Dependent Latent Effects Modeling for Survey Estimation with Application to the Current Employment Statistics Survey[1]

Julie Gershunskaya, Terrance Savitsky
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Suite 4985, Washington, DC, 20212

Abstract

The Current Employment Statistics Survey, administered by the U.S. Bureau of Labor Statistics, publishes total employment estimates for thousands of domains at detailed geographical and industrial levels. Some of these domains do not have adequate sample size for the direct probability sample-based estimates to be reliable. Small area estimation methods are used to integrate information from historical sources and correlated domains to improve estimation efficiency. In this paper, we explore alternatives to the Fay-Herriot two-stage hierarchical model that relax distributional and independence assumptions among random effects indexed by domain and month in order to more fully borrow strength to improve the efficiency of published employment estimates. We compare the performances of our alternative models on both synthetic data and in application to estimates from the Current Employment Statistics survey.

Key Words: Bayesian Hierarchical Modeling, Conditional Autoregressive, Dirichlet process, Probabilistic clustering, Fay-Herriot, Rao-Yu, Variational Bayes, Stan

1. Introduction

Complex surveys usually produce reliable estimates for given sets of population targets. However, at more detailed levels than planned at the design stage, domain-indexed sample sizes may be small, thus rendering the purely sample-based direct domain estimates that are erratic. Model-based approaches to improve estimates in such "unplanned" domains form a wide collection of techniques within Small Area Estimation (SAE) methods (see Pfeffermann, 2002, 2013, Rao and Molina, 2015, for review of SAE).

We consider application of small area modeling techniques to employment estimates from the Current Employment Statistics (CES) survey, which is administered by the U.S. Bureau of Labor Statistics (BLS). CES is a large-scale establishment survey that publishes monthly estimates of employment and other indicators of the US economy at the national total level, as well as for thousands of domains at detailed geographical and industrial levels. Sample is scarce in many of these domains, so that direct sample-based estimates are volatile and unreliable, and model-based estimation is required.

The classical Fay-Herriot model (Fay and Herriot, 1979) has been successfully applied for the CES estimates at the States' major industry levels (supersectors)[2]. In this paper, we

[1] Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

[2] Although the model is fitted using data from all States, the model-based estimates are published only for select States that have small sample in a given supersector.

study estimation at detailed industry levels, where the Fay-Herriot model assumptions may not hold. While striving to maintain the general simplicity and computational scalability of the two-level model, we introduce a suite of alternative formulations: each proposed model relaxes and generalizes some of the Fay-Herriot assumptions.

The Fay-Herriot model may be viewed as a special and relatively simple case of the general linear mixed model. In this sense, the multitude of existing models falling into the class of the general linear mixed models can be regarded as generalizations of the Fay-Herriot model, such as the time series and cross-sectional models of Pfeffermann and Burck (1990), Rao and Yu (1994), Datta et al. (2002) (see Rao and Molina, 2015, for more examples, including the multivariate Fay-Herriot and spatial models.)

We adopt the Bayesian paradigm for model development and view the Fay-Herriot model as a hierarchical Bayes formulation. Our approach develops a sequence of related models, where each model relaxes some assumptions of the base Fay-Herriot model, thus providing a study of how progressively relaxing distributional and independence assumptions of the Fay-Herriot model may improve estimation efficiency. Our models are related by uses of similarly constructed prior distributions for relaxing independence and distributional assumptions to promote ease-of-comprehension and comparison. The Bayesian approach we choose provides associated variance measures for our model-based estimators that account for the uncertainties in the parameter estimates.

Our models are implemented in Stan modeling language (Gelman et al., 2015) using a Variational Bayes (Kucukelbir et al. 2016) algorithm. The Stan code is easy to read and implement. See the computation Section 4.1 for discussion and details.

We begin exposition of our models by first observing that the distribution of the residuals for our CES data expresses heavy tails, so we relax the assumption of normality of the sampling errors by replacing it with the Student's t distribution (which is also considered by Huang and Bell, 2006).

We proceed by further relaxing the assumption of normality of area random effects (where the random effect is a latent response that we extract as denoised estimates) by replacing the normal distribution prior with a finite mixture of the normal distributions. In the case of CES data, it is reasonable to expect that subgroups of industries or localities to express diverging employment trends. If those differences are clear cut, then it may be appropriate to include this information in the model; however, more typically there is no a priori information for defining and labeling the subgroups, since underlying economic drivers of employment continue to evolve. Thus, by assuming a mixture distribution prior, we let the data determine the probabilities of belonging to various clusters for each of the participating domains.

We further consider several multivariate versions of the models, thus borrowing information both cross-sectionally (as is in the Fay-Herriot case) and over time. Rao and Yu (1994), among others, considered similar setup. Their assumptions are different from our approach, most notably in the area random effects distribution, where our model allows for clustered random effects. Similar to the Rao-Yu model, the multivariate formulation also captures autocorrelations in sampling errors.

Our models are applied to the real CES data and results are evaluated using comparison to the external "gold standard" derived from the Quarterly Census of Employment and Wages (QCEW), another BLS program. While timeliness of the estimates is a special feature of the CES survey, QCEW is an important source that provides a census of employment based on the administrative data. The census counts from QCEW become available to the public

on a lagged (6 – 9 month) basis. Due to its administrative nature and the quarterly reporting pattern, QCEW is affected by measurement error, although it does not contain sampling error as does CES. Despite these drawbacks, QCEW plays an important role for the CES survey, providing it with the sampling frame and the annual benchmark levels, as well as the historical series that can be used for research.

Besides using QCEW levels for comparison with the modeling results, we also investigate properties of the models based on simulations. We construct synthetic data by adding noise to the QCEW series, i.e., by treating the QCEW series as truth not affected by the sampling error.

In Section 2, we describe features of the CES estimation. Models are introduced in Section 3. Our models are applied to both CES and synthetic data in Section 4. We conclude with a summary discussion in Section 5.

2. CES Data Construction

Estimates of employment from the CES survey are published every month at the national level (across industries defined by the North American Industry Classification System (NAICS)) as well as for a multitude of domains constructed from combinations geographies and sub-industries.

We now briefly describe how the CES data are transformed to perform estimation. For a given month, t, the target of the CES estimation is the change in employment from the previous to current month, for various domains, $i = 1, \ldots, N$ defined by intersections of industry and geography. The ratio, $R_{i,t}$, constructs the target employment change, specified with,

$$R_{i,t} = \frac{Y_{i,t}}{Y_{i,t-1}}. \tag{1}$$

The associated sample based estimator of the relative change is

$$\hat{r}_{i,t} = \frac{\sum_{j \in s_t^{(i)}} w_j y_{jt}}{\sum_{j \in s_t^{(i)}} w_j y_{j,t-1}}, \tag{2}$$

where y_{jt} is the employment of business j at time t, w_j is the sampling weight of unit j, and $s_t^{(i)}$ is a set of units sampled in domain i that respond and provide positive employment inputs in both previous and current months to provide a "matched" set of respondents. The presence of matched sets of sampled units is typically high from one month to another but there are also unmatched units. In particular, monthly sets of respondents do not include neither emerging ("births") nor the discontinued ("deaths") businesses. The sample estimate $\hat{r}_{i,t}$ expresses a slight bias due to its inability to account for births and deaths. To obtain the estimate of employment level, the following formula is used that includes the bias correction term $\hat{N}_{i,t}$:

$$\hat{Y}_{i,t} = \hat{Y}_{i,t-1}\hat{r}_{i,t} + \hat{N}_{i,t}, \tag{3}$$

where $\hat{Y}_{i,t}$ is the estimate of employment level in domain *i* at month *t*; adjustment $\hat{N}_{i,t}$ is the projected net difference of employment added by births and lost from deaths of businesses in domain i (in the current paper, the set of adjustments $\hat{N}_{i,t}$ is considered fixed and known; for more detail on the net birth-death adjustment see BLS Handbook of Methods 2004.)

The estimated relative change in employment level is

$$\hat{R}_{i,t} = \frac{\hat{Y}_{i,t}}{\hat{Y}_{i,t-1}}. \tag{4}$$

We construct expression (3) to produce estimated employment as,

$$\hat{Y}_{i,t} = Y_{i,0} \prod_{\tau=1}^{t} \hat{R}_{i,\tau}, \tag{5}$$

where $Y_{i,0}$ is a known "benchmark" employment level at month 0, available on a lagged basis from the Quarterly Census of Employment and Wages (QCEW). Every year, the estimation cycle starts at month 0 and after twelve months the CES estimated employment level $\hat{Y}_{i,12}$ is compared to the census data, QCEW, that become available with a lag of about 6 to 9 months. At that point, the levels are revised (in a procedure commonly known as the annual revision), and a new cycle of estimation starts with this true census level as the new month 0. Employment seasonal patterns in the QCEW are affected by the quarterly submission of administrative data provided by units (business establishments). CES estimates are unaffected by this quarterly seasonal influence due to a monthly submission cycle. So we may not compare monthly QCEW and CES estimates, which is discussed, in detail, in Section 4. Nevertheless, the annual levels from QCEW are consider "gold standard" and the quality of the CES employment estimates of levels are judged based on the size of the annual revision that benchmark to the QCEW.

