RESEARCH REPORT SERIES
(Statistics #2016-01)

Estimation of the Difference of Small Area Parameters from Different Time Periods

Ryan Janicki

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: February 25, 2016

Estimation of the difference of small area means from different time periods

Ryan Janicki*

Abstract

The U. S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program provides annual estimates of poverty within school districts, counties and states for different age groups. The SAIPE program's estimates are model-based, and use single year American Community Survey (ACS) data and administrative records. There is potential for improving estimates of the current year's parameters by using previous years' data in the small area model. Two methods for utilizing multiple years of data are compared: the first method uses a bivariate normal distribution for the model errors from different time periods, and the second method assumes an AR1 structure on model parameters. Gains in efficiency using multiple years of survey data for estimation of a small area parameter are investigated using each method. In addition, estimates of the increase or decrease over time of a small area parameter are constructed, as well as credible intervals for the change over time. An example using state-level SAIPE data is presented.

Key Words: small area, time series, EBLUP, interval estimation

1. Introduction

Many surveys collect data at regular intervals, such as monthly or annually. These surveys are designed to provide accurate estimates for quantities of interest on large domains, such as at a national or state level. However, it is often of interest to simultaneously estimate quantities at smaller domains, such as at smaller geographic regions, cross classified by demographic characteristics such as age, race, and sex, for which the original survey was not designed. While the overall sample size for the survey may be very large, there may be insufficient sample size to provide reliable direct estimates for a parameter of interest for certain smaller domains. The term *small area* is used for any domain of interest for which the domain-specific sample size is insufficient to obtain reliable direct estimates. A popular method for obtaining estimates with higher precision than survey-based direct estimates is to introduce small area models which "borrow strength" by connecting different areas and incorporating auxiliary covariate information, such as administrative records. Empirical Bayes and hierarchical Bayes methods are widely used for inference in small area estimation problems, as these methods can explicitly link different small areas through use of random effects and effectively incorporate administrative records.

A popular small area model is the Fay-Herriot model (Fay and Herriot, 1979), given by

$$\begin{aligned} y_i &= \theta_i + e_i \\ \theta_i &= \boldsymbol{x}_i^T \boldsymbol{\beta} + \nu_i \end{aligned} \tag{1}$$

for $i = 1, \ldots, m$. The y_i are the direct estimates of the small area means θ_i, and the sampling errors e_i are independent, mean zero random variables, with known sampling

*Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, Ryan.Janicki@census.gov. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U. S. Census Bureau.

variances V_i. The model errors ν_i are i.i.d. mean zero random variables, independent of the e_i, with common variance σ^2. This model is a cross-sectional model using data from a single time point.

For known σ^2, the best linear unbiased predictor (BLUP) of θ_i is

$$\tilde{\theta}_i = (1 - w_i)y_i + w_i \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}, \tag{2}$$

where $w_i = V_i/(V_i + \sigma^2)$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\left(\sigma^2\right) = diag\left(V_1 + \sigma^2, \ldots, V_m + \sigma^2\right)$, and

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma^2) = \left(\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X} \boldsymbol{\Sigma}^{-1} \boldsymbol{y}. \tag{3}$$

The BLUP $\tilde{\theta}_i$ is thus a weighted average of the regression estimate $\boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}$ and the direct survey estimate y_i. In practice, σ^2 is not known, and must be estimated. The empirical best linear unbiased predictor (EBLUP), $\hat{\theta}_i$ of θ_i, replaces σ^2 with an estimate, $\hat{\sigma}^2$, in equations (2) and (3). The variance component, σ^2, can be estimated with a moment-based estimator (Prasad and Rao, 1990) which does not require an assumption of normality of the error components but is not guaranteed to be positive. The variance component can also be estimated with the maximum likelihood (ML) or restricted maximum likelihood (REML) estimator (Harville, 1977), which both require full specification of a parametric model and must be found numerically.

An important application of the Fay-Herriot model is estimation of income and poverty by the Small Area Income and Poverty Estimates (SAIPE) program. The SAIPE program produces estimates within school districts, counties, and states, for the age groups 0 – 4, 5 – 17, 18 – 64, and 65 and older. The data used are the one year estimates, y_i, from the American Community Survey (ACS) of the proportion in poverty in small area i, and the covariates, $\boldsymbol{x}_i$, include the tax return poverty rate, the tax nonfiler rate, the Supplemental Nutrition Assistance Program (SNAP) participation rate, Supplemental Security Income (SSI) recipiency rate, and the residuals from a regression of the Census 2000 poverty ratios on the previous four covariates.

Many statistical agencies fit the Fay-Herriot model each year, using only the cross-sectional survey data collected during the current time period. This sequence of modeling efforts result in a time series of estimates. From these estimates, patterns, such as the estimated increase or decrease of a small area mean, can be observed, and it is of interest to determine whether these changes are significantly different from zero. There are two challenges to making this type of inference using cross-sectional models. First, because distinct models are fit at distinct time points, estimates for a small area parameter use only data from that time period. There is no explicit linking mechanism in these cross-sectional models which uses data over time or describes how parameters change over time. Second, in small area models, the parameter representing the truth, such as the true poverty rate in a state, is modeled using combination of fixed and random effects. Any interval used for inferring significance needs to account for the randomness of both the estimate and the parameter of interest.

Because estimates are produced at fixed time periods based on cross sectional survey data, it is desirable to make inferences on the difference in parameters from different time periods based on the published estimates. Basel et al. (2010) gives a method for estimating the covariance of model errors in the Fay-Herriot model at distinct time points s and t, using only the estimated parameters $\hat{\boldsymbol{\beta}}_s$ and $\hat{\boldsymbol{\beta}}_t$ and $\hat{\sigma}_s$ and $\hat{\sigma}_t$; that is, the model parameters do not have to be reestimated once new data is collected. This method is based on an implicit assumption of bivariate normality of the model parameters $\nu_{i,s}$ and $\nu_{i,t}$ in model (1), that is, that

$$\begin{pmatrix} \nu_{i,s} \\ \nu_{i,t} \end{pmatrix} \overset{i.i.d.}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \sigma_{s,t} \\ \sigma_{s,t} & \sigma_t^2 \end{pmatrix} \right) \tag{4}$$

for small areas $i = 1, \ldots, m$ and time points s and t.

Under the assumption of bivariate normality of model errors in (4), the marginal distributions at each time point reduce to the Fay-Herriot model (1). The BLUP, $\tilde{\theta}_{i,t}$, of $\theta_{i,t}$ at each time point t, based only on the cross-sectional data, is therefore the same as in (2). The mean square error (MSE) of the difference $\tilde{\theta}_{i,t} - \tilde{\theta}_{i,s}$ under model (4), for known variance components σ_t^2, σ_s^2 and $\sigma_{s,t}$, is

$$
\begin{aligned}
MSE\left(\tilde{\theta}_{i,s} - \tilde{\theta}_{i,t}\right) &= E\left\{\tilde{\theta}_{i,s} - \tilde{\theta}_{i,t} - \theta_{i,s} + \theta_{i,t}\right\}^2 \\
&= E\left\{\tilde{\theta}_{i,s} - \theta_{i,s}\right\}^2 + E\left\{\tilde{\theta}_{i,t} - \theta_{i,t}\right\}^2 - 2E\left\{\left(\tilde{\theta}_{i,s} - \theta_{i,s}\right)\left(\tilde{\theta}_{i,t} - \theta_{i,t}\right)\right\} \\
&= \frac{\sigma_s^2 V_{i,s}}{\sigma_1^2 + V_{i,s}} + \frac{V_{i,s}^2}{\left(\sigma_s^2 + V_{i,s}\right)^2} \boldsymbol{x}_{i,s}^T \left(\boldsymbol{X}_s^T \boldsymbol{\Sigma}_s^{-1} \boldsymbol{X}_s\right)^{-1} \boldsymbol{x}_{i,s} \\
&\qquad + \frac{\sigma_t^2 V_{i,t}}{\sigma_t^2 + V_{i,t}} + \frac{V_{i,t}^2}{\left(\sigma_s^2 + V_{i,t}\right)^2} \boldsymbol{x}_{i,t}^T \left(\boldsymbol{X}_t^T \boldsymbol{\Sigma}_t^{-1} \boldsymbol{X}_t\right)^{-1} \boldsymbol{x}_{i,t} \\
&\qquad - \frac{2\sigma_{s,t} V_{i,s} V_{i,t}}{\left(\sigma_s^2 + V_{i,s}\right)\left(\sigma_t^2 + V_{i,t}\right)} \boldsymbol{l}_i^T \left(\boldsymbol{I} - \boldsymbol{M}_{i,s}\right)\left(\boldsymbol{I} - \boldsymbol{M}_{i,t}\right)^T \boldsymbol{l}_i,
\end{aligned}
\tag{5}
$$

where $\boldsymbol{M}_{i,t} = \boldsymbol{X}_t \left(\boldsymbol{X}_t^T \boldsymbol{\Sigma}_t^{-1} \boldsymbol{X}_t\right)^{-1} \boldsymbol{X}_t^T \boldsymbol{\Sigma}_t^{-1}$ and $\boldsymbol{l}_i$ is a vector of 0s of length m, with a 1 in the ith place.

