Submitted to the Annals of Applied Statistics
arXiv: math.PR/0000000

SUPPLEMENT TO "BAYESIAN NONPARAMETRIC MULTIRESOLUTION ESTIMATION FOR THE AMERICAN COMMUNITY SURVEY"*

By Terrance D. Savitsky†

U.S. Bureau of Labor Statistics

1. Posterior Computation. We implement the posterior computations for the predictor-indexed Dirichlet process mixture model (from which it is easy to derive the computations for the predictor-assisted mixture model), in a sequential scan of parameter blocks from their full conditional posterior distributions in the `growfunctions2` package for **R** (**R** Core Team 2014), which is written in `C++` for fast computation and available from the authors on request. The QCEW predictors used for this analysis, $(\mathbf{X}_\ell)$, include by county-month QCEW employment counts for all counties and municipal civil divisions (cmcds) in the U.S. There are approximately 100 out of 4751 cmcds in our dataset that are not publicly released by BLS due to privacy protection requirements for respondents. Our plan is to examine the possibility for imputing the values for these 100 cmcds under a model that excludes them. Further study within BLS will be required to assess privacy implications. The user may, alternatively, directly access QCEW values where published by BLS and perform the imputation on their own. We intend to publish tables that link cmcds to blocks and years to periods. These tables are used to construct label vectors input to the estimation function in `growfunctions2` to formulate the likelihood statement for each $f_{\ell j}$.

We briefly highlight aspects of our posterior sampling algorithm for the major sets of parameters, below:

1. Model for $\mathbf{Y} = (y_{bq})$
   (a) Sample each block of $P \times T$ random effect coefficients, $(\mathbf{B}_\ell)$, independently, using the elliptical slice sampler (ESS) of Murray et al. (2010) for block-sampling parameters under a multivariate Gaussian prior (that we generalized to matrix variate Gaussian distributions). The ESS generates $(P \times T)$ proposals through a convex combination of a draw from the prior and the previously sampled value. The proposals lie on the ellipse parameterized with a phase angle. The ESS uses a slice sampling algorithm (Neal 2000a) to draw proposals for the phase angle.

*U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E, Washington, D.C. 20212 USA
†Terrance Savitsky is Mathematical Research Statistician in the Office of Survey Methods Research.

Proposals are evaluated with the likelihood,

$$L\left(\mathbf{B}_{\ell}\right)=\prod_{b\in b(\ell)}\prod_{q\in q(b)} f\left(\tilde{y}_{bq,\ell}|\sum_{j\in q}\mathbf{x}_{\ell j}^{'}\boldsymbol{\beta}_{\ell j},\sigma_{bq}^{2}\right), \tag{1}$$

where $b(\ell)$ denotes the (usually multiple) blocks in which county ℓ is nested. Similarly, $q(b)$, denotes the often multiple (multi-year) periods, q, linked to block, b. We define $\tilde{y}_{bq,\ell}=y_{bq}-\sum_{\ell^{'}\neq\ell\in b}\sum_{j\in q}\mathbf{x}_{\ell^{'}j}^{'}\beta_{\ell^{'}j}$ to subtract out estimated functions for all other counties, $\left(\ell^{'}\right)\neq\ell$, which are also linked to y_{bq}.

(b) Sample the posterior distribution for locations of the GP covariance in by-cluster groups, $(\kappa_{dm}^{*})_{d=1,\ldots,D}$, from the subset of counties, $(\mathbf{B}_{\ell})$, assigned to that cluster because $\kappa_{dm}^{*}\perp\!\!\!\perp\kappa_{dm^{'}}^{*}$ for $m^{'}\neq m$, *a posteriori*, in a Metropolis-Hastings scheme using the following log-posterior kernel,

$$\begin{aligned}&\log f\left(\kappa_{dm}^{*}|\boldsymbol{\kappa}_{-dm}^{*},\mathbf{s},\boldsymbol{\Lambda}_{y,m}^{*},\{\mathbf{B}_{\ell}:s_{\ell}=m\}\right)\\&\propto-\frac{1}{2}n_{m}P\,\log\left(\left|\mathbf{C}\left(\kappa_{dm}^{*}\right)\right|\right)-\frac{1}{2}\mathrm{tr}\left[\sum_{\ell:s_{\ell}=m}\mathbf{C}\left(\kappa_{dm}^{*}\right)\mathbf{B}_{\ell}^{'}\boldsymbol{\Lambda}_{y,m}^{*}\mathbf{B}_{\ell}\right]\\&+(a-1)\log(\kappa_{dm}^{*})-b\kappa_{dm}^{*},\end{aligned} \tag{2}$$

where (a,b) are shape and rate hyperparameters of a gamma prior, respectively, which are both set equal to 1. This posterior representation is a relatively straightforward Gaussian kernel of a non-conjugate probability model.

