D_n(\frac{1}{2}-\frac{i}{n})+D_n(\frac{1}{2}+\frac{i}{n})][D_n(\frac{1}{2}-\frac{j}{n})+D_n(\frac{1}{2}+\frac{j}{n})]\Big\} \\
&\xrightarrow{m\to\infty} \frac{1}{n}\int_0^{\frac{1}{2}} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega-\frac{j}{n})d\omega
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\int_0^{\frac{1}{2}} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega+\frac{j}{n})d\omega \\
&+\frac{1}{n}\int_0^{\frac{1}{2}} f(\omega)D_n(\omega+\frac{i}{n})D_n(\omega-\frac{j}{n})d\omega \\
&+\frac{1}{n}\int_0^{\frac{1}{2}} f(\omega)D_n(\omega+\frac{i}{n})D_n(\omega+\frac{j}{n})d\omega \\
= \quad &\frac{1}{n}\int_0^{1} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega-\frac{j}{n})d\omega \\
&+\frac{1}{n}\int_0^{1} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega+\frac{j}{n})d\omega.
\end{aligned}
$$

The final equation comes from the facts:

1. $f(\omega) = f(-\omega) = f(1-\omega)$
2. $D_n(\omega - \frac{i}{n}) = D_n(-\omega - \frac{i}{n}) = D_n(1-\omega-\frac{i}{n})$

Hence we have

$$
\begin{aligned}
&\int_0^{\frac{1}{2}} f(\omega)D_n(\omega+\frac{i}{n})D_n(\omega-\frac{j}{n})d\omega \\
=&\int_0^{\frac{1}{2}} f(1-\omega)D_n(1-\omega-\frac{i}{n})D_n(1-\omega+\frac{j}{n})d\omega \\
=&\int_{\frac{1}{2}}^{1} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega+\frac{j}{n})d\omega
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_0^{\frac{1}{2}} f(\omega)D_n(\omega+\frac{i}{n})D_n(\omega+\frac{j}{n})d\omega \\
=&\int_{\frac{1}{2}}^{1} f(\omega)D_n(\omega-\frac{i}{n})D_n(\omega-\frac{j}{n})d\omega.
\end{aligned}
$$

The asymptotic results come from the lemma and we have

$$
\begin{aligned}
\sigma_{ij}^c &\to \begin{cases} K_n * f(\frac{i}{n}) + \frac{1}{n}D_n(\frac{2i}{n})f(\frac{i}{n}) + M_f \cdot O(\frac{\log n}{n}) & \text{for } i = j \\ \frac{1}{n}[D_n(\frac{i-j}{n}) + D_n(\frac{i+j}{n})]f(\frac{i}{n}) + M_f \cdot O(\frac{\log n}{n}) & \text{for } i \neq j \end{cases} \\
&= \delta_{i-j} \cdot K_n * f(\frac{i}{n}) + M_f \cdot O(\frac{\log n}{n})
\end{aligned}
$$

where

$$
\delta_{i-j} = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j. \end{cases}
$$

References

Aitkin, M. (1991). Posterior Bayes factors. *J. Roy. Statist. Soc. Ser. B* **53**, 111-142.

Aitkin, M. (1997). The calibration of P-values, posterior Bayes factors and the AIC from the posterior distribution of likelihood (with discussion by A. P. Dempster and M. Stone). *Statist. Comput.* **7**, 253-272.

Bell, W. R. and Hillmer, S. C. (1990). The time series approach to estimation for repeated surveys. *Survey Methodology* **16**, 195-215.

Binder, D. A. and Dick, J. P. (1989). Modeling and estimation for repeated surveys. *Survey Methodology* **15**, 29-45.

Carlin, J. B. (1987). Seasonal analysis of economic time series. Ph. D. Thesis, Department of Statistics, Harvard University.

Carlin, J. B. and Dempster, A. P. (1989). Sensitivity analysis of seasonal adjustments: Empirical case studies (with disussion). *J. Amer. Statist. Assoc.* **84**, 6-32.

Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis.* Addison-Wesley, Reading, MA.

Dempster, A. P. (1982). Some formulas useful for covariance estimation with Gaussian linear component models. In *Statistics and Probability: Essays in Honor of C. R. Rao* (Edited by G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh), 213-220. North-Holland Publishing Company.

Dempster, A. P. and Hwang, J. S. (1993). Component models and Bayesian technology for estimation of state employment and unemployment rates (Disc: p582-189). In *Proc. Bur. Census Annual Res. Conf.* US Dept. Commerce, Bur. Census (Washington).

Dempster, A. P. and Hwang, J. S. (1994). Bayesian implementation of a complex hierarchical model. *Multivariate Analysis and its Applications*, IMS Lecture Notes - Monograph Series, Vol. **24**, 221-236.

Dempster, A. P. (1997). The direct use of likelihood for significance testing. *Statist. Comput.* **7**, 247-252.

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**, 269-277.

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

Hannan, E. J. (1970). *Multiple Time Series.* John Wiely, New York.

Hwang, J. S. (1992). Prototype Bayesian estimation of US state employment and unemployment rates. Ph. D. Thesis, Department of Statistics, Harvard University.

Hwang, J. S. (1996). Optimal recursive composite estimators in rotation designs. *J. Chinese Statist. Assoc.* **34**, 239-255.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierachical linear mixed models. *J. Amer. Statist. Assoc.* **91**, 1461-1473.

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**, 853-862.

Pfeffermann, D. (1992). Estimation and seasonal adjustment of population means using data from repeated surveys. *J. Business and Economic Statist.* **9**, 163-175.

Pfeffermann, D. and Bleuer, S. R. (1993). Robust joint modeling of labor force series of small areas. *Survey Methodology,* **19**, 149-163.

Pfeffermann, D. and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology* **16**, 217-237.

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**, 163-171.

Ramos, E. (1988). Resampling methods for time series. Ph. D. Thesis, Department of Statistics, Harvard University.

Rao, J. N. K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canad. J. Statist.* **22**, 511-528.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55**, 3-23.

Tiller, R. (1992) An application of time series methods to labor force estimation using CPS data. *J. Official Statist.*, 149-166.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: jshwang@stat.sinica.edu.tw

Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.

E-mail: dempster@stat.harvard.edu

(Received July 1996; accepted March 1998)