INVITED PAPER

# Small Area Estimation with Spatiotemporal Mixed Model

**Divina Gracia L. Del Prado**
*Philippine Statistics Authority*

**Erniel B. Barrios**
*School of Statistics, University of the Philippines Diliman*

A spatiotemporal model with nested random effects is proposed for small area estimation where sample data are generated from a rotating panel survey. Two methods of estimation are introduced, integrating the backfitting algorithm and bootstrap procedure in two different approaches. Simulation study shows superior predictive ability of the fitted model. The small area estimation methods also produced efficient estimates of parameters in a wide class of population scenarios. The model-based small area estimation procedure is also better over the design-based approach in estimating unemployment rate from the Philippine Labor Force Survey.

*Keywords: spatiotemporal mixed model; small area estimation; backfitting algorithm; bootstrap.*

## 1. Introduction

Many surveys are designed to provide reliable estimates at larger domains due to operational cost and time involved in conducting surveys. As a result, the resulting sample size is inadequate to estimate characteristics of smaller units than survey domains. This necessitate estimation procedures outside the framework of design-based estimation and using auxiliary information to complement survey generated data, e.g., small area estimation.

Small area estimation (SAE) methodology provides estimates of the target indicators for areas where sample size is insufficient to generate estimates with adequate precision using direct or design-based estimators. Direct estimators are dependent on survey data (and the sample design weights based on the induced sampling distribution of the sample selection procedures) collected from the domain. Though direct estimators are unbiased, they yield large sampling errors when the sample size is small especially for indicators with rare occurrence such as unemployment. Indirect estimators like model-based and model-assisted estimators rely on survey data, borrowed information from neighboring areas, and other available data (auxiliary information) for the target area. They are usually biased but have smaller variances than direct estimators.

In the Philippines, official statistics generated from household-based surveys conducted by the Philippine Statistics Authority (PSA), are estimated based on direct methods, i.e., design-unbiased techniques. Sample size allows only for the generation of reliable estimates at the national and regional levels but Local Government Units (LGUs) and policy makers also needed data at smaller geographic unit for a more responsive policies and programs. The PSA came out with official statistics on poverty at the provincial and city/municipality levels but demand for reliable estimates at lower geographic disaggregation does not only involve poverty statistics but also on other indicators that has to do with quality of life.

We developed in this paper a method of small area estimation using data generated from a rotated panel survey. Specifically, we postulate a model that incorporates spatial and temporal dependencies, along with nested random effects, to describe the data generation dynamics in a rotated panel survey. This model is then used in the formulation of a model-based procedure for small area estimation in rotated panel data and used specifically on unemployment rates from the Labor Force Survey.

## 2. Some Statistical Inference Tools in Small Area Estimation

Salvati (2004) classified small area estimation according to the framework of inference as design-based, model-assisted, and model-based. Design-based is also called direct method while model-assisted and model-based are called indirect methods.

The inability of direct estimators to generate reliable statistics at lower levels of geographic disaggregation and for areas with insufficient sample size in national surveys has prompted research on design-based methods. Singh et al. (1994) as cited in Rao (2004) proposed for the replacement of clusters by list frames, use of many strata to control sample size at smaller areas and compromise sample allocation (not proportional to the subgroup population size). Marker (2001) recommended stratification, oversampling in smaller areas and combination of data from multiple years. Rao (2004) pointed out that even after implementing measures at the design stage, sample sizes may still be insufficient to generate reliable estimates for some small areas.

Indirect estimators were used in small area estimation to mitigate the inadequacies observed in direct estimators. Several approaches such as synthetic estimator, composite estimator and generalized regression estimator (GREG) were postulated for model-assisted estimators. Singh and Goel (2000) as cited in Rao (2004) used synthetic estimators to estimate crop yields in India at the tehsil level and Eklund (1998) used composite estimators to estimate net coverage error of the 1997 US Census of Agriculture. The GREG estimator is used in the weighting procedure of the Dutch Labor Force Survey (Brakel and Krieg, 2007). However, since model-assisted estimators are based on fixed effects models, they fail to take into account the variability between areas which contributes to the reliability of the small area estimates.

To consider the random variability between areas, random effects are included in the model in addition to fixed effects contributed by some covariates, leading to the use of linear mixed model in small area estimation. The fixed effects factor assumes that there is a finite set of levels for the factor where all levels are of interest in the study while the random effects factor assumes that there is a population of levels for the factor and only samples of levels are considered. Rao and Yu (1994) postulated a random effects model with error following an Autoregressive (AR) process of order 1 to small area estimation of time series cross-sectional data. To account for the increasing variability over time, the model was later extended by Pereira and Coelho (2007) assuming a heterogeneous covariance structures for the random effects where the heterogeneous compound symmetry (CSH) and first-order autoregressive conditional heteroskedasticity (ARCH (1)) were selected.