To summarize, each month we are interested in estimating the ratios $R_{i,t}$ defined by (1). The estimators $\hat{R}_{i,t}$ are calculated directly from the sample data using formulas (2), (3), and (4). These ratios, along with their respective sampling variances $v_{i,t}$, constitute the domain-level data supplied for the modeling. The domain-level auxiliary information used in the models is the employment ratio, obtained as a forecast from the historical QCEW series. For this paper, we used five-year averages of historic QCEW ratios $R_{i,t}^{H} = \sum_{m=1}^{5} R_{i,(t-12m)} \Big/ 5$ as auxiliary information.

Before fitting the models, we standardized the input values as follows: $y_{i,t} = \left(\hat{R}_{i,t} - \bar{R}\right) \Big/ \sqrt{\bar{v}}$, $x_{i,t} = \left(R_{i,t}^{H} - \bar{R}^{H}\right) \Big/ \sqrt{\bar{v}}$, and $\sigma_{i,t}^{2} = v_{i,t} \big/ \bar{v}$, where $\bar{R}$, $\bar{R}^{H}$, and $\bar{v}$ are the average values of respective original data $\hat{R}_{i,t}$, $R_{i,t}^{H}$, and $v_{i,t}$. To obtain the model estimate $\tilde{R}_{i,t}$, we perform a back transformation.

After we obtain the model-based estimate $\tilde{R}_{i,t}$ of $R_{i,t}$, the estimate of the employment level $\tilde{Y}_{i,t}$ is constructed by successively multiplying the estimates $\tilde{R}_{i,t}$ by the starting level, similar to (5) and as shown in (37) in Section 4.2 below.

As mentioned in the introduction, the Fay-Herriot model is currently implemented in CES for State-level estimates for industry supersectors. The model is fitted separately for each supersector. The model-based estimates are published only for selected States that have small samples in a given supersector.

The current paper describes the endeavor in extending the model for application to the detailed industrial series within States.

3. The proposed models

Our goal is to explore the modeling of domains defined at more detailed levels. Some assumptions of the Fay-Herriot model employed for those broader (coarser) level industry domains may not hold at the lower, sub-industry levels. The different structure of the detailed-level data potentially requires different modeling formulations to more efficiently borrow information across domains as well as among repeated monthly measures. Candidate models must, however, remain computationally scalable to support the tight monthly production process and be relatively straightforward for ease-of-explanation to BLS customers. We, therefore, restrict ourselves to the basic construction of an observed process and a latent response process constructed by the Fay-Herriot.

The available data for each domain $i = 1,...,N$ consist of the direct survey estimate y_i with the sample variance σ_i^2 and the auxiliary information x_i (a vector, in general.) For the multivariate models, the added dimension is denoted by the time index, $t = 1,...,T$.

Note that, although the sample variances are known only approximately, they are customarily treated as fixed and known and supplied as part of the data.

For modeling, the data is standardized as described above in Section 2.

3.1 Model 1 (FH): The Fay-Herriot Model

The first model considered is the usual Fay-Herriot model.

For each area $i = 1,...,N$, assume the following cross-sectional two-level model:

$$y_i = \theta_i + e_i, \tag{6}$$

$$\theta_i = \beta_0 + \beta_1 x_i + u_i , \tag{7}$$

where the sample errors are assumed to be independent and normally distributed with mean zero and variances σ_i^2 ,

$$e_i \overset{ind}{\sim} N\left(0, \sigma_i^2\right). \tag{8}$$

Random effects are independent and identically normally distributed variables with mean zero and variance τ^2 ,

$$u_i \mid \tau^2 \overset{iid}{\sim} N\left(0, \tau^2\right). \tag{9}$$

Parameters β_0 τ, $_1\beta$,^{2} are unknown and need to be estimated from the data.

3.2 Model 2 (FHt): Student's t-distributed sampling errors

In the Fay-Herriot model, an important assumption stated in (8) is that the sampling errors are normally distributed. The normality assumption may not hold, especially when the domain sample sizes are small, thus precluding invocation of the Central Limit Theorem. In addition, the ratio form of the direct survey estimator (4) suggests that survey errors have heavier tails than those of the normal distribution. Indeed, for the estimator used in CES, it has been observed that, while the distribution of the survey errors is nearly symmetric, the tails of the distribution are much heavier than under the normality (see Figure 1.) Therefore, we consider a modification of the Fay-Herriot model by assuming the Student's t distribution over the sampling errors. We introduce latent parameters δ_i to model the t distribution as the scale mixture of the normal distributions (Huang and Bell, 2006, considered similar assumptions).

In summary, Model 2 keeps assumptions (6), (7), and (9) and replaces assumption (8) by (10) and (11):

$$e_i \overset{ind}{\sim} N\left(0, \sigma_i^2 \delta_i\right); \tag{10}$$

parameter δ_i has the inverse gamma distribution with the scale and shape parameters $\nu/2$:

$$\delta_i \sim InvGamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{11}$$

where parameter ν represents the degree of freedoms of the Student's t distribution. Under this construction, integrating over the random effects δ_i induces a marginal t distribution for the errors. (The integration is performed numerically under our estimation procedure discussed in Section 4).

3.3 Model 3 (CFHt): Clustered random effects

Assumption (9) that random effects are independent and identically distributed as normal may not hold.

Sub-industry and sub-State domain CES employment estimates would be expected to be correlated due to underlying similarities in economic factors. That dependence may be captured by a (probabilistic) clustering within each State or larger industry classification. To account for the possibility of dependence across domains, our next model allows for clustering among them by replacing the assumption of independent and identically distributed normal random effects with a mixture of the normal distributions.

Namely, assume the following Model 3:

$$y_i = \theta_i + e_i, \tag{12}$$

$$\theta_i = \beta x_i + u_i, \tag{13}$$

$$e_i \overset{ind}{\sim} N\left(0, \sigma_i^2 \delta_i\right), \tag{14}$$

$$\delta_i \sim InvGamma\left(\frac{\nu}{2},\frac{\nu}{2}\right). \tag{15}$$

Assumptions (12)-(15) are similar to (6), (7), (10), (11), except that the linear model (13) has no intercept term; instead, the distribution of random effects is parameterized with the location parameter vector $\boldsymbol{\mu} = \left(\mu_1,...,\mu_K\right)$.

Random effects u_i assume a K-component mixture of normal distributions with equal variances τ_u^2 and location parameter $\boldsymbol{\mu} = \left(\mu_1,...,\mu_K\right)$:

$$u_i \overset{iid}{\sim} \sum\nolimits_{k=1}^{K} \theta_k N\left(\mu_k, \tau_u^2\right), \tag{16}$$

where θ_k is the probability of belonging to the mixture part k, $k = 1,...,K$; $\sum\nolimits_{k=1}^{K} \theta_k = 1$. Probabilities $\boldsymbol{\theta} = \left(\theta_1,...,\theta_K\right)$ are given a symmetric Dirichlet prior

$$\boldsymbol{\theta} \sim Dirichlet\left(\frac{\alpha}{K},...,\frac{\alpha}{K}\right). \tag{17}$$

Under (17), as the number of clusters K goes to infinity, the concentration parameter α/K approaches zero, and model (16) for the random effect approaches the Dirichlet process mixture model (see Neal 2000). In practice, the maximum number of mixture components K is supplied by the modeler and can be set to some number larger than the number of components expected in the data. As K increases, the inverted bell-shaped form of the Dirichlet distribution becomes more pronounced, thus penalizing complexity by encouraging sparsity in the number of non-zero mixture components estimated by the data. This construction allows the data to learn the number of true clusters or mixture components.

3.4 Model 4 (mFHt): Multiple time points, independent sampling errors

The Fay-Herriot model is simple and easy to implement in the CES production environment. While we would like to avoid unduly complicated models, we weigh the potential improvement in estimation efficiency from extending our modeling options to include multiple time points. This modeling extension to multiple time points would allow borrowing information not only over domains for a given month but also across months.

To formulate the multiple time points model, we introduce additional index $t = 1,...,T$. The assumptions for Model 4 are:

$$y_{i,t} = \theta_{i,t} + e_{i,t}, \tag{18}$$

$$\theta_{i,t} = \beta_t x_{i,t} + u_{i,t}, \tag{19}$$

$$e_{i,t} \overset{ind}{\sim} N\left(0, \sigma_{i,t}^2 \delta_{i,t}\right), \tag{20}$$

$$\delta_{i,t} \sim InvGamma\left(\frac{\nu}{2},\frac{\nu}{2}\right). \tag{21}$$

Assumptions (18)-(21) are similar to the assumptions of the FHt model.

Assume a proper *conditional autoregressive* (CAR) structure for the precision matrix of multivariate Gaussian priors on the random effects and model coefficients:

$$\mathbf{u}_i \overset{iid}{\sim} \mathbf{N}_T\left(\boldsymbol{\mu}, \mathbf{Q}_u^{-1}\right),\ \mathbf{Q}_u = \lambda_u\left(\mathbf{D} - \rho_u \boldsymbol{\Omega}\right), \tag{22}$$

$$\boldsymbol{\beta} \overset{iid}{\sim} \mathbf{N}_T\left(\mathbf{0}, \mathbf{Q}_\beta^{-1}\right),\ \mathbf{Q}_\beta = \lambda_\beta\left(\mathbf{D} - \rho_\beta \boldsymbol{\Omega}\right). \tag{23}$$

In the above formulas, symbol $\mathbf{N}_T$ signifies the multivariate T-dimensional Gaussian distribution.