An estimate of the MSE of the difference of EBLUPs, $\hat{\theta}_{i,t} - \hat{\theta}_{i,s}$, can be obtained by plugging in estimates of the variance components in (5). Moment-based estimates of σ_s^2 and σ_t^2 are well-known (Rao, 2003); alternatively, the maximum likelihood or restricted maximum likelihood estimator can be used. Basel et al. (2010) showed that

$$
\tilde{\sigma}_{s,t} = \frac{1}{r} \sum_{i=1}^{m} \frac{\left(y_{i,s} - \boldsymbol{x}_{i,s}^T \hat{\boldsymbol{\beta}}_s\right)\left(y_{i,t} - \boldsymbol{x}_{i,t}^T \hat{\boldsymbol{\beta}}_t\right)}{\left(\sigma_s^2 + D_i\right)\left(\sigma_t^2 + D_i\right)}, \tag{6}
$$

where

$$
r = tr\left(\boldsymbol{\Sigma}_s^{-1}\left(\boldsymbol{I} - \boldsymbol{M}_s\right)\left(\boldsymbol{I} - \boldsymbol{M}_t\right)^T \boldsymbol{\Sigma}_t^{-1}\right),
$$

is an unbiased estimate of $\sigma_{s,t}$. Estimates $\hat{\sigma}_t^2$ and $\hat{\sigma}_s^2$ can be plugged into the expression for $\tilde{\sigma}_{s,t}$ to obtain an estimator $\hat{\sigma}_{s,t}$ for $\sigma_{s,t}$. This estimator does have the limitation that it can lead to estimated correlations which lie outside the interval $(-1, 1)$.

This method for linking data from different time points is very useful because inference on the change in small area means over time can be made using only point estimates computed separately from each cross-sectional data set, that is, point estimates do not need to be updated with the inclusion of additional data. The disadvantage of this method is that it does not easily generalize to data sets including more than two time periods. Also, because point estimates use only data from a single time frame, there is possible efficiency loss in not using the full data set to compute point estimates.

An alternative to using a bivariate normal distribution for model errors to incorporate data from multiple periods is to introduce an explicit linking mechanism over time within the small area model, in addition to the linking mechanism across small areas. The time-linking within the model involves indexing the small area means θ by both space i and time t, and modeling the relationship of the parameters θ_{i,t_1} and θ_{i,t_2} for different time periods t_1 and t_2. Recently, there has been much written about combining cross-sectional and time series data in small area modeling. Fully Bayesian models were used by Datta et al. (1999), Datta et al. (2002), and Li et al. (2012), with a random walk process linking parameters over time. Franco and Bell (2015) adapted binomial/logit normal models for use with discrete time series. Frequentist methods were used by Saei and Chambers (2003), Torabi and Shokoohi (2012), and Pereira and Coelho (2012), who used an autoregressive process for the time-linking mechanism. Extensions of these models to include a spatio-temporal component were given by Singh et al. (2005), Marhuenda et al. (2013), and Porter et al. (2014). Applications of these methods to estimating poverty can be found in Hawala and Lahiri (2012), Taciak and Basel (2012), and Franco and Bell (2015).

These papers are closely related to the work of Rao and Yu (1994), who gave an extension of the Fay-Herriot model which allows modeling of survey data taken at multiple time points. The extension introduces a first order autoregressive process into the linking model. For $i = 1, \ldots, m$ small areas and $t = 1, \ldots, T$ time points, the model is

$$\begin{aligned} y_{i,t} &= \theta_{i,t} + e_{i,t} \\ \theta_{i,t} &= \boldsymbol{x}_{i,t}^T \boldsymbol{\beta}_t + v_i + u_{i,t} \\ u_{i,t} &= \rho u_{i,t-1} + \varepsilon_{i,t}, \;\; |\rho| < 1 \end{aligned} \tag{7}$$

$$\begin{aligned} \boldsymbol{e}_i = (e_{i,1}, \ldots, e_{i,T})^T &\overset{ind.}{\sim} N_m(\boldsymbol{0}, \boldsymbol{V}_i) \\ v_i &\overset{i.i.d.}{\sim} N\left(0, \sigma_v^2\right) \\ \varepsilon_{i,t} &\overset{i.i.d.}{\sim} N\left(0, \sigma^2\right) \end{aligned}$$

for known $(m \times m)$ sampling covariance matrix $\boldsymbol{V}_i$ for the ith small area. The restriction $|\rho| < 1$ guarantees stationarity of the autoregressive component of the model.

In matrix form, this model can be written as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\nu} + \boldsymbol{u} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Z} = \boldsymbol{I}_m \otimes \boldsymbol{1}_T$, $\otimes$ is the direct product operator, and $\boldsymbol{1}_T$ is a vector of 1's of length T. It follows that

$$Cov\,(\boldsymbol{y}) \equiv \boldsymbol{\Sigma} = \boldsymbol{V} + \sigma^2 \boldsymbol{I}_m \otimes \boldsymbol{\Gamma} + \sigma_\nu^2 \boldsymbol{Z}\boldsymbol{Z}^T, \tag{8}$$

where $\boldsymbol{V}$ is the matrix of sampling variances of the direct estimates $\boldsymbol{y}$ and $\boldsymbol{\Gamma}$ is a $T \times T$ matrix with elements $\rho^{|i-j|}/(1-\rho^2)$.

The advantage of using a small area model which has a time series components is that data sets including any number of time points can be used, and there is potential for increased precision of estimates, provided the model adequately fits the data. However, assessing model fit can be a difficult problem when using small area models. Another issue is that parameter estimates can be difficult to compute, and may take values outside the parameter space. A further, practical problem is that introduction of new data requires a complete refitting of the model, which in turn provides new estimates of previous years' small area parameters, in addition to the new, current year prediction of the small area means.

Extensions of the Fay Herriot model, such as model (4) or (7), which include a linking mechanism over time, are useful not only for point estimation, but also for constructing intervals for the change in small area means over time. A point estimate of the difference of small area parameters can be obtained from model (7) using standard EBLUP theory, and the estimated MSE, along with a normal approximation, can be used to construct intervals $(L_{i,T,\alpha}, U_{i,T,\alpha}) = (L_{i,T,\alpha}(\boldsymbol{Y}, \boldsymbol{X}), U_{i,T,\alpha}(\boldsymbol{Y}, \boldsymbol{X}))$, with the property that

$$P\left(L_{i,T,\alpha} \le \theta_{i,T} - \theta_{i,T-1} \le U_{i,T,\alpha}\right) \ge 1 - \alpha.$$

"Significant change" of small area parameters from different time periods at level α can then be defined to mean the interval $(L_{i,T,\alpha}, U_{i,T,\alpha})$ does not contain 0. It should be noted that here, "significant" does not correspond to a standard hypothesis test, since in small area models such as (1), a linking model using area-specific covariates is employed to relate the true small area means. A standard hypothesis test of $\theta_{i,T-1} = \theta_{i,T}$ is therefore not meaningful, since $P(\theta_{i,T-1} = \theta_{i,T}) = 0$.