We adapt a Metropolis-Hastings algorithm of Wang & Neal (2013) for sampling *each* κ_{dm}^{*} that is designed to speed computation by introducing a lower-dimensional temporary space where the likelihood (e.g. the $T\times T$, Gaussian process covariance matrix, $\mathbf{C}$) is approximated using a subset of the T time-points. We develop a transition / proposal distribution based on composing moves in the lower dimensional, temporary space (using a slice sampler), where computations of the lower-dimensional GP covariance matrix are fast. If the lower dimensional approximations are relatively good, this approach will speed chain convergence by producing draws of lower autocorrelation since each proposal includes a sequence of moves generated in the temporary space for drawing an equivalent effective sample size. See Savitsky (2015) for more details.

(c) Sample location, $\boldsymbol{\Lambda}_{y,m}^{*}$, from a P dimensional Wishart posterior with degrees of freedom, $n_{m}T+(P+1)$ and $P\times P$ inverse scale, $\sum_{\ell:s_{\ell}=m}\mathbf{B}_{\ell}\mathbf{C}(\boldsymbol{\kappa}_{m}^{*})\mathbf{B}_{\ell}^{'}+\mathbb{I}_{P}$.

(d) Sample cluster assignments, $\mathbf{s}=(s_{1},\ldots,s_{N})$, from their full conditionals using the Pólya urn representation, Blackwell & MacQueen (1973),

$$f\left(s_{\ell}=s|\mathbf{s}_{-i},\boldsymbol{\Theta}_{s}^{*},\alpha,\mathbf{B}_{\ell},\boldsymbol{\Delta}_{\ell}\right)\propto\begin{cases}\frac{n_{-\ell,s}}{n-1+\alpha}L\left(\mathbf{B}_{\ell},\boldsymbol{\Delta}_{\ell}\right) & \text{if } 1\leq s\leq M^{-}\\ \frac{\alpha/c^{*}}{n-1+\alpha}L\left(\mathbf{B}_{\ell},\boldsymbol{\Delta}_{\ell}\right) & \text{if } s=M^{-}+h,\end{cases} \tag{3}$$

where $n_{-\ell,s} = \sum_{\ell' \neq \ell} \mathbb{I}(s(\ell') = s)$ is the number of counties, excluding unit ℓ, assigned to cluster s, so that units are assigned to an existing cluster with probability proportional to its "popularity" and M^{-} denotes the total number of clusters when unit ℓ is removed (which is equal to M unless ℓ is a member of singleton cluster). The posterior assigns a county (through s_ℓ) to a new cluster with probability proportional to $\alpha d_0 = \int \mathcal{N}\left(\mathbf{B}|\boldsymbol{\kappa},\ldots\right) G_0(d\boldsymbol{\kappa})$, that requires the likelihood to be integrable in closed form with respect to the base distribution, which is not the case under our non-conjugate parameterization through the GP covariance matrix. So we utilize the auxiliary Gibbs sampler formulation of Neal (2000*b*) and sample $c^* \in \mathbb{N}$ (typically set equal to 2 or 3) locations from base distribution, G_0, *ahead* of any assigned observations, to define $h = M^{-} + c^*$ candidate clusters in an augmented space. We then draw s_ℓ from this augmented space, where any location not assigned units (over a set of draws for $\mathbf{s}$) is dropped.