The T x T adjacency matrix $\boldsymbol{\Omega}$ specifies the dependence structure. It has nonzero elements in positions (i, j) wherever domain i is the "neighbor" of domain j and 0's in all other places. In our model, 1's appear in the first off-diagonals of the adjacency matrix to capture time-based dependence on a line,

$$\boldsymbol{\Omega} = \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & 0 & \dots & \\ & & \dots & \dots & 1 \\ & & & 1 & 0 \end{pmatrix}.$$

The entries of the diagonal matrix $\mathbf{D}$ depend on the number of "neighbors" specified in matrix $\boldsymbol{\Omega}$; namely, each entry in the diagonal equals the sum of the row elements of $\boldsymbol{\Omega}$. In our first-order CAR model, the first and last months of estimation each have 1 less "neighbor" than the other months (which each have a preceding and following month as neighbors). This difference in the number of neighbors is encoded in the diagonal of matrix $\mathbf{D}$, where the first and last entries are 1's and the rest are 2's. The implication is that the estimated precisions for the first and last months are lower compared to other months:

$$\mathbf{D} = \begin{pmatrix} 1 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & \dots & \\ & & & & 1 \end{pmatrix}.$$

Thus, precision matrices $\mathbf{Q}_u$ and $\mathbf{Q}_\beta$ in formulas (22) and (23) are the tridiagonal matrices with zero entries outside the diagonal and first off-diagonals. Parameters λ_u, λ_β are precision parameters that determine the scale of the CAR prior and ρ_u, ρ_β represent the strength of correlations between the adjacent months, and also serve to ensure that the precision matrix is of full rank (Jin et al., 2005, discuss the structure of the precision matrices.) The support of these strength of correlation parameters is between 0 and 1, where a value near 0 indicates there is no dependence and the prior on the random effects and regression coefficients revert to independent normals with precisions controlled by the number of neighbors and the precision parameters. As the strength of correlation

parameters approaches 1, the autocorrelation increases. These parameters receive uniform priors $\rho_u \sim U(0,1)$ and $\rho_\beta \sim U(0,1)$, which are updated by the data.

The same specifications for T x T $\mathbf{\Omega}$ and $\mathbf{D}$ matrices are used in all remaining models that follow in the sequel.

3.5 Model 5 (mFHtc): Multiple time points, correlated sampling errors

In Model 4 above, we assume existence of a dependence structure among the T repeated measures in time for each latent response (indexed by domain), which we model under a CAR multivariate Gaussian prior. The sampling errors are, however, assumed to be independent in Model 4. The sampling errors may express correlation, however, since the sets of sampled units that compose each domain largely overlap across the set of repeated measures. In the case of CES, previous research (Scott et al., 2012, Scott and Sverchkov, 2005, Gershunskaya, 2015.) shows that there is small negative correlation between the sample estimates of relative employment change in adjacent months. The negative sign of the correlation is related to the construction of the form for the estimator as the ratio of highly correlated estimates of employment levels.

Briefly, let $\hat{Y}_{i,t} = Y_{i,t} + \eta_{i,t}$, where $\eta_{i,t}$ represents the sampling error for the estimate of level in domain i at month t. Suppose sampling errors follow autoregressive scheme AR(1) with autocorrelation ρ_η, and let the variance V_i be approximately the same for months that are close in time. For simplicity, consider estimates of monthly changes, $\Delta\hat{Y}_{i,t} = \hat{Y}_{i,t} - \hat{Y}_{i,t-1}$ (the values of ratios of employment estimates are in the neighborhood of 1.) The variance of the change estimate is $Var\left(\Delta Y_{i,t}\right) \approx 2V_i \left(1-\rho_\eta\right)$, the covariance is $Cov\left(\Delta\hat{Y}_{i,t}, \Delta\hat{Y}_{i,t+1}\right) \approx -\left(1-\rho_\eta\right)^2 V_i$. Thus, correlation between the sampling errors of changes for the adjacent months is $-\frac{1}{2}\left(1-\rho_\eta\right)$; i.e., when ρ_η is close to 1, correlation between estimates of changes is small and negative.

Model 5 was constructed from Model 4 by modifying assumption (20) of independence of sampling errors assumptions.

$$y_{i,t} = \theta_{i,t} + e_{i,t}\left(\sigma_{i,t}^2 \delta_{i,t}\right)^{1/2}, \tag{24}$$

$$\theta_{i,t} = \beta_t x_{i,t} + u_{i,t}, \tag{25}$$

$$\mathbf{e}_i = \left(e_{i,1},...,e_{i,T}\right) \overset{iid}{\sim} \mathbf{N}_T\left(\mathbf{0}, \mathbf{Q}_e^{-1}\right), \qquad \mathbf{Q}_e = \mathbf{D} - \rho_e \mathbf{\Omega}, \tag{26}$$

$$\delta_{i,t} \sim InvGamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right), \tag{27}$$

$$\mathbf{u}_i \overset{iid}{\sim} \mathbf{N}_T\left(\boldsymbol{\mu}, \mathbf{Q}_u^{-1}\right),\ \mathbf{Q}_u = \lambda_u\left(\mathbf{D} - \rho_u \mathbf{\Omega}\right), \tag{28}$$

$$\boldsymbol{\beta} \overset{iid}{\sim} \mathbf{N}_T\left(\mathbf{0}, \mathbf{Q}_\beta^{-1}\right),\ \mathbf{Q}_\beta = \lambda_\beta\left(\mathbf{D} - \rho_\beta \mathbf{\Omega}\right). \tag{29}$$

Statement (26) imposes the CAR correlation structure model for sampling errors. In this case, we set the precision parmeter equal to 1 since the known variances provide the scale. Strength-of-correlation parameter ρ_e is assigned prior $\rho_e \sim U(-1,0)$ based on the previous work that discovered a negative correlation of ratio survey estimates in adjacent months. We set the support of the strength of correlation parameter for the latent random effects ρ_u to $(0,1)$ under our assumption of adjacency-dependence among the multiple latent response values over the 12 months for each domain. By setting the support range for parameters ρ_e and ρ_u to not overlap, we establish their mutual identifiability.

3.6 Model 6 (mCFHtc): Multiple time points, clustered random effects and correlated sampling errors

The last model considered may be viewed as the multivariate generalization of Model 3. We assume that random effects come from a mixture of multivariate normal distributions. The other assumptions of Model 6 are the same as in Model 5.

$$y_{i,t} = \theta_{i,t} + e_{i,t}\left(\sigma_{i,t}^2 \delta_{i,t}\right)^{1/2}, \tag{30}$$

$$\theta_{i,t} = \beta_t x_{i,t} + u_{i,t}, \tag{31}$$

$$\mathbf{e}_i = \left(e_{i,1},\ldots,e_{i,T}\right) \overset{ind}{\sim} \mathbf{N}_T\left(\mathbf{0},\mathbf{Q}_e^{-1}\right), \qquad \mathbf{Q}_e = \mathbf{D} - \rho_e \mathbf{\Omega}, \tag{32}$$

$$\delta_{i,t} \sim InvGamma\left(\frac{\nu}{2},\frac{\nu}{2}\right). \tag{33}$$

Assume that random effects come from the mixture distribution:

$$\mathbf{u}_i \overset{iid}{\sim} \sum_{k=1}^{K} \theta_k \mathbf{N}_T\left(\boldsymbol{\mu}_k,\mathbf{Q}_{uk}^{-1}\right),\ \mathbf{Q}_{uk} = \lambda_{uk}\left(\mathbf{D} - \rho_{uk}\mathbf{\Omega}\right), \tag{34}$$

with

$$\boldsymbol{\mu}_k \overset{iid}{\sim} \mathbf{N}_T\left(\mathbf{0},\mathbf{Q}_\mu^{-1}\right),\ \mathbf{Q}_\mu = \lambda_\mu\left(\mathbf{D} - \rho_\mu \mathbf{\Omega}\right), \tag{35}$$

$$\boldsymbol{\beta} \overset{iid}{\sim} \mathbf{N}_T\left(\mathbf{0},\mathbf{Q}_\beta^{-1}\right),\ \mathbf{Q}_\beta = \lambda_\beta\left(\mathbf{D} - \rho_\beta \mathbf{\Omega}\right). \tag{36}$$

We note that each mixture component, k, is indexed by both its own mean and variance under this construction, which is more flexible than the use of a common variance under Model 3.

In fitting the above models, for the rest of the parameters, we choose proper priors with hyperparameter values that are easily updated by the data. Namely, the prior distribution for the degrees of freedom parameter ν is $InvGamma(1,1)$; priors for precision parameters λ, τ^{-2} are set to $Gamma(1,1)$; the prior for the concentration parameter α of the Dirichlet distribution is also $Gamma(1,1)$; strength of correlation parameters ρ are sampled from the uniform distribution with appropriate support intervals, $(-1,0)$ or $(0,1)$, as discussed in Section 3.5.

4. Applications to CES data and simulation study

4.1 Computation

We used the Automatic Differentiation Variational Inference (ADVI) algorithm (Kucukelbir et al., 2015) implemented in RStan V2.9.0 package, which is the R interface for the Stan modeling language (Gelman et al., 2015), to implement our models. The variational inference ADVI scheme, implemented in Stan, factors the joint posterior distribution for the set of model parameters and employs a normal distribution for each distribution component after transforming the support for the associated parameter to the real line. The approximate posterior is then estimated using gradient-descent method to minimize the "error lower bound" (ELBO) derived from minimizing the Kullback-Leibler divergence between the approximate density and the true distribution. The Jacobians for the inverse transformations of the parameters are carried along in the objective function.