There are two main goals of this paper. The first is to explore the benefit of using multiple years of survey data to predict a small area mean in the current time period using model (7), compared to estimation using only cross sectional data and the Fay-Herriot model (1). The second goal is to estimate the increase or decrease of small area parameters over time, and to construct valid intervals for this difference, using model (4) or (7). In Section 2, estimation of model parameters in small area models which combine cross-sectional and time-series data are discussed, and new moment-based estimators for estimating the variance components are introduced. In Section 3, SAIPE data from 2007 – 2012 are analyzed, and the EBLUPs using only cross-sectional data are compared to EBLUPs using the full data set. Credible intervals centered on these EBLUPs are constructed based on estimated MSE using models (4) and (7), and compared to naive intervals based only on direct estimates. A numerical example is presented in Section 4 to investigate finite sample properties of the estimators and of the different intervals. Concluding remarks are given in 5.

2. Parameter estimation

In model (7), the BLUP of $\theta_{i,t}$ depends on the variance parameters ρ, σ^2 and σ_v^2, which would be unknown in practice. Rao and Yu (1994) gave unbiased moment estimators of σ^2 and σ_v^2 when ρ is known, and different plug-in estimators of ρ were explored; the validity of these moment estimators depends only on correct specification of the model mean and covariance. The issue with these moment estimators is that they depend on estimation of ρ, which is particularly difficult in model (7), and the estimated values are very sensitive to ρ. In addition, moment estimates of ρ were shown in Rao and Yu (1994) to often fall outside the admissible range of $(-1, 1)$, and to underestimate the true value of ρ. In this section, new moment estimators of the variance components are presented as alternatives to those given in Rao and Yu (1994), which are biased, but do not depend on any other estimated parameters.

Let $a_{i,t} = y_{i,t} - \boldsymbol{x}_{i,t}^T\boldsymbol{\beta}$ and let $\hat{a}_{i,t} = y_{i,t} - \boldsymbol{x}_{i,t}^T\hat{\boldsymbol{\beta}}$ be the residuals based on the least squares estimator $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ of $\boldsymbol{\beta}$. For simplicity of presentation, assume the sampling errors $e_{i,t}$ are uncorrelated both over time t and across small areas i. Note that

$$E\left(\left(a_{i,t+2} - a_{i,t}\right)\left(a_{i,t+1} - a_{i,t}\right)\right) = \sigma^2 + V_{i,t},$$

independent of the parameters ρ and σ_v^2. If $T > 2$, the residuals $\hat{a}_{i,t}$ can be used to obtain an estimator of σ^2:

$$\tilde{\sigma}^2 = \frac{1}{m(T-2)}\sum_{i=1}^{m}\sum_{t=1}^{T-2}\left\{\left(\hat{a}_{i,t+2} - \hat{a}_{i,t}\right)\left(\hat{a}_{i,t+1} - \hat{a}_{i,t}\right) - V_{i,t}\right\} \tag{9}$$

and $\hat{\sigma}^2 = \max\left(\widetilde{\sigma}^2, 0\right)$. This estimator is biased due to the estimation of the regression parameter $\boldsymbol{\beta}$, but does not depend on an initial estimate of the parameter ρ.

An estimator for ρ can be obtained in a similar way. Note that

$$E\left\{(a_{i,t+2} - a_{i,t+1})(a_{i,t} - a_{i,t+1})\right\} = \frac{\sigma^2}{1-\rho^2}(1-\rho)^2 + V_{i,t+1}$$

and

$$E\left\{\left(a_{i,t+1} - a_{i,t}\right)^2\right\} = \frac{2\sigma^2}{1-\rho^2}\left(1-\rho\right) + V_{i,t+1} + V_{i,t}.$$

Combining these two expressions and replacing the $a_{i,t}$ with their estimators $\hat{a}_{i,t}$ gives

$$\hat{\rho} = \frac{\sum_{i=1}^{m} \sum_{t=1}^{T-2} \{(\hat{a}_{i,t+1} - 2\hat{a}_{i,t+2} + \hat{a}_{i,t})(\hat{a}_{i,t} - \hat{a}_{i,t+1}) + V_{i,t+1} - V_{i,t}\}}{\sum_{i=1}^{m} \sum_{t=1}^{T-2} \{(\hat{a}_{i,t+1} - \hat{a}_{i,t})^2 - V_{i,t+1} - V_{i,t}\}} \tag{10}$$

as an estimator of ρ. This estimator is biased, and not guaranteed to be in the range $(-1, 1)$, but will be consistent under general conditions.

It is more difficult to isolate the parameter σ_v^2, but an approximation can be made using the following identities:

$$E\left(\sum_{t=1}^{T} a_{i,t}\right)^2 = T^2\sigma_v^2 + \frac{\sigma^2}{1-\rho^2}\frac{T(1-\rho^2) - 2\rho(1-\rho^T)}{(1-\rho)^2} + \sum_{t=1}^{T} V_{i,t}$$

and

$$E\left(\sum_{t=1}^{T} a_{i,t}^2\right) = T\sigma_v^2 + \frac{T\sigma^2}{1-\rho^2} + \sum_{t=1}^{T} V_{i,t}.$$

Combining these two equations gives

$$\frac{1}{mT(T-1)}\sum_{i=1}^{m} E\left\{\left(\sum_{t=1}^{T} a_{i,t}\right)^2 - \sum_{t=1}^{T} a_{i,t}^2\right\} = \sigma_v^2 + \frac{2\rho\sigma^2(\rho^T - T\rho - 1)}{T(T-1)(1-\rho^2)(1-\rho)^2} = \sigma_v^2 + O(T^{-1}).$$

We then have

$$\hat{\sigma}_v^2 = \frac{1}{mT(T-1)}\sum_{t=1}^{m}\left\{\left(\sum_{t=1}^{T} \hat{a}_{i,t}\right)^2 - \sum_{t=1}^{T} \hat{a}_{i,t}^2\right\} \tag{11}$$

as an estimator of σ_v^2. The estimator $\hat{\sigma}_v^2$ has the attractive property that it is nonnegative, and, as with the other moment estimators derived in this section, does not depend on estimates of other variance components.

Alternatively, the unknown parameters can be estimated numerically by maximizing the likelihood or the restricted likelihood function when the full distribution is specified. The value of using a moment estimator over REML or ML estimators is that the moment estimators do not depend on the assumption of normally distributed sampling or model errors; they instead depend only on the correctness of the first and second moment specification in the models. In addition, numerical routines for optimizing likelihood or restricted likelihood functions can be sensitive to the starting value, and reliable moment estimators are useful as initial values.

3. Data example

This section analyzes the state-level ACS poverty data for the age group 5 – 17 related children for the years 2007 – 2012. In this example, 2012 is the target year for obtaining predictions for the small area (state) poverty rates. The direct survey estimates from the ACS are compared to model-based estimates obtained using the Fay-Herriot model (1) and the Rao-Yu model (7). The Fay-Herriot model is fit using cross-sectional data from each year from 2007 – 2012, and the Rao-Yu model is fit on the full six year data set, as well as on 4 years of data (2009 – 2012) and 5 years of data (2008 – 2012).

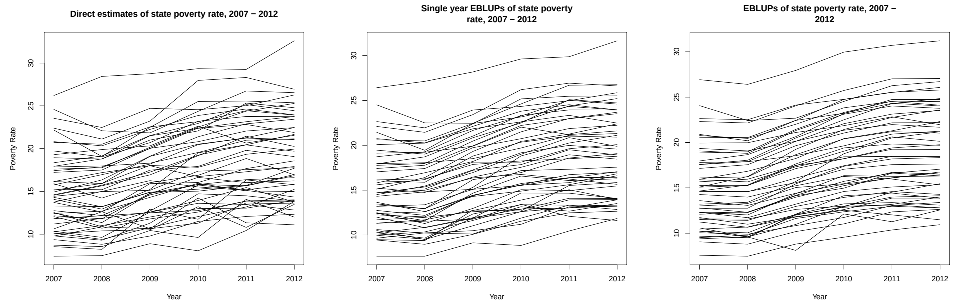

Figure 1: State poverty estimates over time

Figure 1 shows the time series of estimates of poverty rates for each of the 50 states from 2007 – 2012. From left to right, the plots are of the direct survey estimates, the predicted values using cross-sectional data from each year and the Fay-Herriot model, and the predicted values using all six years of data and the Rao-Yu model. The plots of the direct estimates indicate a clear overall trend of increase in poverty over the years 2007 – 2012; this is most noticeable on the subintervals 2008 – 2009 and 2010 – 2011. For the years 2009 – 2010 and 2011 – 2012, the poverty estimates are noisier and do not display the same strong trend as over other time periods. The second graph in Figure 1 shows the EBLUPs of poverty rates for each state using only cross-sectional data. At each year, the EBLUPs shrink the direct estimates to the regression estimate of poverty rate (equation (2)), but there is no smoothing of the direct estimates over time. From the plot of these model-based estimates, a clearer time trend of the estimated poverty rates can be observed. The final plot is the plot of the EBLUPs of poverty rates using the Rao-Yu model, which "borrows stregth" both across the small areas as well as over time. Compared to the first and second set of plots, the third set of graphs exhibits the most smoothing.