2. Model for $\mathbf{X}_\ell = (\mathbf{x}_{\ell j})$
   (a) Sample $P \times T$, $\boldsymbol{\Delta}_\ell$, independently, by stacking the transpose of the P ,$T \times 1$ rows of $\boldsymbol{\Delta}_\ell$ to form the $PT \times 1$, $\boldsymbol{\delta}_{v,\ell}$, from which we perform a draw from the following conjugate Gaussian posterior,

$$f\left(\boldsymbol{\delta}_{v,\ell}|\mathbf{X}_\ell, \mathbf{H}_x, \mathbf{s}, \boldsymbol{\Lambda}^*_{x,s_\ell}, \mathbf{Q}\left(\tau^*_{x,s_\ell}, \rho^*_{x,s_\ell}\right)\right) = \mathcal{N}_{PT}\left(\mathbf{h}_\delta, \boldsymbol{\phi}_\delta^{-1}\right), \tag{4}$$

   where we define $PT \times 1$, $\mathbf{e}_\delta = \mathbf{H}_{x,T}\mathbf{x}_{v,\ell}$, with $\mathbf{H}_{x,T} = (\mathbf{H}_x \otimes \mathbb{I}_T)$, while $\mathbf{x}_{v,\ell}$ is formed by stacking the transpose of the rows of $\mathbf{X}_\ell$. Posterior precision, $\boldsymbol{\phi}_\delta = \mathbf{H}_{x,T} + \left(\boldsymbol{\Lambda}^*_{x,s_\ell} \otimes \mathbf{Q}\left(\tau^*_{x,s_\ell}, \rho^*_{x,s_\ell}\right)\right)$. Finally, compose $\mathbf{h}_\delta = \boldsymbol{\phi}_\delta^{-1}\mathbf{e}_\delta$.
   (b) Sample the location parameters, $\left(\tau^*_{x,m}\right)$, of the $T \times T$ CAR precision matrix, $\mathbf{Q}$, from the Gamma distribution,

$$f\left(\tau^*_{x,m}|(\boldsymbol{\Delta}_\ell : s_\ell = m), \rho^*_{x,m}\right) = \mathcal{G}a\left(a_1, b_1\right), \tag{5}$$

   with shape, $a_1 = 0.5 n_m TP + a$, and rate, $b_1 = 0.5\text{tr}\left[\sum_{\ell:s_(\ell)=m} \mathbf{R}^*_m \boldsymbol{\Delta}'_\ell \boldsymbol{\Lambda}^*_{x,m} \boldsymbol{\Delta}_\ell + b\right]$, where $\mathbf{R}^*_m = (\mathbf{D}_x - \rho^*_m \boldsymbol{\Omega}_x)$.
   Next, sample $\rho^*_{x,m}$ using a slice sampler with the following posterior evaluation kernel,

$$\begin{aligned} &\log f\left(\rho^*_{x,m}|(\boldsymbol{\Delta}_\ell : s_\ell = m), \tau^*_{x,m}\right) \\ &\propto 0.5 n_m P \log|R^*_m| + 0.5\tau^*_{x,m}\rho^*_{x,m}\text{tr}\left[\sum_{\ell:s_\ell=m} \boldsymbol{\Omega}_x \boldsymbol{\Delta}'_\ell \boldsymbol{\Lambda}^*_{x,m} \boldsymbol{\Delta}_\ell\right]. \end{aligned} \tag{6}$$

2. Posterior Mixing. Figure 1 presents post-burnin MCMC trace plots for randomly-selected $(f_{\ell j})$ for counties of varying-sized populations. The mixing appears to be very good and stable.