Data from surveys usually exhibit spatial dependence or correlation of the same attributes in neighboring areas. With the limitation of linear mixed model to address the spatial nature of survey data, spatial models were included among classes of models postulated for small area estimation. Salvati (2004) postulated an area level model that considers the spatial dependence of the data by incorporating spatially correlated error structure in a standard regression model for small area estimation called Spatial Empirical Best Linear Unbiased Predictor (Spatial EBLUP). With spatial EBLUP, there is an appreciable improvement of the statistical properties of the small area estimators. Coelho and Pereira (2011) proposed a spatial unit level model that takes into account the spatial correlation as a function of distance between small areas in the context of small area estimation problem. The estimator considers auxiliary information about some known variables in the population and a non-diagonal structure for the variance-covariance matrices of the random effects. The estimator yields higher precision than synthetic estimators for very small domains and reduces bias relative to estimators that ignore spatial correlation. The proposed model allows estimation at the lowest possible geographic domain and for small domains that are not represented in the sample.

Spatial models fail to account for the temporal dependency of data produced from panel surveys. Several works have confirmed the advantage of spatiotemporal models over spatial models for time series data. Singh et al. (2005) proposed a spatiotemporal model in the framework of mixed effects linear model with constant regression parameter and autocorrelation parameter in the context of Kalman filters to improve small area direct estimators at any point in time. The monthly consumption expenditure data from 1993-2000 generated by the National Sample Survey Organization of India confirmed the gains obtained in incorporating temporal effects in the model as compared with the direct estimates, ordinary least squares, spatial model and various mixed effects models. Landagan and Barrios (2006) developed a regression model that accounts for spatial and temporal dependencies with parameter estimation procedure following the backfitting algorithm described by Hastie and Tibshirani (1990), incorporating the Cochranne-Orcutt procedure.

The backfitting algorithm is an iterative procedure for fitting additive models, in which, at each step, one component is estimated holding the effects of other components constant and iterating until convergence (Ansley and Kohn, 1992). The convergence of backfitting algorithm was shown by Buja et al. (1989) for bivariate additive model and Ansley and Kohn (1992) in the context of spline smoothers and structural time series models when the additive components are estimated by penalized least squares. Landagan and Barrios (2006) also noted the convergence of backfitting algorithm embedded with Cochranne-Orcutt procedure after second or third iteration. Consistency and asymptotic properties of backfitting for nonparametric mixed effects model were exhibited by Park and Wu (2004). Guarte (2009) established the consistency and asymptotic normality of the basic estimators of the spatial and temporal parameters estimated through the backfitting algorithm.

The bootstrap method introduced by Efron in 1979, is used to obtain approximation of the empirical distribution of certain statistics. Bootstrap, along with other resampling methods, were used to obtain estimate of bias and dispersion of the estimator. A three-step Monte Carlo algorithm for bootstrap estimation of standard error was given by Efron and Tibshirani (1986). The bootstrap estimate $\hat{\sigma} = \sigma(\hat{F})$ is also called maximum likelihood estimate (MLE) of $\sigma$ since $\hat{F}$ is the nonparametric MLE of the unknown distribution $F$. To obtain the bootstrap estimates of the covariate, spatial and temporal effects, Landagan and Barrios (2006) used the average estimates for the $N$ locations with regards to the covariate and spatial effects and average of the $T$ time points for the temporal effect. Dumanjug et al. (2010) modified this bootstrap estimation procedure by utilizing the $N$ and $T$ estimates for the parameters mentioned. The procedure was shown to be robust to structural change and work better specially for longer time series data.

Shao and Tu (1995) have shown that bootstrap estimators for sampling distribution when $X_1, X_2, \ldots, X_n$ are i.i.d. and $X_1$ have finite mean ($E\|X_1\| < \infty$) is strongly consistent. Guarte (2009) also established consistency and asymptotic normality of bootstrap estimators in the spatiotemporal model. Bootstrap estimators are less susceptible to model violation but not when resampling based on the residuals (Shao and Tu, 1995), confirmed by Guarte (2009). In order to determine the appropriate number of bootstrap replicates, Efron and Tibshirani (1986) provided a formula that expresses the coefficient of variation (CV) of the bootstrap standard error estimator and the number of bootstrap replicates given as:

$$CV\left(\hat{\sigma}_B\right) = \sqrt{CV(\hat{\sigma})^2 + \frac{E\{\hat{\delta}\}+2}{4B}} \tag{1}$$

where $\delta$ is the kurtosis of the distribution of the statistic $\hat{\theta}^*$. Furthermore, in estimating standard errors, an adequate number of bootstrap replicates ranges from 50 to 200 for most situations.