There are two primary interests in this analysis. The first is to estimate the poverty rates of children ages 5 – 17 in each state in 2012. The second is to estimate the change in poverty rates from 2011 to 2012 and to construct intervals for this change. Visual inspection of Figure 1 indicates that each of the three methods can produce estimates which vary substantially. For example, the top line in each graph in Figure 1, which corresponds to poverty estimates in Mississippi, have similar estimates of poverty in 2012. However, the estimated change in poverty from 2011 to 2012 is quite different using each of the three methods, with the AR1 model (7) predicting little change between 2011 and 2012, and the cross-sectional methods predicting steeper increases.

3.1 Parameter estimation

To fit the Rao-Yu model (7), the unknown variance components σ^2, σ_v^2, and ρ need to be estimated. Table 1 compares moment estimates of the variance components using the methods of Section 2 with the moment estimates derived in Rao and Yu (1994) and with REML estimates, when the model is fit over four, five, or six years.

Table 1a compares the Rao and Yu (1994) moment estimate of ρ in column 1 with the moment estimate in equation (10) and the REML estimate. The estimates in Table 1a show the difficulty in estimating the autoregressive parameter ρ, as each estimation method produces very different results. The two moment estimators are negative, which is counterintuitive, and does not agree with the plots in Figure 1; the REML estimates are more reasonable. In addition, the estimates of ρ are sensitive to the number of time periods used, with the REML estimates varying from 0.92 using 4 years of data, to 0.51 using 5

Table 1: Parameter estimation of the variance components ρ, σ^2, and σ_v^2

Years	$\hat{\rho}_{ry}$	$\hat{\rho}$	$\hat{\rho}_{REML}$
2007 – 2012	-0.215	-0.135	0.656
2008 – 2012	-0.574	-0.238	0.511
2009 – 2012	-0.494	-0.199	0.920

(a) Estimation of ρ. The moment estimator $\hat{\rho}_{ry}$ is from Rao and Yu (1994), and $\hat{\rho}$ is the estimator given in (10). The REML estimator $\hat{\rho}_{REML}$ is found numerically.

Years	$\hat{\sigma}^2(0)$	$\hat{\sigma}^2(0.25)$	$\hat{\sigma}^2(0.5)$	$\hat{\sigma}^2(0.75)$	$\hat{\sigma}^2(\hat{\rho}_{ry})$	$\hat{\sigma}^2$	$\hat{\sigma}^2_{REML}$
2007 – 2012	0.312	0.291	0.283	0.288	0.344	0.141	0.172
2008 – 2012	0.365	0.356	0.352	0.356	0.416	0.191	0.192
2009 – 2012	0.316	0.348	0.388	0.438	0.270	0.308	0.118

(b) Estimation of σ^2. The moment estimators $\hat{\sigma}^2(\cdot)$ are from Rao and Yu (1994), and depend on ρ, and $\hat{\sigma}^2$ is the estimator given in (9). The REML estimator $\hat{\sigma}^2$ is found numerically.

Years	$\hat{\sigma}_v^2(0)$	$\hat{\sigma}_v^2(0.25)$	$\hat{\sigma}_v^2(0.5)$	$\hat{\sigma}_v^2(0.75)$	$\hat{\sigma}_v^2\,(\hat{\rho}_{ry})$	$\hat{\sigma}_v^2$	$\hat{\sigma}^2_{v,REML}$
2007 – 2012	0.341	0.315	0.249	0.001	0.352	0.473	0.477
2008 – 2012	0.534	0.508	0.442	0.175	0.554	0.449	0.506
2009 – 2012	0.530	0.459	0.314	0.000	0.600	0.490	0.000

(c) Estimation of σ_v^2. The moment estimators $\hat{\sigma}_v^2(\cdot)$ are from Rao and Yu (1994), and depend on ρ, and $\hat{\sigma}_v^2$ is the estimator given in (11). The REML estimator $\hat{\sigma}_v^2$ is found numerically.

years of data. The unreasonableness and instability of moment estimates of ρ over time show the usefulness of having estimators for the other parameters that do not depend on ρ.

Table 1b compares estimates of σ^2 using each of the three methods. The moment estimate $\hat{\sigma}^2(\cdot)$ depends on the unknown parameter ρ, and was shown in Rao and Yu (1994) to be unbiased when the true value of ρ is used. The first four columns are the estimates of σ^2 as ρ ranges from 0 to 0.75 and the fifth column uses the moment estimate $\hat{\rho}_{ry}$ of ρ. This moment estimate is stable, both as a function of ρ as well as over different time periods. The moment estimator $\hat{\sigma}^2$ in Table 1b is calculated using equation (9) and does not depend on ρ. It is interesting to note the close agreement of $\hat{\sigma}^2$ with the REML estimate $\hat{\sigma}^2_{REM}$ when five or six years of data are used

Table 1c compares estimates of σ_v^2 using each of the three methods. The estimate $\hat{\sigma}_v^2(\cdot)$, which depends on ρ, varies considerably, both over different values of ρ and over different time periods. It is interesting to note that as ρ approaches 0.75, $\hat{\sigma}_v^2(\cdot)$ approaches 0. The fitted model with $\rho = 0.75$ then has a deterministic AR1 component, and all variation is over the small areas, and not over time. The estimate $\hat{\sigma}_v^2$ from equation (11) is stable over different time periods, and is close to the REML estimate $\hat{\sigma}^2_{v,REML}$ when five or six years of data are used. When four years of data are used, the REML estimate is exactly equal to 0. It seems reasonable to conclude in this example, that four years of data is too few time points to obtain reliable estimates.

Table 2 shows the estimates in Figure 1 for the first five states in more detail. The first columns in Table 2 show the direct ACS survey estimates, with the estimated sampling error in parentheses. The next two columns show the cross sectional EBLUPs and estimated root mean square error in parentheses. Note that the cross-sectional EBLUPs are based on the Fay-Herriot model, and only use data at a single time point. The final two columns are the predicted values using all six years of data and the Rao-Yu model, with the estimated root mean square error in parentheses; the EBLUPs of the poverty rates using all six years data in each of the 50 states are constructed using REML estimates for the variance components.

For larger states, like California, Table 2 shows that all three methods produce very similar results, in terms of both point estimates and estimated MSE, due to the relatively small sampling variances. For the other states in Table 2, there is noticeable reduction in the estimated MSE of the EBLUPs using cross-sectional data, and a further reduction when the full data set is used.

3.2 Credible intervals

In this section, credible intervals for the difference of small area means from 2011 to 2012 are compared. One simple method for constructing a valid interval for $\theta_{i,T} - \theta_{i,T-1}$ is to use only the direct estimates and the estimated sampling variances:

$$y_{i,T} - y_{i,T-1} \pm z_{\alpha/2}\sqrt{V_{i,T} + V_{i,T-1}}.$$

When the samples from different years are taken independently, and the direct estimates are normally distributed, this coverage probability is exactly $1 - \alpha$. However, if the sampling variances V_i are large, as is typically the case in small area estimation problems, this interval will be too large to make useful inferences.

An improvement over using only the direct estimates is to use the cross-sectional EBLUPs along with an estimate of the model covariance between years as in (4). Estimates of the model covariance $\sigma_{s,t}$, using equation (6), between different years from 2007 to 2012 are shown in Table 3. The credible interval based on the cross-sectional EBLUPs is

$$\hat{\theta}_{i,T} - \hat{\theta}_{i,T-1} \pm z_{\alpha/2}\sqrt{M\hat{S}E\left(\hat{\theta}_{i,T} - \hat{\theta}_{i,T-1}\right)},$$

where $M\hat{S}E\left(\hat{\theta}_{i,T} - \hat{\theta}_{i,T-1}\right)$ is given in equation (5).