Full Stan scripts for Models 1-6 are given in Appendix. The same script can be used to run Markov Chain Monte Carlo (MCMC) and ADVI algorithms in Stan. The advantage of the ADVI over an MCMC algorithm is that it is relatively fast and scalable. It must be noted that the ADVI algorithm is susceptible to local optima. We conducted a simulation study (not included) to ensure ADVI returns correct parameter values under our prior formulations. In the study, we used parameter values similar to what is expected in CES. A pilot run on the CES data showed that both algorithms returned the same values for our sets of true parameters.

We set the truncation level for maximum number of clusters, K , to 20 in both clustering models (3 and 6), which is sufficiently high since only about 3 to 5 of the cluster probabilities θ_k's are estimated as non-zero.

4.2 Application to CES

Our data are defined for a set of $N = 2233$ sub-industry-by-State domains; the data series are based on September 2008 as the starting point. We chose this particular year estimation cycle because of the non-trivial employment pattern that occurred during the period of the "great recession", which induced a marked shift in employment trends from previous years.

The survey direct estimates are given by (4). Variances of $\hat{R}_{i,t}$ are assumed to be known, as earlier discussed. In the current application, similar to the classical case of Fay and Herriot (1979), variances were obtained using a generalized variance function approach.

We obtained model estimates $\tilde{R}_{i,t}$ (based on standardized employment counts) for relative monthly changes for each month over the 12-month estimation period. For the cross-sectional Models 1-3, we fit one month of estimates at a time; for Models 4-6, the preceding twelve months of data were used to derive the estimate for a given month $t,\ t = 1,...,T$.

The estimates of employment levels at month t are obtained from the set of $\tilde{R}_{i,\tau}$, $\tau = 1,...,t,$ as

$$\tilde{Y}_{i,t} = Y_{i,0} \prod_{\tau=1}^{t} \tilde{R}_{i,\tau}. \tag{37}$$

Due to different seasonality patterns between the employment series derived from the administrative QCEW data and CES, the most meaningful comparison of the two series is

after 12 months of estimation. Results for each major industry and overall, presented in Table 1, are based on the mean absolute deviation (MAD):

$$MAD = N^{-1}\sum_{i=1}^{N}\left|\tilde{Y}_{i,12} - Y_{i,12}\right|, \tag{38}$$

where $Y_{i,12}$ comes from the (QCEW) census data and is used as "the gold standard" for the estimates.

Table 1: Real data results (levels after 12 months of estimation, MAD)

Industry	N	Direct	FH	FHt	CFHt	mFHt	mFHtc	mCFHtc
1000	50	804	**734**	1077	829	1146	1093	769
2000	141	2152	1990	1825	**1798**	1933	1942	1903
3100	234	1112	**1022**	1220	1140	1269	1118	1035
3200	140	955	877	1156	971	925	878	**856**
4100	124	1485	1270	1232	**1198**	1229	1241	1234
4200	286	**1439**	1836	1468	1443	1484	1471	1488
4300	194	1310	1099	915	947	933	**900**	937
5000	83	1204	1122	810	812	794	792	**739**
5500	149	1473	1121	1081	1099	**1068**	1085	1141
6054	150	1447	1173	1223	1168	1113	**1102**	1127
6055	45	1066	849	807	863	**793**	844	892
6056	115	2344	**1935**	2235	2154	2193	2048	1977
6561	59	1901	1782	1511	1618	**1386**	1429	1428
6562	214	1551	1334	**1230**	1257	1282	1289	1313
7071	59	2047	1362	**1016**	1306	1115	1029	1335
7072	80	1912	1672	1753	**1509**	1659	1554	1638
8000	110	1773	1355	**1085**	1094	1093	1203	1212
Overall	2233	1502	1355	1292	1262	1285	1257	**1256**

Table 1 shows values of MAD for the survey direct estimator and the model-based estimators for each major industry and overall; for each line, the smallest among methods are bold-faced. Judging from the overall results, any of the models considered is an improvement compared to the direct estimator. All models except Model 1 (FH) are designed to curb the effects from outliers occurring in the direct survey estimates. Their overall performances are similar and better than the performance of the FH model. There is a small, but notable, incremental improvement for multivariate models that account for a negatively correlated error structure. We explore this improvement in the simulation study section that follows.

The results also vary by industry. The FHt model is better than FH in most industries, but does worse in some industries, most notably Mining (1000), both Manufacturing Durable (3100) and Nondurable (3200) Goods, and Administrative and Support and Waste Management and Remediation Services (6056). This may happen because, in the year considered, the auxiliary history-based value of x_i was not a good predictor for the current economy. As a result, it is likely that the linear part $\beta_0 + \beta_1 x_i$ of both FH and FHt was misspecified. If this is the case, the variance of random effects would be estimated as large and this, in turn, would have the effect of placing more weight on the direct estimator in the usual weighted average form of the FH estimator. The overall result is that FH is more robust to model misspecification, in the sense that it returns the value

that is close to the direct estimator rather than to the synthetic part $\hat{\beta}_0 + \hat{\beta}_1 x_i$ of the weighted average. In FHt, the assumption is that the direct estimator has the t-distributed errors, that is, a priori, the direct estimator is treated with less trust. The t distribution on sampling errors would cause lower weight on the direct estimator and more weight on the model synthetic estimate. In the case when the model is wrong, this may lead to unfavorable results.

Interestingly, this situation is somewhat corrected in CFHt, the other univariate model that accounts for possible clustering of the domains. Suppose the true model is not linear $\beta_0 + \beta_1 x_i$ but has, say, the parabolic form. The clustering assumption for the distribution of the random effects would compensate for the misspecification of the functional form of the model.

Turning to multivariate models, mFHt and mFHtc also performed worse than FH in the two industries of Mining and Durable Goods Manufacturing, but performed uniformly better across other industries.

Conclusions on the relative performances of our models are limited by fitting to the CES data because of differences in seasonality, the presence of measurement error in the QCEW we use as the gold standard and because our CES data are a single realization. We next employ a simulation study that corrects for these limitations.

4.3 Simulation study

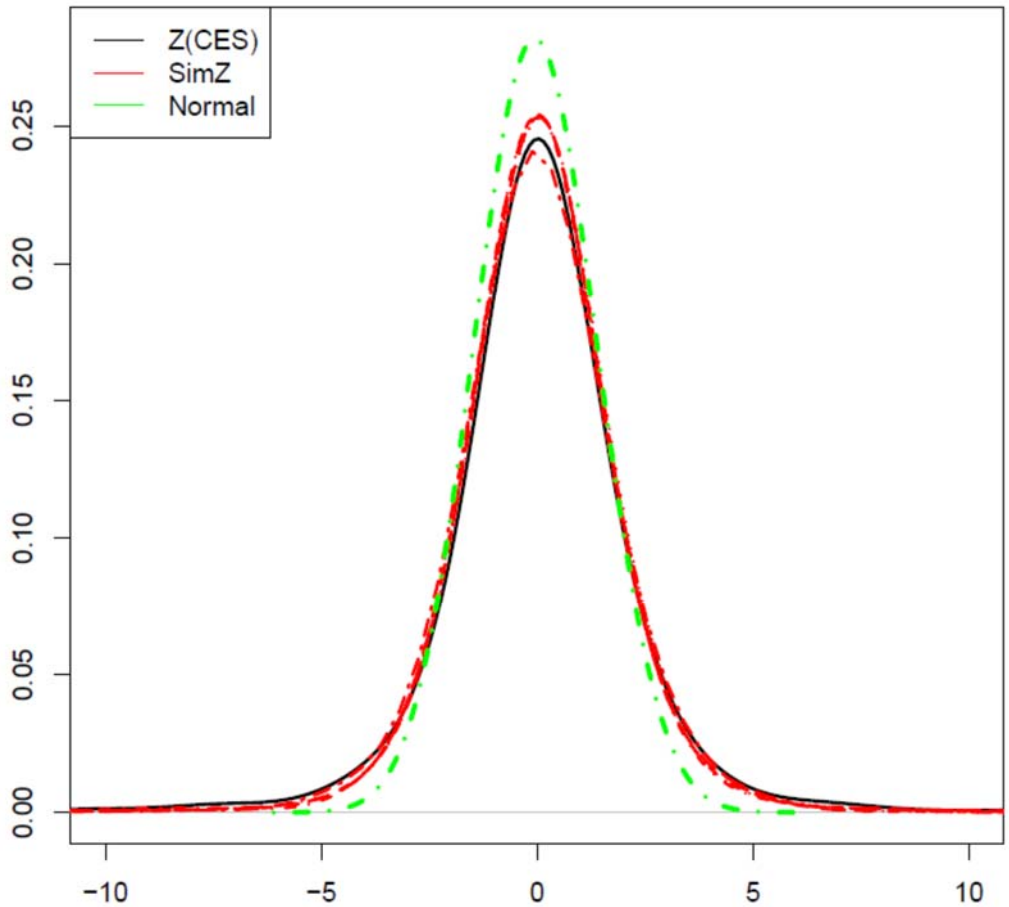
We generated synthetic data based on the "true" census (QCEW) series. We added the Student's t-distributed and correlated noise to QCEW-based ratios $R_{i,t}$ for the 2008 estimation cycle time period of interest. To generate the data, we used the given sample variances and assumed small negative correlation $\rho_e = -0.1$ in the sampling errors between the adjacent months (as also noted in Section 3.5, this choice of value for the autocorrelation is based on previous research.)