Chernick (1999) as cited by Guarte (2009), suggested 30 as the minimum sample size in applying bootstrap. In the context of spatiotemporal model, Guarte (2009) confirmed previous findings that a sample size of 50 is sufficient to yield correct inference in testing for constant temporal effects.

## 3. Spatiotemporal Modeling in Rotated Panel Data

The proposed spatio-temporal mixed model is a modification of the spatiotemporal model of Landagan and Barrios (2006), taking into account the sampling design typically used in large scale national surveys for the generation of official statistics. These surveys usually adopt a multi-stage sampling design and a rotation scheme in the selection of samples every survey period (e.g., quarterly).

The nested random effects and the spatial and temporal dependencies of panel data are considered as follows:

$$\begin{aligned} &Y_{it} = \alpha + \beta X_i + u_i + \mathrm{v}_{it} + \varepsilon_{it} \\ &i = 1, \ldots, N,\ t = 1, \ldots, T \end{aligned} \tag{2}$$

where $Y_{it}$ is a continuous response variable for unit $I$ at time $t$, $X_i$ is the set of spatial indicators for unit $i$ which is common to all units in the group and at all time points, $\beta$ is the constant spatial effect across time, $u_i$ is the first level random effect of area $i$, $v_{it}$ is the second level random effect of area $i$ at time $t$ where $v_{it}$ is nested within $u_i$.

The assumptions for the random effect components are: $u_i \sim IID(0,\sigma_u^2)$ and $v_{it} \sim IID(0,\sigma_v^2)$. The error term $\varepsilon_{it}$ follows an autoregressive process of order 1, i.e., $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + a_{it}$ where $\rho$ is the constant temporal effect across areas, $|\rho| < 1$, $a_{it} \sim IID(0,\sigma_a^2)$ and $\varepsilon_{i0} \sim N\left(0,\sigma_a^2/\left(1-\rho^2\right)\right)$. The random effects ($u_i$ and $v_{it}$) and $\varepsilon_{it}$ are independent.

In matrix form, the model in (2) can be represented as a general linear mixed model:

$$\boldsymbol{Y} + \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{\psi} + \boldsymbol{\varepsilon} \qquad (3)$$

where $\boldsymbol{Y}$ is the $NT$ x 1 vector of responses, $\boldsymbol{X}$ is the $NT$ x $p$ matrix of spatial indicators, $\boldsymbol{\beta}$ is the $p$ x 1 vector of parameters for the spatial effects, $\boldsymbol{Z}$ is the $NT$ x $q$ design matrix, $\boldsymbol{\psi}$ = [**u v**]' is the $q$ x 1 vector of random effects and $\varepsilon$ is the $NT$ x 1 vector of random errors which follow an AR(1) process or

$$\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T} \\ \vdots \\ \varepsilon_{N1} \\ \varepsilon_{N2} \\ \vdots \\ \varepsilon_{NT} \end{bmatrix} = \rho \begin{bmatrix} \varepsilon_{10} \\ \varepsilon_{11} \\ \vdots \\ \varepsilon_{1,T-1} \\ \vdots \\ \varepsilon_{N0} \\ \varepsilon_{N1} \\ \vdots \\ \varepsilon_{N,T-1} \end{bmatrix} + \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1T} \\ \vdots \\ a_{N1} \\ a_{N2} \\ \vdots \\ a_{NT} \end{bmatrix}$$

The assumptions for $\psi$ and $\varepsilon$ are $\begin{bmatrix} \psi \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D & 0 \\ 0 & \Sigma \end{bmatrix}\right)$, where **D** and **Σ** are the variance-covariance matrices for $\boldsymbol{\psi}$ and $\boldsymbol{\varepsilon}$, respectively which are given as $D = \begin{bmatrix} \sigma_u^2 I_N & 0 \\ 0 & \sigma_v^2 I_{NT} \end{bmatrix}$, $\Sigma = \left(\sigma_a^2 / 1 - \rho^2\right) I_{NT}$. The variance for $\boldsymbol{Y}$ is $\boldsymbol{V} = \boldsymbol{ZDZ}' + \boldsymbol{\Sigma}$.

Data from the LFS can be characterized by the spatiotemporal mixed model in equation (3). The two-stage random selection of sample units (PSU in the first stage and rotation group or group of housing units within sampled PSU in the second stage) in the LFS is described in the proposed model by the two-level random effects that are nested. In the LFS, sample housing units are rotated quarterly to mitigate response-burden problem as an effect of repeatedly interviewing the same households in every survey rounds. The rotation is reflected in the manner the second-level random effects are coded quarterly in the dataset. The collection of data in successive intervals often results to correlated errors over time, hence, the temporal effect.