An interval which is constructed using all 6 years of data is

$$\hat{\theta}_{i,diff} \pm z_{\alpha/2}\sqrt{M\hat{S}E\left(\hat{\theta}_{i,diff}\right)},$$

where $\hat{\theta}_{i,diff}$ is the EBLUP of $\theta_{i,T} - \theta_{i,T-1}$ under model (7), and $M\hat{S}E\left(\hat{\theta}_{i,diff}\right)$ is an estimate of the mean square error of $\hat{\theta}_{i,diff}$, given by

$$\begin{aligned} M\hat{S}E\left(\hat{\theta}_{i,diff}\right) = & \frac{2\hat{\sigma}^2(1-\hat{\rho})}{1-\hat{\rho}^2} - \left(\frac{\hat{\sigma}^2(1-\hat{\rho})}{1-\hat{\rho}^2}\right)^2 \hat{\boldsymbol{\gamma}}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\gamma}} \\ & + \left(\boldsymbol{x}_{i,T} - \boldsymbol{x}_{i,T-1} + \boldsymbol{W}\right)\left(\boldsymbol{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X}\right)^{-1} \left(\boldsymbol{x}_{i,T} - \boldsymbol{x}_{i,T-1} + \boldsymbol{W}\right)^T, \end{aligned}$$

where $\hat{\boldsymbol{\gamma}}^T = \left(-\rho^{T-2}, -\rho^{T-3}, \ldots, -\rho, -1, 1\right)$, $\boldsymbol{\Sigma}_i$ is the ith block of the block diagonal covarince matrix $\boldsymbol{\Sigma}$ in (7), and

$$\boldsymbol{W} = \frac{\hat{\sigma}^2\left(1-\hat{\rho}\right)}{1-\hat{\rho}^2} \boldsymbol{\gamma}^T \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_i.$$

Table 4 compares interval estimates using each of the three methods, using $\alpha = 0.05$, for the first five states. The first column shows intervals based only on the direct estimates, the second column shows intervals which use cross-sectional EBLUPs and the estimated

Table 2: Comparison of estimates of the proportion in poverty in select states over 2007 – 2012. The estimated root mean square error for each estimate is in parentheses.

State	Year	Direct		Bivariate		AR1	
AL	2012	25.358	(0.534)	25.513	(0.461)	25.789	(0.400)
AL	2011	25.559	(0.736)	25.468	(0.582)	25.539	(0.420)
AL	2010	25.509	(0.672)	25.175	(0.570)	24.753	(0.409)
AL	2009	22.484	(0.605)	22.401	(0.509)	22.440	(0.395)
AL	2008	19.087	(0.691)	19.443	(0.565)	20.190	(0.428)
AL	2007	22.171	(0.726)	21.447	(0.571)	20.892	(0.455)
AK	2012	11.989	(0.912)	13.252	(0.643)	12.568	(0.541)
AK	2011	14.067	(1.523)	13.334	(0.808)	11.264	(0.560)
AK	2010	9.613	(1.037)	11.882	(0.743)	12.124	(0.521)
AK	2009	10.428	(1.130)	10.110	(0.758)	8.104	(0.580)
AK	2008	8.548	(0.968)	10.214	(0.683)	9.588	(0.509)
AK	2007	8.714	(0.877)	10.604	(0.634)	10.229	(0.519)
AZ	2012	24.778	(0.583)	24.720	(0.497)	24.741	(0.430)
AZ	2011	25.319	(0.798)	25.089	(0.628)	24.745	(0.467)
AZ	2010	22.070	(0.735)	22.486	(0.616)	23.189	(0.445)
AZ	2009	20.998	(0.582)	20.840	(0.516)	20.644	(0.434)
AZ	2008	18.799	(0.571)	18.712	(0.507)	18.661	(0.416)
AZ	2007	17.930	(0.584)	17.925	(0.512)	17.968	(0.442)
AR	2012	26.272	(0.792)	25.863	(0.608)	26.076	(0.502)
AR	2011	25.016	(0.838)	24.970	(0.639)	25.525	(0.482)
AR	2010	24.547	(0.830)	24.331	(0.664)	24.755	(0.471)
AR	2009	24.709	(0.859)	23.931	(0.643)	24.130	(0.477)
AR	2008	22.469	(0.834)	21.943	(0.645)	22.511	(0.487)
AR	2007	23.544	(0.752)	22.646	(0.594)	22.634	(0.492)
CA	2012	22.458	(0.239)	22.394	(0.233)	22.315	(0.224)
CA	2011	21.273	(0.247)	21.250	(0.241)	21.289	(0.227)
CA	2010	20.462	(0.268)	20.435	(0.262)	20.428	(0.242)
CA	2009	18.190	(0.228)	18.194	(0.223)	18.235	(0.210)
CA	2008	16.954	(0.247)	16.936	(0.240)	16.831	(0.221)
CA	2007	15.760	(0.232)	15.810	(0.226)	15.949	(0.216)

Table 3: Moment estimator of $\sigma_{s,t}$ in equation (6)

year	2007	2008	2009	2010	2011	2012
2007						
2008	0.714					
2009	0.586	0.630				
2010	0.793	0.673	0.622			
2011	0.372	0.336	0.645	0.618		
2012	0.542	0.579	0.640	0.676	0.725	

parameter $\hat{\sigma}_{s,t}$, and the third column gives intervals which use the time series model with AR1 structure. Overall, the average length of the confidence intervals using only the direct estimates and the estimated survey variances is 4.066. The average length of the 50 credible intervals based on the bivariate model (4) is 2.480, compared to an average length of 1.838 using model (7). However, there is a wide range in the lengths among the different states under each method. It can be seen in Table 4 that for large states such as California, the intervals using each of the three methods are very similar. Also, for California, the increase in poverty from 2011 to 2012 can be defined to be a significant increase under any of the three methods, since the intervals do not contain 0. For small states, such as Alaska, there can be dramatic differences in the lengths of the intervals. Note that the decrease in poverty rates from 2011 to 2012 in Alaska would be defined as insignificant using only the direct estimates or the Fay-Herriot model, but there would be an estimated significant *increase*, using the Rao-Yu model.

The estimates for poverty rates for Alaska in Table 2, along with the associated credible intervals in Table 4 highlight one of the practical issues that need to be considered when combining cross-sectional and time-series data. The direct estimates of the proportion in poverty in Alaska decrease from 14.1% to 12% from 2011 to 2012. Compare this to the cross-sectional EBLUPs which decrease only slightly from 13.4% to 13.3% over the same time period. However, the EBLUPs of poverty rates using model (7) actually *increase* from 2011 to 2012 from 11.3% to 12.6%, and from Table 4, this would be defined to be a significant increase. While these results seem counterintuitive at first, there is no logical contradiction, as the direct estimates for Alaska have relatively high estimated sampling variances, and therefore will "borrow strength" substantially from related areas, using both cross-sectional and time-series information.

Table 4: Comparison of credible intervals for the difference of small area means from 2011 to 2012 for select states.

State	Direct	Bivariate	AR1
AL	(-1.984, 1.580)	(-1.204, 1.292)	(-0.622, 1.122)
AK	(-5.557, 1.401)	(-1.506, 1.343)	(0.335, 2.274)
AZ	(-2.478, 1.396)	(-1.713, 0.976)	(-0.977, 0.971)
AR	(-1.004, 3.517)	(-0.497, 2.282)	(-0.436, 1.540)
CA	(0.512, 1.858)	(0.502, 1.785)	(0.423, 1.629)

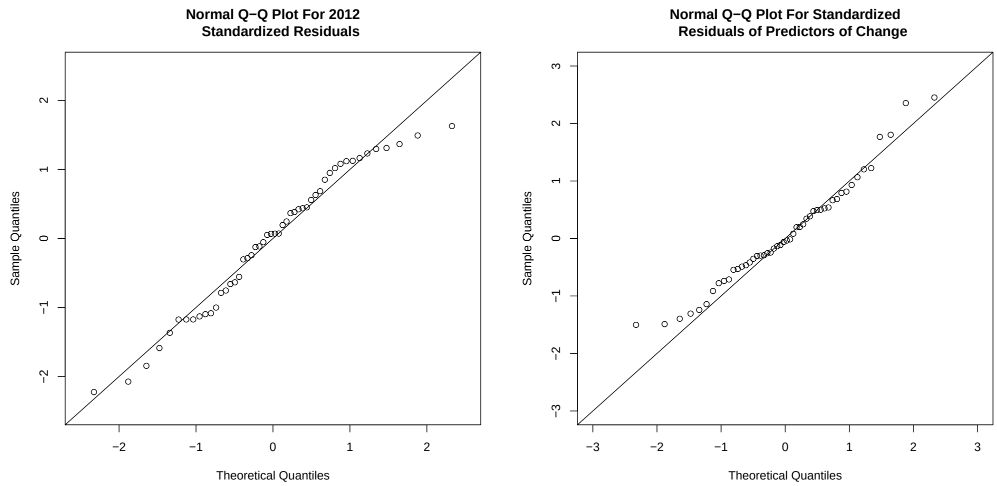

Figure 2: Q-Q plots of 2012 standardized residuals using predicted propoprtions of related children 5 – 17 in poverty and standardized residuals using predicted changes in poverty rates from 2011 – 2012.