Figure 1 plots the density of the standardized CES data (black solid line) and the standardized simulated data (red lines) against the normal density (green line). The CES data is standardized as $Z_{i,t} = \left(\hat{R}_{i,t} - R_{i,t}\right) \Big/ \sqrt{v_{i,t}}$, and the same standardization is used for the simulated data. The plots demonstrate that the CES data density has heavier tails than the tails of the normal distribution and that we have effectively replicated this feature in our simulated data.

A different approach to design the simulation study would be to generate both signal and noise from a model; however, since the true model for signal is unknown, we decided to avoid simulating the signal and use the census data at hand.

There are several advantages in evaluation based on our simulation procedure to generate the synthetic data: (1) simulations allow us to circumvent differences in seasonality between the census and CES series; (2) with the real data, we only *assume* that the census series represent the truth; for the simulated data, they become *in fact* the true values; (3) with the real data, sampling variances are only *assumed* to be known; for simulations, variances *really are* known; (4) evaluation now can be based on multiple simulation runs rather than on a single realization of the real data.

Figure 1. Standardized direct estimates vs normal vs simulated densities

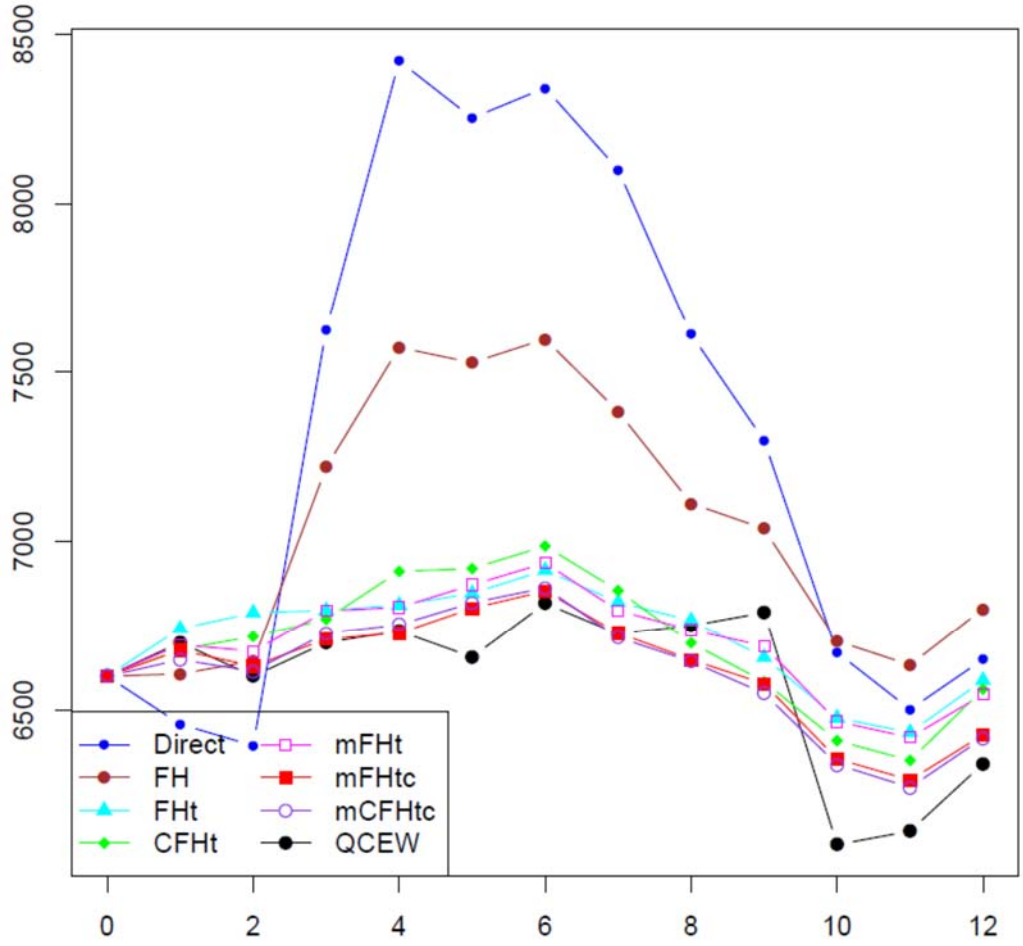


In Figure 2, we plot posterior mean model estimates for one example domain, out of the total of 2233 domains considered in the study. This example (domain #55 in Education, supersector 6561) serves as illustration of the CES (annual) estimation cycle and how the estimators perform over this period.

The months are plotted on the x-axis, the employment levels are on the y-axis. The estimation procedure starts at month t=0, which is September, 2008, in our case. The black line with larger black circles shows the true QCEW-based series. At the time of estimation, the truth is, of course, unknown – QCEW data become available with the lag of 6 to 9 months. The blue line represents the direct sample-based estimator. As we can see, the values of the monthly direct estimator may be extreme, most notably the changes in months 2-to-3 and 3-to-4. The direct estimates are notorious for containing extreme values. However, by the 12-th month of estimation, the direct estimator comes back closer to the QCEW.

Apparently, the extreme changes, as in months 3 and 4, affect the results for the FH estimator (the brown line). This is corrected in the FHt model, as well as in all the other models.

Figure 2. Estimation in Domain #55 of Industry 6561



The largest improvement from our models comes from this correction to extreme values in the direct sample estimates. The multivariate models perform consistently better in this case than the univariate models, though the improvement is much less. The mCFHtc model best captures the dips in employment at months 2 and 10.

Figure 3 presents boxplots that display the distribution (within 95% credible intervals) of the modeled estimates around the true QCEW-based monthly ratios. Simulations as we designed them – using the monthly QCEW values as truth – give us opportunity to compare the results of modeling for each month over the estimation period. Panels in Figure 3 represent months from 1 to 12 (i.e., October through September, with September 2008 as the benchmark month 0). The boxplots are given for four estimators: FH, FHt, CFHt, and mCFHtc. The true value in each month is subtracted from the set of posterior draws for each model. Models other than FH generally perform similarly – and better than FH – in the efficient coverage of the truth. The distribution of the FHt model estimators in most months has narrower credible intervals; however, in about half of the months the 25-75% quantiles of the FHt distribution do not include the true values; the clustering models perform relatively better in covering the true values, although in most months their distribution is slightly wider than the distribution of the FHt estimator.

Figure 3. Posterior distributions (within 95% credible intervals) in Domain #55 of Industry 6561

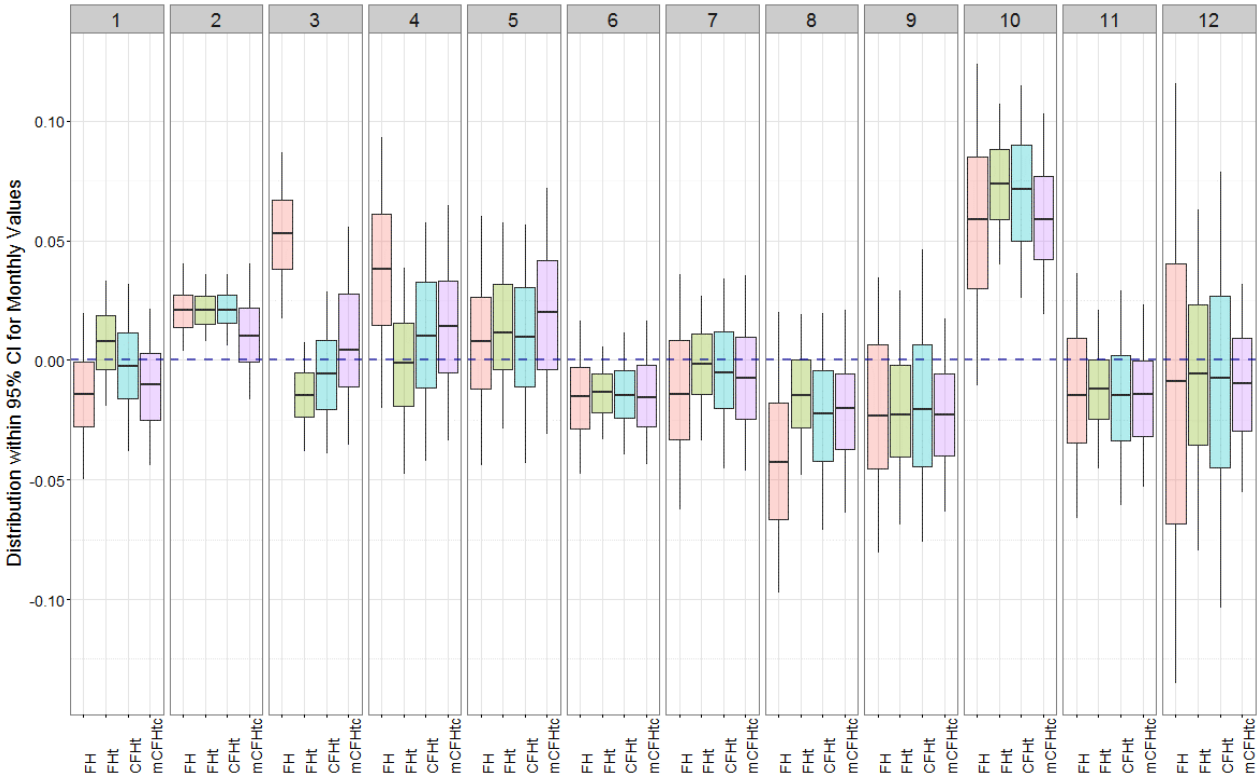


Table 2: Simulation Study (levels after 12 months of estimation, MAD)