In the case of unemployment data generated from the LFS, the city/municipality per capita income is a suitable indicator that defines the neighborhood system. It reflects the financial capability of the Local Government Units to support projects and other priority needs in their locality, including the generation of employment. Thus, barangays or Primary Sampling Units (PSUs) within the same city/municipality will likely have a similar employment situation. The city/municipality per capita income is used as one of the implicit stratification variables for the LFS because of its significant correlation to the proportion of poor households.

This paper adopts the backfitting algorithm in estimating the parameters of the spatiotemporal mixed model. Algorithm 1 presents the steps of the backfitting algorithm in estimating the parameters of the model in equation (3). In the algorithm, the parameters for the spatial effect are estimated first, followed by the random effects and variance components, and the temporal effect and error variance in the last step.

### 3.1 Estimation of fixed effects

Ignoring the random effects $u_i$ and $v_{it}$ and the temporal parameter $\rho$, the parameters $\alpha$ and $\beta$ are estimated using Ordinary Least Squares (OLS) from the model:

$$Y_{it} = \alpha + \beta X_i + \varepsilon_{it}^{(*)} \tag{4}$$

then compute the residuals $\varepsilon_{it}^{(1)} = Y_{it} - \hat{Y}_{it} = Y_{it} - (\hat{\alpha} + \hat{\beta} X_i)$. In matrix form equation (4) is expressed as:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}^{(*)} \tag{5}$$

The least squares estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$, which is biased since $\varepsilon^{(*)}$ still contains information on the random and temporal effects.

### 3.2 Estimation of random effects

The variance components and the random effects are estimated from the residuals after the initial estimates of the fixed effects $\boldsymbol{\beta}$ in Section 3.1. Thus,

$$Y_{it}^{(1)} = \varepsilon_{it}^{(1)} = Y_{it} - (\hat{\alpha} + \hat{\beta} X_i) = u_i + v_i + \varepsilon_{it}^{(**)} \tag{6}$$

is a random effects model with no constant. In matrix form, equation (6) can be written as:

$$Y^{(1)} = Z\psi + \varepsilon^{(**)} \tag{7}$$

where $\begin{bmatrix} \psi \\ \varepsilon^{(2)} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D & 0 \\ 0 & \Sigma \end{bmatrix} \right)$ and $\psi = \begin{bmatrix} u \\ v \end{bmatrix}$, $Var(\psi) = D = \begin{bmatrix} \sigma_u^2 I_N & 0 \\ 0 & \sigma_v^2 I_{NT} \end{bmatrix}$, $\boldsymbol{Y}^{(1)} \sim N(0, \boldsymbol{V}^{(1)})$ and $\boldsymbol{V}^{(1)} = \boldsymbol{ZDZ'} + \boldsymbol{\Sigma}$. The variance components $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$ are estimated through Restricted Maximum Likelihood (REML) while the random effects $\boldsymbol{\psi}$ is estimated through the Empirical Best Unbiased Predictor (EBLUP) given by:

$$\hat{\boldsymbol{\psi}}_{EBLUP} = \hat{\boldsymbol{D}}\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}(\hat{\boldsymbol{Y}}^{(1)} - \boldsymbol{X}\hat{\boldsymbol{B}}) \tag{8}$$

Note that $\hat{\boldsymbol{\psi}}_{EBLUP}$ is also biased since $\varepsilon^{(**)}$ in equation (7) contains information on the temporal effect $\rho$.

### *3.3 Estimation of temporal parameters*

Computing the residuals after estimating β and ψ in Sections 3.1 and 3.2,

$$Y_{it}^{(2)} = Y_{it} - (\hat{\alpha} + \hat{\beta}X_i + \hat{\mu}_i + \hat{\upsilon}_{it}) = \varepsilon_{it}^{(**)} = \rho\varepsilon_{i,t-1}^{(**)} + a_{it},\ |\rho| < 1 \tag{9}$$

The parameters $\rho$ and $\sigma_a^2$ are estimated from equation (9) via Maximum Likelihood Estimation (MLE) method.

### 3.4 Iterative Procedure

To address the bias of the estimates in Sections 3. 1 to 3. 3, the backfitting procedure is implemented in Algorithm 1.

*Algorithm 1*

***Initialization phase***

Step 1: The parameters $\alpha$ and $\beta$ are simultaneously estimated from the regression equation in (5) using OLS. Compute $\hat{Y}_{it} = \hat{\alpha} + \hat{\beta}X_i$. Recall that estimates of $Y_{it}$ in this step are biased since the resulting residuals still contain information on the random and temporal effects.