3.3 Model diagnostics

It is clear from the analysis in subsections 3.1 and 3.2 that there is improvement, in terms of reduction in estimated MSEs of the predicted small area means and length of interval estimates, in using the Rao-Yu model (7) and the full data set over only cross-sectional data, provided that model (7) fits the SAIPE data set reasonably well. Assessing model fit is known to be a difficult problem in small area estimation. For the Fay-Herriot model, common diagnostic procedures include testing significance of regression coefficients for selection of fixed effects, and analysis of residuals for checking normality. Note that, conditional on the parameters, the response variables are independent under the Fay-Herriot model. In contrast, when an autoregressive structure is introduced into the small area model, the data are not independent, due to the correlation over time. However, certain subsets of the data will be independent, conditional on the parameters. For example, data across small areas within a given time period are independent (cross-sectional subsets). Choosing independent subsets allows for residual analysis within each smaller set of data.

Figure 2 shows normal QQ plots for residuals using the 5 to 17 related children data set over the time periods 2007 – 2012. The first plot is a normal QQ plot of the 50 state-level standardized residuals in the 2012 time period. The second plot is a normal QQ plot of the standardized residuals based on the EBLUP of the *differences* of small area parameters in each of the 50 states between the year 2011 and the year 2012. Both plots of standardized residuals are close to the diagonal line, giving no indication of departures from normality of the model error terms.

Residual analysis using different subsets of the data allows for a graphical detection of departures from normality of the distribution of the error terms in the small area model. However, it does not give an indication as to whether the linking mechanism over time within the small area model is appropriate. The main difference between the standard Fay-Herriot model and the extension given by Rao and Yu (1994) is the inclusion of an autoregressive process, parameterized by ρ. When $\rho = 0$, there is independence of the response variables over different time periods, and the marginal distribution at year T is similar to the Fay-Herriot model. A test of the null hypothesis $H_0 : \rho = 0$ is therefore of interest.

Table 5: Likelihood ratio test statistic and p-value

T	Years	Λ	p-value
4	2009 – 2012	3.801	0.051
5	2008 – 2012	3.535	0.060
6	2007 – 2012	9.003	0.003

Denote the set of variance parameters by $\boldsymbol{\theta}^T = \left(\rho, \sigma_\nu^2, \sigma^2\right)$. The restricted maximum likelihood (REML) estimator is the estimator $\hat{\boldsymbol{\theta}}$ which maximizes the function (Harville, 1977)

$$\begin{aligned} & l_{REML}\left(\boldsymbol{\theta}; \boldsymbol{y}\right) \\ & \quad = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log\left|\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right| - \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\right)^T \boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\right), \end{aligned} \tag{12}$$

where $\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$ is as in equation (8). Denote by $\hat{\boldsymbol{\theta}}_0$ the maximizer of l_{REML} in (12) when $\rho = 0$. A likelihood ratio test (LRT) statistic based on the REML estimator is given by

$$\Lambda = -2\left(l_{REML}\left(\hat{\boldsymbol{\theta}}_0; \boldsymbol{y}\right) - l_{REML}\left(\hat{\boldsymbol{\theta}}; \boldsymbol{y}\right)\right),$$

which is asymptotically χ_1^2 under the null hypothesis.

Table 5 gives the LRT statistic and the associated p-values based on a χ_1^2 approximation using models fit with 4, 5, or 6 years of data. The p-values for LRT based on 4 and 5 years of data are 0.051 and 0.06, respectively, indicating significance at the 0.1 level. However, when the data from 2007 are included in the analysis, the p-value decreases to 0.003, indicating strong significance to the inclusion of the autoregressive process in the linking model.

4. Simulation study

In this section, two numerical examples are presented to investigate the properties of the moment estimators of variance components discussed in Section 2, and to understand finite sample performance of the EBLUP and associated credible intervals when multiple years of data are used. For the first example, to have similarity with the data analysis done in Section 3, $T = 6$ time periods and $m = 50$ small areas are used. Data are simulated from the Rao-Yu model (7), with parameter values set to $\boldsymbol{\beta} = (-2.5, 0.75, 0.2, 0.2, 0.5)^T$, $\sigma^2 = 1$, $\sigma_v^2 = 1$, and $\rho = 0.5$. Real administrative records data are used for the covariance matrix $\boldsymbol{X}$ and the sampling variances $V_{i,t}$ are the actual estimated sampling variances used by SAIPE. The second example uses the same model formulation as the first, except that $\rho = 0$, so that there is independence of the model errors $u_{i,t}$ in (7) over time. For each of the two model specifications, 1000 simulated data sets were generated. All computational work in this paper was done using R.

4.1 Estimates of model parameters

Table 6 shows the performance of different estimates of the variance components σ^2, σ_v^2, and ρ using three methods: the moment estimators introduced in Subsection 2, the moment estimators in Rao and Yu (1994), and the REML estimates. Table 6 summarizes the performance of each estimation method in terms of bias, standard error, and mean square error (MSE).

Table 6: Performance of estimates of variance components

		$\hat{\rho}_{ry}$	$\hat{\rho}$	$\hat{\rho}_{REML}$
	mean	0.034	0.023	0.002
$\rho = 0$	s.e.	0.227	0.220	0.127
	MSE	0.053	0.049	0.016
	mean	0.564	0.583	0.504
$\rho = 0.5$	s.e.	0.325	0.287	0.131
	MSE	0.110	0.089	0.017

(a) Estimation of ρ. The estimator $\hat{\rho}_{ry}$ is the moment estimator of ρ in Rao and Yu (1994) and $\hat{\rho}$ is the moment estimator in (10). The estimator $\hat{\rho}_{REML}$ is the REML estimator, and is found numerically.

		$\hat{\sigma}^2(0)$	$\hat{\sigma}^2(0.25)$	$\hat{\sigma}^2(0.5)$	$\hat{\sigma}^2(0.75)$	$\hat{\sigma}^2(\hat{\rho}_{ry})$	$\hat{\sigma}^2$	$\hat{\sigma}^2_{REML}$
	mean	1.000	1.125	1.328	1.614	1.012	0.849	0.992
$\rho = 0$	s.e.	0.162	0.188	0.237	0.303	0.185	0.211	0.144
	MSE	0.026	0.051	0.164	0.469	0.034	0.067	0.021
	mean	0.974	0.956	1.006	1.129	1.055	0.854	0.997
$\rho = 0.5$	s.e.	0.168	0.171	0.196	0.240	0.204	0.200	0.142
	MSE	0.029	0.031	0.039	0.074	0.045	0.061	0.020

(b) Estimation of σ^2. The moment estimators $\hat{\sigma}^2(\cdot)$ are from Rao and Yu (1994) and depend on ρ, and $\hat{\sigma}^2$ is the moment estimator in (9). The REML estimator $\hat{\rho}_{REML}$ is found numerically using the correctly specified model.