Industry	*N*	*Direct*	*FH*	*FHt*	*CFHt*	*mFHt*	*mFHtc*	*mCFHtc*
1000	50	726	650	733	596	741	617	**561**
2000	141	1794	1365	1473	1400	1355	1318	**1282**
3100	234	858	720	994	937	791	713	**662**
3200	140	661	555	752	634	586	556	**521**
4100	124	1143	805	844	778	740	691	**672**
4200	286	3863	1787	1247	1227	**1165**	1168	1212
4300	194	1831	1152	839	833	799	**795**	827
5000	83	721	583	632	**579**	617	620	602
5500	149	1102	752	757	732	687	**675**	680
6054	150	1196	905	985	925	872	821	**820**
6055	45	771	590	547	559	**485**	503	507
6056	115	2266	1853	1852	1766	1940	1802	**1738**
6561	59	3674	2633	1313	1628	**1256**	1330	1446
6562	214	1181	844	854	819	751	**748**	757
7071	59	2738	1805	1235	1393	**1170**	1221	1279
7072	80	1748	1358	1418	1398	1336	**1321**	1345
8000	110	1747	1020	**920**	939	968	1025	1048
Overall	2233	1722	1122	1032	1002	950	**927**	**927**

Numerical evaluation is presented in Table 2. The results are based on average over $S = 5$ sets of simulated data. For employment levels, after 12 months of estimation, the following mean absolute deviation measure for each industry was used:

$$MAD = S^{-1}N^{-1}\sum_{s=1}^{S}\sum_{i=1}^{N}\left|\tilde{Y}_{i,12} - Y_{i,12}\right|.$$

As in the case of the real data, any of the models is an improvement relative to the direct estimator. Similar to the real data results, FHt model is worse than FH in a few industries, notably, in Mining and Manufacturing. As noted in Section 4.2, FHt gives more weight to the auxiliary historical value that was not a good predictor in that year, given the linear model considered. Similarly to the real data case, this unfortunate effect is alleviated by the CFHt model.

The multivariate models give results that are better than the results from the univariate models; accounting for correlation in sampling errors helped in most industries. An interesting result of note is that, similar to the univariate case, clustering in the multivariate model improved estimates in the Goods Producing industries.

5. Summary and Discussion

In this paper, we presented a collection of hierarchical models of similar construction to the Fay-Herriot that estimate a de-noised latent response process. Each model relaxed some of the assumptions of the base Fay-Herriot model formulation to allow the data to better learn the distribution and dependence structure in the survey direct estimates. The computations were implemented by application of the ADVI algorithm with the use of the Stan modeling language, such that our models may be readily employed by both BLS and among practitioners. The computations are relatively fast and easy to implement in the production environment.

Relaxing the assumption of normality of the sampling errors and replacing it with the Student's t distribution lead to significantly improved estimators. The other enhancements displayed small but notable improvements. Accounting for possible clustering of random effects benefited several industries, especially in the Goods Producing sector; accounting for correlation in sampling errors also improved estimates in some industries.

This paper's results may be regarded as a pilot study. The models need to be tested on multiple years of the CES data, as well as for the other CES estimation structure that is currently used for publication at detailed geography. Our simulation study, however, does provide a more general indication that our multivariate and probabilistic clustering models would be expected to out-perform on multiple years of CES data because we generated multiple samples under conditions where we know the true values and have removed seasonality differences.

Appendix

A. Model 1 (FH)

```
data{
 int<lower=1> N; // number of domains
 row_vector[N] y; // set of N observations
 row_vector<lower=0>[N] sigma_y; // known standard deviations
 row_vector[N] x; // set of predictors
} /* end data block */
```

```
parameters{
  row_vector[N] u; /* random effects */
  real mu; /* mean (global intercept) of u_i */
  real<lower=0> tau_mu; /* precision in prior for mu */
  real<lower=0> lambda;
  real<lower=0> lambda_u;
  real beta;
} /* end parameters block */


transformed parameters{
  row_vector[N] fitted_values; /* fitted values */
   { /* block to declare local variables, xb */
    row_vector[N] xb;
    for( i in 1:N )
    {
       xb[i]              <- x[i] * beta;
       fitted_values[i]      <- xb[i] + u[i];
    } /* end loop i over domains */
   } /* end local block to declare xb */
 } /* end transformed parameters block */

model{
  // prior N random effects, u
  { /* local block for parameters */
    lambda_u    ~ gamma( 1.0, 1.0 );
    tau_mu      ~ gamma( 1.0, 1.0 );
    mu   ~ normal(0,inv(sqrt(tau_mu)));
    u    ~ normal(mu,inv(sqrt(lambda_u))); /* N x 1 vectorized */
  } /* end local block for parameters */

  // prior on precision, lambda
  lambda                  ~ gamma( 1.0, 1.0 );
  // prior for fixed effect, beta
  beta                  ~ normal(0,inv(sqrt(lambda)));

  // observed response likelihood
  y        ~ normal( fitted_values, sigma_y);

} /* end model{} block */
```

B. Model 2 (FHt)

```
data{
  int<lower=1> N; // number of domains
  row_vector[N] y; // set of N observations
  row_vector<lower=0>[N] sigma_y; // known standard deviations
  row_vector[N] x; // set of predictors
} /* end data block */

parameters{
  row_vector[N] u; /* random effects */
```

```
  real mu; /* mean (global intercept) of u_i */
  real<lower=0> tau_mu; /* precision in prior for mu */
  real<lower=0> lambda;
  real<lower=0> lambda_u;
  real beta;
  real<lower=0> nu;
  row_vector<lower=0>[N] delta;
} /* end parameters block */

transformed parameters{
  row_vector[N] fitted_values; /* fitted values */
  row_vector<lower=0>[N] sqrt_delta;
   { /* block to declare local variables, xb */
    row_vector[N] xb;
    for( i in 1:N )
    {
       xb[i]              <- x[i] * beta;
       fitted_values[i]      <- xb[i] + u[i];
       sqrt_delta[i]        <- sqrt( delta[i] );
    } /* end loop i over domains */
  } /* end local block to declare xb */
 } /* end transformed parameters block */

model{
  // prior N random effects, u
  { /* local block for parameters */
    lambda_u    ~ gamma( 1.0, 1.0 );
    tau_mu      ~ gamma( 1.0, 1.0 );
    mu   ~ normal(0,inv(sqrt(tau_mu)));
    u    ~ normal(mu,inv(sqrt(lambda_u))); /* N x 1 vectorized */
  } /* end local block for parameters */

  // prior on precision, lambda
  lambda                  ~ gamma( 1.0, 1.0 );
  // prior for fixed effect, beta
  beta                   ~ normal(0,inv(sqrt(lambda)));

  // observed response likelihood
  nu       ~ inv_gamma( 1.0, 1.0 );
  delta     ~ inv_gamma( 0.5*nu, 0.5*nu ); /* N x 1 vectorized */
  y        ~ normal( fitted_values, sigma_y .* sqrt_delta );

} /* end model{} block */
```

C. Model 3 (CFHt)

```
data{
  int<lower=1> N; // number of domains
  row_vector[N] y; // set of N observations
  row_vector<lower=0>[N] sigma_y; // known standard deviations
  row_vector[N] x; // set of predictors
  int<lower=1> K; // *maximum* - truncated number of clusters
} /* end data block */
```

```
transformed data{
  vector<lower=0>[K] ones_K;
  ones_K <- rep_vector(1,K); /* dirichlet prior on alpha has equal shapes */
} /* end transformed data block */

parameters{
  row_vector[N] u; /* random effects */
  row_vector[K] mu; /* cluster centers */
  simplex[K] theta; /* mixture probabilities */
  real alpha; /* DP concentration parameter */
  real<lower=0> lambda_u;
  real<lower=0> tau_mu; /* precision in prior for mu */
  real<lower=0> lambda;
  real beta;
  real<lower=0> nu;
  row_vector<lower=0>[N] delta;
} /* end parameters block */

transformed parameters{
  row_vector[N] fitted_values; /* fitted values */
  row_vector<lower=0>[N] sqrt_delta;
   { /* block to declare local variables, xb */
    row_vector[N] xb;
    for( i in 1:N )
    {
       xb[i]                <- x[i] * beta;
       fitted_values[i]     <- xb[i] + u[i];
       sqrt_delta[i]        <- sqrt( delta[i] );
    } /* end loop i over domains */
  } /* end local block to declare xb */
 } /* end transformed parameters block */

model{
  // priors for mean and covariance cluster locations
  alpha ~ gamma( 1.0, 1.0 ); /* DP concentration parameter */
  theta ~ dirichlet( alpha/K * ones_K ); /* instantiate a truncated DP prior */
 { /* local block for priors for K sets of  cluster centers and variances*/
    tau_mu       ~ gamma( 1.0, 1.0 );
    for( k in 1:K )
    {
      mu[k]          ~ normal(0,inv(sqrt(tau_mu)));
    }
 } /* end local block for cluster parameters */

{
    /* mixture prior for u[i], i = 1 ,.., N */
    lambda_u      ~ gamma( 1.0, 1.0 );
    for( i in 1:N ) /* by row of N x T, u */
     {
      real ps[K];
      for( k in 1:K )
```

```
    {
     /* increment log posterior for clustering model on u */
     ps[k]       <- log(theta[k]) + normal_log(u[i], mu[k], inv(sqrt(lambda_u)));
    }
    increment_log_prob( log_sum_exp(ps) );
  } /* end loop over case observation, i */
 }

 // prior for fixed effect, beta
 // prior on precision, lambda
 lambda                ~ gamma( 1.0, 1.0 );
 beta                  ~ normal(0,inv(sqrt(lambda)));

 // observed response likelihood
 nu      ~ inv_gamma( 1.0, 1.0 );
 delta  ~ inv_gamma( 0.5*nu, 0.5*nu ); /* N x 1 vectorized */
 y       ~ normal( fitted_values, sigma_y .* sqrt_delta );

} /* end model{} block */
```