Step 2: Compute $\varepsilon_{it}^{(1)} = Y_{it} - \hat{Y}_{it}$. This can be represented as a random effects model in equation (7). Estimate the variance components of the random effects $\sigma_u^2$ and $\sigma_v^2$ and predict the random effects $u_i$ and $v_{it}$ using EBLUP then compute $\hat{\varepsilon}_{it}^{(1)} = \hat{\mu}_i + \hat{v}_{it}$. Residuals at this point contain the information on the temporal effect.

Step 3: Compute $\varepsilon_{it}^{(2)} = \varepsilon_{it}^{(1)} - \hat{\varepsilon}_{it}^{(1)}$. This can be modeled as an AR(1) process as in equation (9) to estimate the temporal parameter $\rho$ and variance $\sigma_a^2$.

***Iteration phase***

Step 4: Compute new residuals, denoted by $Y_{it}'$, by removing from $Y_{it}$ the effects of the random and temporal effects using their latest estimates. That is, $Y_{it}' = Y_{it} - \hat{\mu}_i - \hat{v}_{it} - \hat{\varepsilon}_{it}$ where $\hat{\varepsilon}_{it} = \hat{\rho}\varepsilon_{i,t-1}$ and $\varepsilon_{i,t}$ are the latest residuals after removing from $Y_{it}$ the latest estimates of fixed and random effects.

From the $Y'_{it}$, obtain new estimates of $\hat{\alpha}$ and $\hat{\beta}$. The bias incurred in Step 1 is mitigated gradually at this point.

Step 5: Recompute $\varepsilon_{it}^{(1)}$, denoted by $\varepsilon_{it}^{(1)\prime}$, by subtracting from $Y_{it}$ the effects of the spatial and temporal components using their latest estimates, i.e., $\varepsilon_{it}^{(1)\prime} = Y_{it} - \hat{\alpha} - \hat{\beta} X_i - \hat{\varepsilon}_{it}$ where $\hat{\varepsilon}_{it} = \hat{\rho}\varepsilon_{i,t-1}$. Compute new predictions of the random effects $u_i$ and $v_{it}$ and new estimates of the variance components from $\varepsilon_{it}^{(1)\prime}$. Again, bias from Step 2 is lessened at this step.

Step 6: Recompute $\varepsilon_{it}^{(2)}$, denoted by $\varepsilon_{it}^{(2)\prime}$, by subtracting from $Y_{it}$ the effects of the spatial and random components using their latest estimates, i.e., $\varepsilon_{it}^{(2)\prime} = Y_{it} - \hat{\alpha} - \hat{\beta} X_i - \hat{\mu}_i - \hat{\upsilon}_{it}$. Compute new estimates of $\rho$ and $\sigma_a^2$ from $\varepsilon_{it}^{(2)\prime}$.

Step 7: Repeat steps 4 to 6 until convergence.

Algorithm 1 implements the hybrid of OLS, EBLUP, and MLE in the backfitting framework known to be consistent and convergent, see for example Buja et al (1989) and Opsomer (1999).

## 4. Model-Based Small Area Estimation Procedure

The spatiotemporal mixed model presented in Section 3 is used in small area estimation of continuous data such as rates, ratios and averages denoted as $\overline{Y}_{dt}$ and expressed as:

$$\overline{Y}_{dt} = N_d^{-1} Y_{dt} = N_d^{-1} \sum_{i=1}^{N_d} Y_{it} \tag{10}$$

where $N_d$ is the total number of units in small area $d$ and $Y_{dt}$ is the sum of the realizations of the target indicator in small area $d$ at time $t$. The sum $Y_{dt}$ can be decomposed as a sum of the observed and unobserved values where the subscript $s$ pertains to the sampled units and $n$ to the non-sampled units as follows:

$$Y_{dt} = Y_{dt,s} + Y_{dt,n} \tag{11}$$

where $Y_{dt,s}$ is the observed sum from the samples within the small area $d$ at time $t$ and $Y_{dt,n}$ is the unobserved sum of the non-sampled units in small area $d$ at time $t$. Since $Y_{dt,s}$ is known, the small area estimator for $Y_{dt}$ is:

$$\hat{Y}_{dt} = Y_{dt,s} + \hat{Y}_{dt,n} \tag{12}$$

Hence, to obtain a small area estimate of the total, we need to estimate the unobserved sum $\hat{Y}_{dt,n}$.

### 4.1 General form of the estimator for the unobserved sum $Y_{dt,n}$

Under the spatiotemporal mixed model in equation (3), the unobserved sum $Y_{dt,n}$ expressed as:

$$Y_{dt,n} = \sum_{i=1}^{N_d - n_d} (\alpha + \beta X_i + \mu_i + \nu_{it} + \rho\varepsilon_{i,t-1} + a_{it}) \tag{13}$$

where $n_d$ is the number of sampled units in domain $d$, $N_d$ is the total number of units in domain $d$ and $N_d - n_d$ is the number of non-sampled units in domain $d$.