		$\hat{\sigma}_v^2(0)$	$\hat{\sigma}_v^2(0.25)$	$\hat{\sigma}_v^2(0.5)$	$\hat{\sigma}_v^2(0.75)$	$\hat{\sigma}_v^2(\hat{\rho}_{ry})$	$\hat{\sigma}_v^2$	$\hat{\sigma}^2_{v,REML}$
	mean	0.992	0.862	0.522	0.010	0.974	0.878	0.985
$\rho = 0$	s.e.	0.325	0.330	0.344	0.064	0.328	0.236	0.264
	MSE	0.105	0.128	0.347	0.983	0.108	0.071	0.070
	mean	1.334	1.229	0.974	0.203	1.152	1.188	0.898
$\rho = 0.5$	s.e.	0.414	0.413	0.422	0.326	0.444	0.305	0.424
	MSE	0.283	0.223	0.179	0.741	0.220	0.128	0.190

(c) Estimation of σ_v^2. The moment estimators $\hat{\sigma}_v^2(\cdot)$ are from Rao and Yu (1994) and depend on ρ. The REML estimator $\hat{\sigma}^2_{v,REML}$ is found numerically using the correctly specified model.

Table 6a gives a comparison of estimators of ρ. The estimator $\hat{\rho}_{ry}$ is the moment estimator of ρ from Rao and Yu (1994), $\hat{\rho}$ is the moment estimator in equation (10), and $\hat{\rho}_{REML}$ is the REML estimate of ρ. When the true value of ρ is 0, all three estimators had similar performance in terms of bias, and all three estimators had low MSE, with the REML estimator being the lowest. When the true value of ρ is 0.5, the two moment estimators overestimated ρ, with $\hat{\rho}$ having the lower MSE.

One potential issue with using moment estimators for the autoregressive parameter ρ is that they are not guaranteed to be in the interval $(-1, 1)$. With a the true value of $\rho = 0.5$, the estimator $\hat{\rho}_{ry}$ was greater than 1 in 16.5% of the simulations, while $\hat{\rho}$ was greater than 1 in 14.7% of the simulations. In these cases, the estimator was truncated to 1. When $\rho = 0$, neither estimator was outside the range $(-1, 1)$.

Table 6b gives a comparison of estimators of σ^2. The Rao-Yu moment estimator $\hat{\sigma}^2\,(\rho)$ was shown to be unbiased when the true value of ρ is used – this can be seen in Table 6b

Table 7: Comparison of the average mean square error of the direct estimator, cross-sectional EBLUP, and 6 year EBLUP over all 50 small areas (states).

		Direct	Bivariate	AR1
$\rho = 0$	MSE	0.615	0.440	0.386
$\rho = 0.5$	MSE	0.648	0.473	0.398

when $\rho = 0$ or $\rho = 0.5$. The moment estimator $\hat{\sigma}^2$ from equation (9) underestimated the true value of σ^2 in both examples, for $\rho = 0$ or $\rho = 0.5$; this moment estimator does not perform nearly as well as the Rao-Yu moment estimator when a plug-in for ρ close to the true value is used. However, for values of ρ further from the true value, there was a large increase in bias and MSE.

Table 6c gives a comparison of estimators of σ_v^2. The REML estimate $\hat{\sigma}_v^2$ was very accurate when $\rho = 0$, with small bias and MSE. However, performance was worse when $\rho = 0.5$, with the REML estimate having large increase in both bias and MSE. In these examples the moment estimator $\hat{\sigma}_v^2(\cdot)$ varied considerably over ρ, particularly for the case of $\rho = 0.5$. $\hat{\sigma}_v^2(\hat{\rho}_{ry})$ was nearly unbiased when $\rho = 0$, but overestimated the true value of $\sigma_v^2 = 1$ when $\rho = 0.5$, with a large increase in MSE. The estimate $\hat{\sigma}_v^2$ in equation (11) was biased in both examples, overestimating the true value of $\sigma_v^2 = 1$ when $\rho = 0$, and underestimating σ_v^2 when $\rho = 0.5$. However, the MSE of $\hat{\sigma}_v^2$ was the smallest of all estimators, even compared to the REML estimate $\hat{\sigma}^2_{v,REML}$.

4.2 Current year estimates

In this section, accuracy of predictions for the small area mean in the current time period are investigated. Table 7 shows the average mean square error of the direct estimate, the cross-sectional EBLUP, and the EBLUP using all 6 years of data and the (correctly specified) Rao-Yu model (7). Clearly, from Table 7, there is a large reduction in the average MSE using a small area model, over using only the direct estimates. There is also a noticeable reduction in the average MSE using the Rao-Yu model, compared to only cross-sectional data and the Fay-Herriot model, particularly for the case of $\rho = 0.5$.

Table 8: Comparison of the mean square error of the direct estimates, cross-sectional EBLUPs, and 6 year EBLUPs, in selected individual small areas (states). Average mean square error over 1000 simulations is given, along with the standard error (in parentheses).

	State	AK	DE	LA	MS	NC	RI	TX	WY
$\rho = 0$	Direct	0.794	1.847	0.600	1.190	0.211	1.752	0.066	2.176
		(1.082)	(2.437)	(0.854)	(1.549)	(0.301)	(2.306)	(0.091)	(3.068)
	Bivariate	0.605	0.986	0.506	0.821	0.191	1.035	0.066	1.244
		(0.858)	(1.338)	(0.756)	(1.128)	(0.276)	(1.396)	(0.089)	(1.805)
	AR1	0.520	0.809	0.461	0.699	0.182	0.850	0.065	1.021
		(0.725)	(1.079)	(0.667)	(0.940)	(0.263)	(1.126)	(0.086)	(1.472)
$\rho = 0.5$	Direct	0.893	2.073	0.605	0.792	0.280	1.806	0.092	2.300
		(1.193)	(2.841)	(0.863)	(1.202)	(0.399)	(2.538)	(0.133)	(3.086)
	Bivariate	0.681	1.206	0.514	0.630	0.247	1.114	0.090	1.280
		(0.934)	(1.759)	(0.748)	(0.890)	(0.347)	(1.631)	(0.130)	(1.807)
	AR1	0.580	0.908	0.457	0.564	0.227	0.896	0.085	1.044
		(0.791)	(1.319)	(0.700)	(0.814)	(0.319)	(1.250)	(0.124)	(1.486)

Table 9: Average, over all 50 small areas, of length and coverage of credible intervals for year-to-year change.

		$\alpha = 0.10$			$\alpha = 0.05$		
		Direct	Bivariate	AR1	Direct	Bivariate	AR1
$\rho = 0$	Coverage	0.901	0.896	0.897	0.950	0.948	0.948
	Length	3.412	2.754	2.660	4.066	3.282	3.169
$\rho = 0.5$	Coverage	0.901	0.918	0.899	0.950	0.949	0.949
	Length	3.391	2.722	2.437	4.041	3.244	2.904

Table 8 compares the MSE of the different predictors of the current year small area mean for select small areas. The states presented in Table 8 are chosen to show a variety in the size (or estimated sampling variance) of the small areas, ranging from very small (Alaska and Deleware) to very large (Texas). As would be expected, the MSE of the predictions, regardless of which estimation method is used, or which value of ρ represents the truth, are nearly identical for the large state, Texas. For the smaller states, like Alaska or Deleware, there is a noticeable reduction in the MSE of the cross-sectional EBLUPs, and a further large gain in precision using the entire data set and a small area model with an AR1 process linking parameters over time.

4.3 Interval estimation for year-to-year change

This section looks at interval estimates for the parameters $\theta_{i,T} - \theta_{i,T-1}$ for each of the 50 small areas. The same three methods for interval estimation discussed in Section 3.2 are considered here. The first is based only on the direct estimates $y_{i,t}$ and their sampling variances $V_{i,t}$, where the sampling scheme is assumed independent between years. The second method uses EBLUPs based only on cross sectional data, and an assumed bivariate normal structure on the model errors over time, along with the moment estimator for the model covariance, given in equation (3). The third method uses EBLUPs based on the entire data set, and the Rao-Yu model (7). These methods are denoted Direct, Bivariate, and AR1, respectively, and the results are presented in Tables 9 and 10.

Table 9 shows the average coverage and average interval length over all 50 small areas in the simulation study, with values of $\alpha = 0.1$ and $\alpha = 0.05$. All three methods produce intervals which have appropriate coverage, on average (with the exception that the bivariate

Table 10: Average length and coverage of credible intervals for year-to-year change for select states ($\alpha = 0.05$).