D. Model 4 (mFHt)

```
functions{
   // sum a matrix, by rows
   vector row_sums(matrix X)
   {
     vector[rows(X)] s ;
     s <- X * rep_vector(1, cols(X)) ;
     return s;
   }
} /* end user functions{} block */

data{
 int<lower=1> N; // number of domains
 int<lower=1> T; // number of times points - months
 row_vector[T] y[N]; // set of N, T x 1 multivariate observations
 row_vector<lower=0>[T] sigma_y[N]; // known area standard deviations
 matrix[N,T] x; // matrix of (time-indexed) predictors
 matrix<lower=0>[T,T] Omega; // T x T CAR adjacency matrix
} /* end data block */

transformed data{
 vector[T] zros_T;
 vector<lower=0>[T] d;
 matrix[T,T] D;
 zros_T <- rep_vector(0,T); /* vector of zeros for sampling Tx1 fixed effects, beta */
 d <- row_sums(Omega); /* diagonal entries of precision matrix */
 D <- diag_matrix(d);
} /* end transformed data block */

parameters{
 row_vector[T] u[N]; /* random effects */
 row_vector[T] mu; /* T x 1 mean of u_i */
```

```
  real<lower=0> tau_mu; /* precision in prior for mu */
  real<lower=0,upper=1> rb; /* strength of correlation parameter */
  real<lower=0,upper=1> r_u; /* strength of correlation parameter */
  /* for Q = lambda *(D - rb*Omega) in beta ~ N_{T}(0,Q^{-1}) */
  real<lower=0> lambda;
  real<lower=0> lambda_u;
  vector[T] beta;
  real<lower=0> nu;
  row_vector<lower=0>[T] delta[N];
} /* end parameters block */

transformed parameters{
  row_vector[T] fitted_values[N]; /* fitted values */
  row_vector<lower=0>[T] sqrt_delta[N];
  { /* block to declare local variables, xb */
    matrix[N,T] xb;
    for( i in 1:N )
    {
      for( j in 1:T )
      {
        xb[i,j]                <- x[i,j] * beta[j];
        fitted_values[i,j]  <- xb[i,j] + u[i,j];
        sqrt_delta[i,j]       <- sqrt( delta[i,j] );
      } /* end loop j over time points */
    } /* end loop i over domains */
  } /* end local block to declare xb */
} /* end transformed parameters block */

model{
  // prior N x T random effects, u
  { /* local block for priors for T x T CAR precision */
    matrix[T,T] Q; /* T x T CAR precision matrix for mu[k] */
    r_u              ~ beta( 1.0, 1.0 );
    lambda_u      ~ gamma( 1.0, 1.0 );
    tau_mu         ~ gamma( 1.0, 1.0 );
    Q                <- lambda_u * (D - r_u * Omega);
    /* prior for T x 1 mu */
    mu          ~ normal(0,inv(sqrt(tau_mu))); /* vectorized */
    /* CAR prior for u[i], i = 1 ,.., n */
    for( i in 1:N )
    {
      u[i]                  ~ multi_normal_prec( mu, Q );
    } /* end loop i over domains */
  } /* end local block for cluster parameters */

  // prior for time-indexed fixed effects, T x 1,beta
  { /* local variable block for Q in prior for T x 1, beta */
    matrix[T,T] Q; /* T x T CAR precision matrix for beta */
    /* T x T precision matrix, Q, for T x 1 vector regression coefficients, beta */
    Q                    <- lambda * (D - rb * Omega);
    /* TxT, Q <- tb * (D-rb*Omega) = CAR precision matrix */
    rb                   ~ uniform( 0, 1 );
```

```
   // prior on P x P covariance matrix, Lambda
   lambda                ~ gamma( 1.0, 1.0 );
   /* Implement a beta ~ N_{T}(0,Q^{-1})  */
   beta                  ~ multi_normal_prec( zros_T, Q );
 } /* end local variable blocks for Q, Sigma_beta and Lbeta */

 // observed response likelihood
 nu                  ~ inv_gamma( 1.0, 1.0 );
 for( i in 1:N ) /* by row of y */
 {
   to_vector(delta[i])      ~ inv_gamma( 0.5*nu, 0.5*nu );
   y[i] ~ normal( fitted_values[i], sigma_y[i] .* sqrt_delta[i] ); // vectorized for T x 1 response
 }
} /* end model{} block */
```

E. Model 5 (mFHtc)

```
functions{
   // sum a matrix, by rows
   vector row_sums(matrix X)
   {
     vector[rows(X)] s ;
     s <- X * rep_vector(1, cols(X)) ;
     return s;
   }
} /* end user functions{} block */

data{
 int<lower=1> N; // number of domains
 int<lower=1> T; // number of times points - months
 row_vector[T] y[N]; // set of N, T x 1 multivariate observations
 row_vector<lower=0>[T] sigma_y[N]; // known area standard deviations
 matrix[N,T] x; // matrix of (time-indexed) predictors
 matrix[T,T] Omega; // T x T CAR adjacency matrix for prior on cluster centers
} /* end data block */

transformed data{
 vector[T] zros_T;
 vector<lower=0>[T] d;
 matrix[T,T] D;
 zros_T <- rep_vector(0,T); /* vector of zeros for sampling Tx1 fixed effects, beta */
 d <- row_sums(Omega); /* diagonal entries of precision matrix */
 D <- diag_matrix(d);
} /* end transformed data block */

parameters{
 row_vector[T] u[N]; /* latent random effects receiving Gaussian mixture prior */
 row_vector[T] mu; /* T x 1 mean of u_i */
 real<lower=0> tau_mu; /* precision in prior for mu */
 real<lower=0,upper=1> rb; /* strength of correlation parameter */
 real<lower=0,upper=1> r_u; /* strength of correlation parameter */
```

```
  /* for Q = lambda *(D - rb*Omega) in beta ~ N_{T}(0,Q^{-1}) */
  real<lower=0> lambda;
  real<lower=0> lambda_u;
  vector[T] beta;
  real<lower=0> nu;
  row_vector<lower=0>[T] delta[N];
  real<lower=-1,upper=0> r_e;
} /* end parameters block */

transformed parameters{
  row_vector[T] fitted_values[N]; /* fitted values */
  row_vector<lower=0>[T] sqrt_delta[N];
  { /* block to declare local variables, xb */
    matrix[N,T] xb;
    for( i in 1:N )
    {
      for( j in 1:T )
      {
        xb[i,j]                <- x[i,j] * beta[j];
        fitted_values[i,j]   <- xb[i,j] + u[i,j];
        sqrt_delta[i,j]       <- sqrt( delta[i,j] );
      } /* end loop j over time points */
    } /* end loop i over domains */
  } /* end local block to declare xb */
} /* end transformed parameters block */

model{
  // prior N x T random effects, u
  { /* local block for priors for T x T CAR precision */
    matrix[T,T] Q; /* T x T CAR precision matrix for mu[k] */
    r_u              ~ beta( 1, 1 );
    lambda_u      ~ gamma( 1.0, 1.0 );
    tau_mu         ~ gamma( 1.0, 1.0 );
    Q                <- lambda_u * (D - r_u * Omega);
    /* prior for T x 1 mu */
    mu          ~ normal(0,inv(sqrt(tau_mu))); /* vectorized */
    /* CAR prior for u[i], i = 1 ,.., n */
    for( i in 1:N )
    {
      u[i]                  ~ multi_normal_prec( mu, Q );
    } /* end loop i over domains */
  } /* end local block for cluster parameters */

  // prior for time-indexed fixed effects, T x 1,beta
  { /* local variable block for Q in prior for T x 1, beta */
    matrix[T,T] Q; /* T x T CAR precision matrix for beta */
    /* T x T precision matrix, Q, for T x 1 vector regression coefficients, beta */
    Q                    <- lambda * (D - rb * Omega);
    /* TxT, Q <- tb * (D-rb*Omega) = CAR precision matrix */
    rb                   ~ uniform( 0, 1 );

    // prior on P x P covariance matrix, Lambda
```

```
    lambda                  ~ gamma( 1.0, 1.0 );
    /* Implement a beta ~ N_{T}(0,Q^{-1})  */
    beta                    ~ multi_normal_prec( zros_T, Q );
  } /* end local variable blocks for Q, Sigma_beta and Lbeta */

  // observed response likelihood
  nu                  ~ inv_gamma( 1.0, 1.0 );
  { /* local block to build T x T error covariance matrix */
    matrix[T,T] Q;
    matrix[T,T] L;
    Q          <- (D - r_e* Omega);
    L          <- cholesky_decompose(inverse_spd(Q));
    for( i in 1:N ) /* by row of y */
    {
      matrix[T,T] L_i;
      to_vector(delta[i])     ~ inv_gamma( 0.5*nu, 0.5*nu );
      L_i   <- diag_post_multiply(L ,(sigma_y[i] .*  sqrt_delta[i]));
      y[i]  ~ multi_normal_cholesky( fitted_values[i], L_i); /* T x 1 */
    } /* end loop i over areas */
  } /* end local block */
} /* end model{} block */
```