Summing up equation (13) by component,

$$Y_{dt,n} = \left( (N_d - n_d)\alpha + \beta \sum_{i=1}^{N_d - n_d} X_i \right) + \left( \sum_{i=1}^{N_d - n_d} (\mu_i + \nu_{it}) \right) + \left( \rho \sum_{i=1}^{N_d - n_d} \varepsilon_{i.t-1} \right) + \sum_{i=1}^{N_d - n_d} a_{it} \tag{14}$$

Hence, the general form of the estimator for $Y_{dt,n}$ is:

$$\hat{Y}_{dt,n} = \left( (N_d - n_d)\hat{\alpha}_s + \hat{\beta} \sum_{i=1}^{N_d - n_d} X_{i,n} \right) + \left( \sum_{i=1}^{N_d - n_d} (\hat{\mu}_i + \hat{\nu}_{it}) \right) + \left( \hat{\rho}_s \sum_{i=1}^{N_d - n_d} \varepsilon_{i,t-1} \right) \tag{15}$$

where the subscript $s$ in the estimators indicates that estimates for these parameters are based on the available information from the sampled units. The covariate $X_{i,n}$ indicates the value of the $i^{\text{th}}$ non-sampled unit for the spatial variable $X$.

Two methods are proposed for the estimation of $Y_{dt,n}$ with estimator in equation (15). Both methods adopt the backfitting and bootstrap procedures in two different approaches. *Method 1* uses backfitting and bootstrap procedure in obtaining the estimates for the spatial and temporal effects denoted as $\hat{\alpha}_s, \hat{\beta}_s$ and $\hat{\rho}_s$ in equation (15). The estimates for the variance components of the random effects $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$, and the variance of the error $\hat{\sigma}_a^2$ component are also obtained using bootstrap methods. The estimates of the variance components are used in the prediction of the random effects $\hat{\mu}_i$ and $\hat{v}_{it}$ for the non-sampled units and $\hat{\sigma}_a^2$ in the reconstruction of the error structure $\varepsilon_{it}$. *Method 2* uses the backfitting algorithm in obtaining the values for $\hat{\alpha}_s$, $\hat{\beta}_s$, $\hat{\rho}_s$, $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_a^2$. The unobserved sum $Y_{dt,n}$ is estimated using bootstrap procedure where the unobserved sum for each bootstrap replicate $Y_{dt,n}$ (b) is computed using equation (15).

For both methods, the two-level random effects $\hat{\mu}_i$ and $\hat{v}_{it}$ for the non-sampled units are independently generated by random sampling from their estimated distributions. This is done since the EBLUP for the random effects cannot be generated in the absence of the observed values for the response variable in the non-sampled units. The $\varepsilon_{it}$'s are reconstructed from the fitted AR(1) where values for $a_{it}$'s are simulated from the estimated distribution of the residuals $a_{it}$.

## *4.2 Method 1 (Bootstrap estimators for the model parameters)*

In Method 1, parameters are estimated using bootstrap methods summarized in Algorithm 2.

***Algorithm 2***

Step 1: Draw $B$ bootstrap replicates of size $N$ with replacement from the dataset with $N$ observations.

Step 2: For each bootstrap replicate, estimate the parameter $\theta$ using Algorithm 1.

Step 3: Calculate the bootstrap estimate of $\theta$ using the estimator:

$$\hat{\theta}_B = B^{-1}\sum_{b=1}^{B}\hat{\theta}(b) \tag{16}$$

where $\theta$ refers to the parameter, $\hat{\theta}_B$ is the bootstrap estimator for $\theta$ and $\hat{\theta}(b)$ is the estimator of $\theta$ in bootstrap replicate $b$.

The prediction of the sum for the non-sampled units $Y_{dt,n}$ using Method 1 is obtained by the following equation:

$$\hat{Y}_{dt,n} = \left((N_d - n_d)\hat{\alpha}_{s,B} + \hat{\beta}_{s,B}\sum_{i=1}^{N_d-n_d} X_{i,n}\right) + \left(\sum_{i=1}^{N_d-n_d}(\hat{\mu}_i + \hat{\upsilon}_{it}\right) + \left(\hat{\rho}_{S,B}\sum_{i=1}^{N_d-n_d}\varepsilon_{i,t-1}\right) \tag{17}$$

where the subscript $B$ in the estimators refers to bootstrap estimators. This is similar to equation (15) but with a more specific notation for the estimators of the parameters $\alpha_s$, $\beta_s$ and $\rho_s$. In Method 1, the bootstrap estimates for $\alpha_s$, $\beta_s$, $\rho_s$, $\sigma_u^2$, $\sigma_v^2$ and $\sigma_a^2$ are obtained after generating the basic estimates for these parameters using backfitting procedures described in Algorithm 1. Algorithm 3 describes the procedure in estimating $Y_{dt,n}$ using the equation in (15).