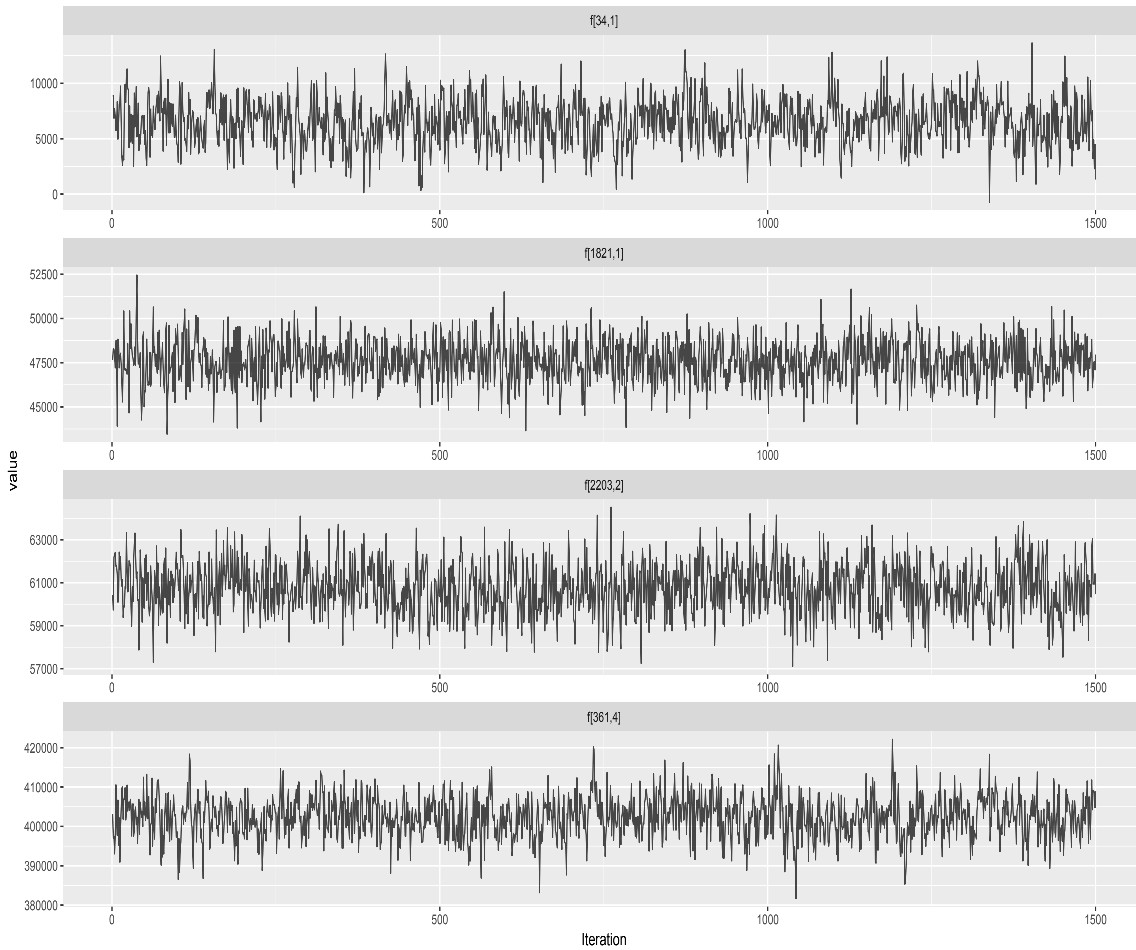

Fig 1: Post-burnin trace plots for randomly selected latent functions, $(f_{\ell j})$.

3. Simulation Study Results for 5− year Counties. Similarly to the 3− year county result, Figure 2 presents typical results for a county with only a single, 5− year estimate available in the case where that county is nested in a group relatively near to it in size. As earlier mentioned, this situation is typical for MCD's, which by construction (in New England) are nested within counties. While we see that the fitted result expresses more smoothness than the truth, it does generally follow local features in the true trend and the credible interval is wider than those for counties with published 3− year period estimates. We note that the bottom end of the credible interval in Figure 2 crosses below zero. All of the posterior means and modes are greater than 0. In the ACS application, 10 of the 4751 counties express credible intervals that cross 0 for the smallest-sized, far-nested 5− year counties, though again all posterior means and modes are greater than 0. It is generally more difficult to stably estimate tail quantiles of posterior distributions than the mean, median or mode. Adding data for upcoming years will bring in additional 5− year period estimates, which are expected to improve the quality of estimation for these far-nested counties by borrowing strength over (overlapping) multi-year periods.