	State	AK	DE	LA	MS	NC	RI	TX	WY
$\rho = 0$	Direct	0.944	0.951	0.935	0.954	0.955	0.938	0.952	0.961
		6.958	7.913	3.779	5.380	2.797	6.873	1.675	7.892
	Bivariate	0.954	0.957	0.933	0.955	0.954	0.934	0.957	0.947
		4.623	4.923	3.324	4.361	2.561	4.756	1.620	5.116
	AR1	0.952	0.947	0.934	0.953	0.957	0.944	0.959	0.941
		4.331	4.663	3.241	4.200	2.507	4.517	1.606	4.839
$\rho = 0.5$	Direct	0.953	0.936	0.954	0.944	0.940	0.946	0.959	0.951
		5.186	7.945	4.225	5.628	2.742	7.281	1.605	8.643
	Bivariate	0.946	0.927	0.950	0.934	0.943	0.945	0.965	0.948
		3.987	4.843	3.607	4.504	2.517	4.842	1.558	5.253
	AR1	0.947	0.939	0.954	0.927	0.943	0.958	0.964	0.954
		3.462	4.062	3.276	3.962	2.350	4.082	1.517	4.384

method was somewhat conservative in the example with $\rho = 0.5$ and $\alpha = 0.1$), but the intervals based only on the direct estimates have an average interval length which is much larger than the other two methods. When the true value of the autoregressive parameter ρ is 0, there is little gain constructing intervals using 6 years of data, as opposed to using only the current year's cross sectional data, as both methods achieve the nominal rate of coverage and the average length of the intervals are nearly the same. However, when $\rho = 0.5$, there is some gain in terms of average length of the credible intervals, with on average, a 10% reduction in the length of the interval.

The reduction in interval length can be more dramatic in individual small areas. Table 10 gives the coverage rate and average length of the credible intervals for each of the three methods for select small areas. Table 10 shows that both model-based methods produce intervals with reduced length in states with small and moderate sample size (there is little gain with larger states, such as Texas), in both examples for $\rho = 0$ or $\rho = 0.5$. The intervals based on 6 years of data have greater gains in terms of average length over the model-based intervals based only on a single year's data for smaller states (or states with large sampling variance) such as Delaware or Wyoming when $\rho = 0.5$. There is only a small reduction in interval length when using all 6 years of data when $\rho = 0$, even for the smallest states.

5. Conclusion

There were two main goals of this paper. The first was to investigate potential gains in efficiency using multiple years of survey data for estimation of a small area parameter. The second goal was to estimate the increase or decrease over time of a small area parameter, and to construct valid credible intervals for the change over time.

The direct survey estimates, along with associated estimated sampling variances, can be used to make such inferences. However, when the sampling variances are large, as is typically the case for some of the small areas, the lengths of the intervals are too large to be useful. Extensions of the Fay-Herriot model can be used to estimate the difference of small area parameters over time, and to construct credible intervals which are shorter in length than those based only on the direct estimates and sampling variances.

The examples in this paper focused on data collected from six time points or less. It seems important to have a reasonably long time horizon to capture a time trend in the small areas. In particular, the data example showed difficulty in estimating the autoregressive parameter ρ, with all methods producing unreasonable estimates when using only 4 years of data.

An issue that was observed in the data analysis is that it can be the case that direct estimates in a small area (or the single year EBLUPs) over time do not match the pattern, in terms of increase or decrease, of the multi-year EBLUPs. This is an important practical issue that needs to be considered, since introducing new data changes the EBLUPs of all small areas over all previous years when using the Rao-Yu model. Statistical agencies may be reluctant to revise estimates from previous years, and if a small area model incorporates a time series component, it is not clear what meaning the previous year's estimates will have if they are not updated when new survey data is introduced.

Using the estimated MSE as was done in this paper to construct credible intervals will guarantee that the intervals are first order correct, that is, have coverage equal to the nominal level, with error of order $1/m$, where m is the number of small areas (Chatterjee et al., 2008). With the Fay-Herriot model, it was shown that corrections to the intervals can be made to guarantee accuracy to within order of $m^{-3/2}$, or 'second order' correct (Diao et al., 2014; Yoshimori and Lahiri, 2014). Future work involves investigating corrections to the credible intervals for change in parameters over time which reduce the length of the interval estimate, while maintaing coverage probabilities. Other areas of future work include generalizations to spatio-temporal models, generalized linear models, and multivariate extensions.

References

Wesley Basel, Sam Hawala, and David Powers. Serial comparisons in small domain models: a residual-based approach. Technical report, U. S. Census Bureau, 2010. Available at `http://www.census.gov/did/www/saipe/publications/files/BaselHawalaPowers2010asa.pdf`.

Snigdhansu Chatterjee, Partha Lahiri, and Huilin Li. Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Ann. Statist.*, 36(3):1221 – 1245, 2008.

G. S. Datta, P. Lahiri, T. Maiti, and K. L. Lu. Hierarchical Bayes estimation of unemployment rates for the states of the U. S. *J. Amer. Statist. Assoc.*, 94(448):1074 – 1082, 1999.

G. S. Datta, P. Lahiri, and T. Maiti. Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *J. Statist. Plann. Inference*, 102:83 – 97, 2002.

Lixia Diao, David D. Smith, Gauri Sankar Datta, Tapabrata Maiti, and Jean D. Opsomer. Accurate confidence interval estimation of small area parameters under the Fay-Herriot model. *Scand. J. Statist.*, 41:497 – 515, 2014.

R. E. Fay and R. A. Herriot. Estimates of income for small places: an application of James-Stein procedure to census data. *J. Amer. Statist. Assoc.*, 74:269 – 277, 1979.

Carolina Franco and William R. Bell. Borrowing information over time in binomial/logit normal models for small area estimation. *Statist. Trans.*, 16(4):563 – 584, 2015.

D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.*, 72:320 – 338, 1977.

Sam Hawala and Partha Lahiri. Hierarchical Bayes estimation of poverty rates. Technical report, U. S. Census Bureau, 2012. Available at `http://www.census.gov/did/www/saipe/publications/files/hawalalahirishpl2012.pdf`.

Guangquan Li, Nicky Best, Anna Hansell, Ismail Ahmed, and Sylvia Richardson. BaySTDetect: detecting unusual temporal patterns in small area data via Bayesian model choice. *Biostatistics*, 13(4):695 – 710, 2012.

Yolanda Marhuenda, Isabel Molina, and Domingo Morales. Small area estimation with spatio-temporal Fay-Herriot models. *Comput. Statist. Data Anal.*, 58:308 – 325, 2013.

S. G. Pantula and K. H. Pollock. Nested analysis of variance with autocorrelated errors. *Biometrics*, 41:909 – 920, 1985.

Luis N. Pereira and Pedro S. Coelho. Small area estimation using a spatio-temporal linear mixed model. *REVSTAT*, 10(3):285 – 308, 2012.

Aaron Porter, Scott H. Holan, Christopher K. Wikle, and Noel Cressie. Spatial Fay-Herriot models for small area estimation with functional covariates. Technical report, 2014. Available at `http://arxiv.org/pdf/1303.6668v3.pdf`.

N. G. N. Prasad and J. N. K. Rao. The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, 85:163 – 171, 1990.

J. N. K. Rao. *Small Area Estimation*. Wiley, Hoboken, New Jersey, 2003.

J. N. K. Rao and Mingyu Yu. Small-area estimation by combining time-series and cross-sectional data. *Canad. J. Statist.*, 22(4):511 – 528, 1994.

Ayoub Saei and Ray Chambers. Small area estimation under linear and generalized linear mixed models with time and area effects. Technical report, Southampton Statistical Sciences Research Institute, 2003. Available at `http://eprints.soton.ac.uk/8165/1/8165-01.pdf`.

Bharat Bhushan Singh, Girja Kant Shukla, and Debasis Kundu. Spatio-temporal models in small area estimation. *Survey Methodology*, 31(2):183 – 195, 2005.

Jasen Taciak and Wesley Basel. Time-series cross-sectional approach for small area poverty models. Technical report, U. S. Census Bureau, 2012. Available at `http://www.census.gov/did/www/saipe/publications/files/JTaciakBaseljsm2012.pdf`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

Mahmoud Torabi and Farhad Shokoohi. Likelihood inference in small area estimation by combining time-series and cross-sectional data. *J. Multivariate Anal.*, 111:213 – 221, 2012.

Masayo Yoshimori and Partha Lahiri. A second-order efficient empirical Bayes confidence interval. *Ann. Statist.*, 42(4):1233 – 1261, 2014.