F. Model 6 (mCFHtc)

```
functions{
    // sum a matrix, by rows
    vector row_sums(matrix X)
    {
      vector[rows(X)] s ;
      s <- X * rep_vector(1, cols(X)) ;
      return s;
    }

data{
  int<lower=1> N; // number of domains
  int<lower=1> K; // *maximum* - truncated number of clusters
  int<lower=1> T; // number of times points - months
  row_vector[T] y[N]; // Set of N, T x 1 multivariate observations
  row_vector<lower=0>[T] sigma_y[N]; // known area standard deviations
  matrix[N,T] x; // matrix of (time-indexed) predictors
  matrix<lower=0>[T,T] Omega; // T x T CAR adjacency matrix for prior on cluster centers
} /* end data block */

transformed data{
  vector<lower=0>[K] ones_K;
  vector[T] zros_T;
  vector<lower=0>[T] d;
  matrix[T,T] D;
  ones_K <- rep_vector(1,K); /* dirichlet prior on alpha has equal shapes */
  zros_T <- rep_vector(0,T); /* vector of zeros for sampling Tx1 fixed effects, beta */
  d <- row_sums(Omega); /* diagonal entries of precision matrix */
  D <- diag_matrix(d);
} /* end transformed data block */
```

```
parameters{
  row_vector[T] u[N]; /* latent random effects receiving Gaussian mixture prior */
  simplex[K] theta; /* mixture probabilities after marginalizing over inclusions*/
  row_vector[T] mu[K]; /* cluster centers */
  real<lower=0> tau[K]; /* cluster precision parameters for T x T Rw(1) CAR precision matrix
*/
  real<lower=0,upper=1> rho[K]; /* cluster strength of correlation parameters */
  real alpha; /* DP concentration parameter */
  real<lower=0,upper=1> rb; /* strength of correlation parameter */
  real<lower=0,upper=1> rmu; /* strength of correlation parameter */
  /* for Q = lambda *(D - rb*Omega) in beta ~ N_{T}(0,Q^{-1}) */
  real<lower=0> lambda;
  real<lower=0> lambda_mu;
  vector[T] beta;
  real<lower=0> nu;
  row_vector<lower=0>[T] delta[N];
  real<lower=0> eta; /* squared exponential formula precision parameter */
  real<lower=-1,upper=0> r_e;
} /* end parameters block */

transformed parameters{
  row_vector[T] fitted_values[N]; /* fitted values */
  row_vector<lower=0>[T] sqrt_delta[N];
  { /* block to declare local variables, xb */
    matrix[N,T] xb;
    for( i in 1:N )
    {
      for( j in 1:T )
      {
        xb[i,j]                <- x[i,j] * beta[j];
        fitted_values[i,j]  <- xb[i,j] + u[i,j];
        sqrt_delta[i,j]      <- sqrt( delta[i,j] );
      } /* end loop j over time points */
    } /* end loop i over domains */
  } /* end local block to declare xb */
} /* end transformed parameters block */

model{
  // priors for mean and covariance cluster locations
  alpha ~ gamma( 1.0, 1.0 ); /* DP concentration parameter */
  theta ~ dirichlet( alpha/K * ones_K ); // instantiate a truncated DP prior
  { /* local variable block for mu[k] and invSigma[k] */
    matrix[T,T] Q; /* T x T CAR precision matrix for mu[k] */
    /* prior for T x 1, mu[k] */
    /* T x T precision matrix, Q, for T x 1 vector regression coefficients, beta */
    rmu              ~ uniform( 0, 1 );
    lambda_mu     ~ gamma( 1.0, 1.0 );
    Q                <- lambda_mu * (D - rmu * Omega);
    for( k in 1:K )
    {
      mu[k]        ~ multi_normal_prec( zros_T, Q ); /* T x 1 cluster centers */
```

```
  } /* end loop k over clusters */
} /* end construction of local variables, invSigma */

// prior for latent response, u_{i} ~ prod_{k=1}^{K}theta_k *N(u_{i},mu_k,Sigma_k)
{ /* local block for invSigma[k], k = 1,...,K */
  matrix[T,T] invSigma[K]; /* invSigma[k] <- tau[k] * (D-rho[k]*Omega) */
  rho ~ uniform( 0, 1 ); /* K x 1 vector of cluster-indexed strength of correlation */
  /* priors for invSigma[k] */
  tau ~ gamma( 1.0, 1.0 ); /* K x 1 vector of cluster-indexed precision parameters */
  for( k in 1:K  )
  {
    /* Construct set of K (cluster-indexed) T x T CAR precision matrices, invSigma */
    /* for mixture distribution for N x T, u */
    invSigma[k]          <- tau[k] * (D - rho[k] * Omega);
  } /* end loop over k */

  for( i in 1:N ) /* by row of N x T, u */
  {
    real ps[K];
    for( k in 1:K )
    {
     /* increment log posterior for clustering model on u */
      ps[k]      <- log(theta[k]) + multi_normal_prec_log(u[i], mu[k], invSigma[k]);
    }
    increment_log_prob( log_sum_exp(ps) );
  } /* end loop i over case observations */
} /* end local block in mixture prior for u */

// prior for time-indexed fixed effects, T x 1,beta
// and TxT precision, Q (constructed as a CAR, RW(1))
/* sample parameters that compose PxP, Lambda and TxT Q */
{ /* local variable block for Q */
  matrix[T,T] Q; /* T x T CAR precision matrix for beta */
  /* T x T precision matrix, Q, for T x 1 vector regression coefficients, beta */
  Q                    <- lambda * (D - rb * Omega);
  /* TxT, Q <- tb * (D-rb*Omega) = CAR precision matrix */
  rb                   ~ uniform( 0, 1 );

  // prior on P x P covariance matrix, Lambda
  lambda                ~ gamma( 1.0, 1.0 );
  /* Implement a beta ~ N_{T}(0,Q^{-1})  */
  beta                  ~ multi_normal_prec( zros_T, Q );
} /* end local variable blocks for Q, Sigma_beta and Lbeta */

// observed response likelihood
nu                  ~ inv_gamma( 1.0, 1.0 );
{ /* local block to build T x T error covariance matrix */
  matrix[T,T] Q;
  matrix[T,T] L;
  Q          <- (D - r_e* Omega);
  L          <- cholesky_decompose(inverse_spd(Q));
  for( i in 1:N ) /* by row of y */
```

```
  {
    matrix[T,T] L_i;
    to_vector(delta[i])    ~ inv_gamma( 0.5*nu, 0.5*nu );
    L_i                <- diag_post_multiply(L ,(sigma_y[i] .*  sqrt_delta[i]));
    y[i]               ~ multi_normal_cholesky( fitted_values[i], L_i); /* T x 1 */
  } /* end loop i over areas */
 } /* end local block */
} /* end model{} block */
```

References

Banerjee, S., Wall, M. and Carlin, B.P. (2003). Frailty modelling for spatially correlated survival data with application to infant mortality in Minnesota. Biostatistics 4123-142.

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

Gelman, A., Lee, D., and Guo, J. (2015) **Stan: A probabilistic programming language for Bayesian inference and optimization**. In press, *Journal of Educational and Behavior Science*.

Gershunskaya, J. (2015), "Combining Time Series and Cross-sectional Data for Current Employment Statistics Estimates" *2015 Proceedings of the American Statistical Association*, Survey Research Methods Section, Alexandria, VA: American Statistical Association.

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.

Huang, E. T. and Bell, W. R. (2006), "Using the t-Distribution in Small Area Estimation: An Application to SAIPE State Poverty Models," *2006 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association

Jin, X., Carlin, B. P. and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. Biometrics 61, 950-961.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., and Blei, D. M. (2016), Automatic differentiation variational inference. arXiv preprint arXiv:1603.00788.

Neal, R. M. (2000) ``Markov chain sampling methods for Dirichlet process mixture models", *Journal of Computational and Graphical Statistics*, vol. 9, pp. 249-265

Pfeffermann, D. (2002). Small area estimation - new developments and directions. *Int. Statist. Rev.* 70 125-143.

Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*.28. 40–68.

Pfeffermann, D. and Burck, L. (1990). Robust Small Area Estimation Combining Time Series and Cross- Sectional Data. Survey Methodology, 16, 217-237.

Rao, J.N.K. and Molina, I. (2015), *Small Area Estimation, Second Edition.* John Wiley & Sons, Hoboken.

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

Scott, S. and Sverchkov, M. (2005), Variance Measures for X-11 Seasonal Adjustment: A Summing Up of Empirical Work. *ASA Proceedings of the Joint Statistical Meetings.*

Scott, S., Pfeffermann, D., and Sverchkov, M. (2012). Estimating Variance in X-11 Seasonal Adjustment. In *Economic Time Series: Modeling and Seasonality*, edited by William R. Bell, Scott H. Holan, and Tucker S. McElroy, 185–210. London: Chapman and Hall.

Stan Development Team. (2015). Stan modeling language users guide and reference manual, version 2.6.0 [Computer software manual]. Retrieved from **http://mc-stan.org/**