***Algorithm 3***

Step 1: Using the data from the sampled units, obtain the basic estimates of the constant $\alpha$, spatial effect $\beta$, temporal effect $\rho$, the variance components of the random effects $\sigma_u^2$ and $\sigma_v^2$ and the variance of the residual $\sigma_a^2$ using backfitting procedure as described in Algorithm 1.

Step 2: Obtain bootstrap estimates for the parameters in Step 1 using Algorithm 2.

Step 3: Predict the random effects $u_i$ and $v_{it}$ for the non-sampled units by drawing samples from $u_i \sim N\left(0, \hat{\sigma}_{u,s,B}^2\right)$ and $v_{it} \sim N\left(0, \hat{\sigma}_{v,s,B}^2\right)$, respectively, where $\hat{\sigma}_{u,s,B}^2$ and $\hat{\sigma}_{v,s,B}^2$ are computed in Step 2.

Step 4: The $\varepsilon_{i,t}$'s are computed from $\varepsilon_{i,t} = \hat{\rho}_{s,B}\varepsilon_{i,t-1} + a_{it}$ where the $a_{it}$'s are simulated from and $\varepsilon_{i,0}$ from $\varepsilon_{i,0} \sim N\left(0, \frac{\hat{\sigma}_{a,s,B}^2}{1-\hat{\rho}_{s,B}^2}\right)$. $\hat{\sigma}_{a,s,B}^2$ is the bootstrap estimate of the variance of the residual $a_{it}$. The values for the bootstrap estimators and are computed in Step 2.

Step 5: Obtain the sum of the sampled units in the domain ($Y_{dt,s}$). Estimate the sum for the non-sampled units in the domain ($\hat{Y}_{dt,n}$) using equation (15) and the overall sum for the domain ($\hat{Y}_{dt}$) (using equation (14). Estimate the overall mean for the domain using $\hat{\bar{Y}}_{dt} = N^{-1}\hat{Y}_{dt}$.

### *4.3 Method 2 (Bootstrap estimator for the unobserved $Y_{dt,n}$)*

The procedure in estimating $Y_{dt,n}$ using Method 2 is presented in Algorithm 4.

***Algorithm 4***

Step 1: For each bootstrap replicate, obtain the estimates of the constant α, spatial effect $\beta$, temporal effect $\rho$, the variance components of the random effects $\sigma_u^2$ and $\sigma_v^2$ and the variance of the residual $\sigma_a^2$ using backfitting procedure in Algorithm 1.

Step 2: Predict the random effects $u_i$ and $v_{it}$ by drawing samples from $\mu_i \sim N\left(0, \hat{\sigma}^2_{u,s,b}\right)$ and $v_i \sim N\ \ 0, \hat{\sigma}^2_{v,s,b}$ , respectively.

Step 3: The $\varepsilon_{i,t}$'s are computed from $\varepsilon_{i,t} = \hat{\rho}_{s,b}\varepsilon_{i,t-1} + a_{it}$ where the $a_{i,t}$'s are drawn from $a_{it} \sim N(0, \hat{\sigma}^2_{a,s,b})$ and $\varepsilon_{i,0}$ from $\varepsilon_{i,0} \sim N\left(0, \dfrac{\hat{\sigma}^2_{a,s,b}}{1-\hat{\rho}^2_{s,b}}\right)$.

Step 4: Compute $\hat{Y}_{dt,n}(b)$ using equation (19), the bootstrap estimate of the sum of the non-sampled units ($\hat{Y}_{dt,n}$) using equation (18) and the overall sum for the domain ($\hat{Y}_{dt}$) using equation (14). Estimate the overall mean for the domain using $\hat{\bar{Y}}_{dt} = N^{-1}\hat{Y}_{dt}$.