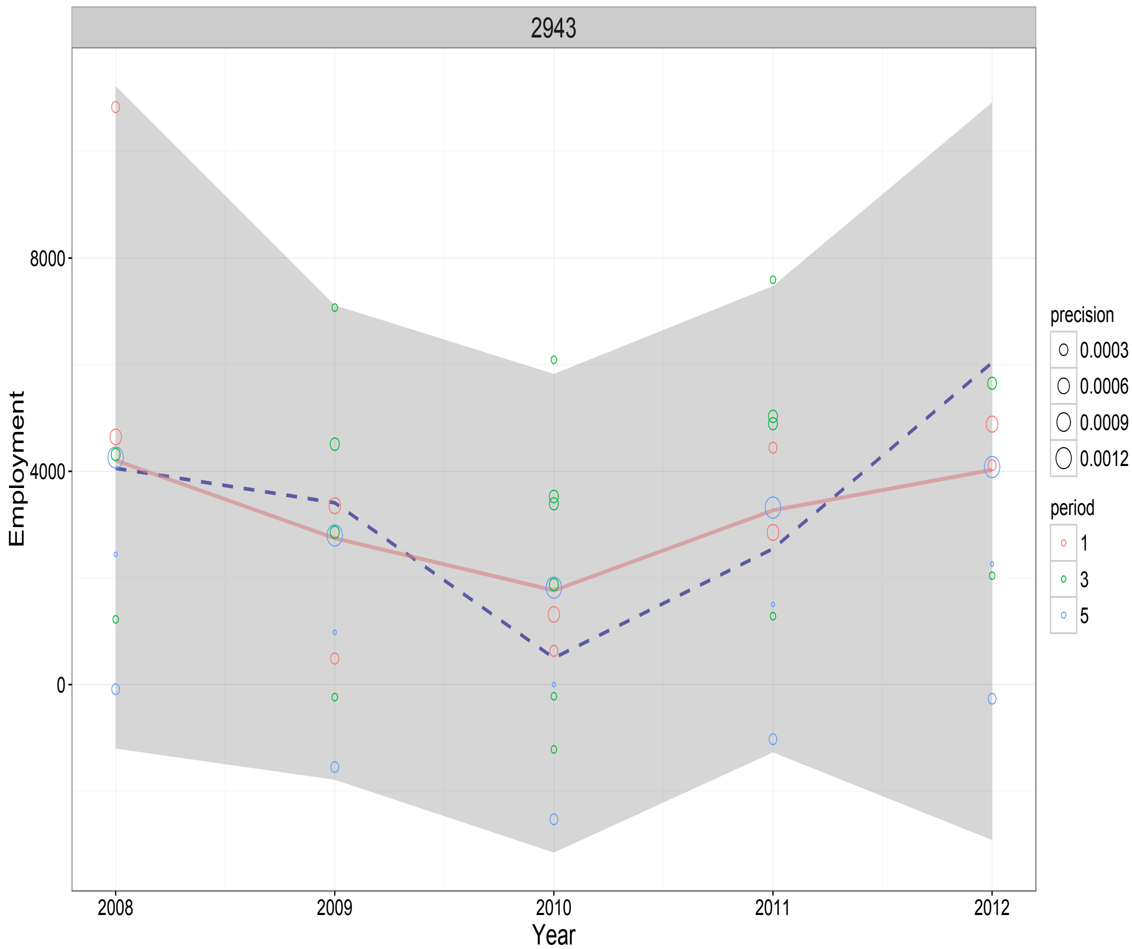

Fig 2: Fitted versus Data values for simulated 5− year county linked to one or more groups of similar size.

References.

Blackwell, D. & MacQueen, J. B. (1973), 'Ferguson distributions via Pólya urn schemes', *The Annals of Statistics* **1**, 353–355.

Murray, I., Adams, R. P. & MacKay, D. J. (2010), 'Elliptical slice sampling', *Journal of Machine Learning Research: Workshop & Conference Proceedings* **9**, 541–548.

Neal, R. (2000*a*), 'Slice sampling', *The Annals of Statistics* **31**, 705–767.

Neal, R. M. (2000*b*), 'Markov chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**(2), 249–265.

R Core Team (2014), ***R**: A Language and Environment for Statistical Computing*, **R** Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Savitsky, T. D. (2015), 'Bayesian Nonparametric Functional Mixture Estimation for Time-Series Data, With Application to Estimation of State Employment Totals', *ArXiv e-prints* .
URL: *http://adsabs.harvard.edu/abs/2015arXiv150800615S*

Wang, C. & Neal, R. M. (2013), 'MCMC methods for Gaussian Process models using fast approximations for the likelihood'.
URL: *http://arxiv.org/abs/1305.2235v1*

2 Massachusetts Ave. N.E, Washington, D.C. 20212 USA, E-mail: Savitsky.Terrance@bls.gov