For Method 2, $Y_{dt,n}$ is estimated from:

$$\hat{Y}_{dt,n} = B^{-1}\sum_{b=1}^{B}\hat{Y}_{dt,n}(b) \tag{18}$$

where $B$ is the total number of bootstrap replicates and $\hat{Y}_{dt,n}(b)$, $b = 1,\ldots, B$, is the sum in each bootstrap replicate which follows the general estimator $\hat{Y}_{dt,n}$ in equation (15). Specifically,

$$\hat{Y}_{dt,n}(b) = \left((N_d - n_d)\hat{\alpha}_{s,b} + \beta_{s,b}\sum_{i=1,b}^{N_d-n_d} X_{i,n}\right) + \left(\sum_{i=1,b}^{N_d-n_d}(\hat{\mu}_i + \hat{v}_{it})\right) + \left(\hat{\rho}_{s,b}\sum_{i=1,b}^{N_d-n_d}\varepsilon_{i,t-1}\right) \tag{19}$$

where the subscript $b$ is a bootstrap index and $\hat{\alpha}_{s,b}, \hat{\beta}_{s,b}$ and $\hat{\rho}_{s,b}$ are the backfitting estimators for the bootstrap replicate $b$ which are obtained using the information from the sampled units.

Method 2 from Method 1 are differentiated in terms of the estimation procedures for the overall sum of the non-sampled units $Y_{dt,n}$; the model parameters $\alpha_s$, $\beta_s$ and $\rho_s$; and the variances $\sigma_u^2, \sigma_v^2$ and $\sigma_a^2$. In Method 2, the predicted sum for $Y_{dt,n}$ is obtained through bootstrap while Method 1 is not. The estimators for the model parameters $\alpha_s$, $\beta_s$ and $\rho_s$; and the variances $\sigma_u^2, \sigma_v^2$ and $\sigma_a^2$ are obtained through backfitting procedure in Method 2 while backfitting and bootstrap procedures in Method 1. The procedure in predicting the random effects and estimating the residual component are the same for both methods.

## 4.4 Precision of estimates

The SE for the small area estimate of $\overline{Y}_{dt}$ using Method 1 is model-based while Method 2 uses the bootstrap SE. Note that under the proposed spatiotemporal mixed model in (3), the variance of $Y_{it}$ is given by:

$$\sigma^2(Y_{it}) = \sigma_u^2 + \sigma_v^2 + \frac{\sigma_a^2}{1-\rho^2} \tag{20}$$

since $u_i$, $v_{it}$ and $\varepsilon_{it}$ are independent. Estimator for the variance of is $\hat{Y}_{it}$:

$$\hat{\sigma}^2(\hat{Y}_{it}) = \hat{\sigma}_u^2 + \hat{\sigma}_v^2 + \frac{\hat{\sigma}_a^2}{1-\hat{\rho}^2} \tag{21}$$

Hence, the variance of the small area estimate for $\overline{Y}_{dt}$ with Method 1 is $\hat{\sigma}^2\left(\hat{\overline{Y}}_{dt}\right) = Var\left(N_d^{-1}(\hat{Y}_{dt})\right) = Var\left(N_d^{-2}\left(\sum_{i=1}^{N_d}\hat{Y}_{it}\right)\right) = N_d^{-2}Var\left(\sum_{i=1}^{N_d}\hat{Y}_{it}\right)$.

By the independence of the $Y_{it}$'s and by equation (21), we have $\hat{\sigma}^2\ \hat{\overline{Y}}_{dt} = N_d^{-2}\sum_{i=1}^{N_d} Var\ \hat{Y}_{it} = N_d^{-2}\sum_{i-1}^{N_d}\left(\sigma^2_{u,s,B} + \hat{\sigma}^2_{v,s,B} + \frac{\hat{\sigma}^2_{a,s,B}}{1-\hat{\rho}^2_{s,B}}\right)$ and $\hat{\sigma}^2\left(\hat{\overline{Y}}_{dt}\right) = N_d^{-1}\left(\hat{\sigma}^2_{u,s,B} + \hat{\sigma}^2_{v,s,B} + \frac{\hat{\sigma}^2_{a,s,B}}{1-\hat{\rho}^2_{s,B}}\right)$. Hence, the standard error (SE) of $\hat{\overline{Y}}_{dt}$ is:

$$\hat{\sigma}\left(\hat{\overline{Y}}_{dt}\right) = \sqrt{N_d^{-1}\left(\hat{\sigma}^2_{u,s,B} + \hat{\sigma}^2_{v,s,B} + \frac{\hat{\sigma}^2_{a,s,B}}{1-\hat{\rho}^2_{s,B}}\right)} \tag{22}$$

where $\hat{\sigma}^2_{u,s,B}$, $\hat{\sigma}^2_{v,s,B}$, $\hat{\sigma}^2_{a,s,B}$ and $\hat{\rho}^2_{s,B}$ are the bootstrap estimates for and ρ, respectively, using the sample data.

For Method 2, since $\hat{Y}_{dt}$ is obtained through bootstrap, the SE of $\hat{Y}_{dt}$ is obtained using Monte Carlo SE given by:

$$\hat{\sigma}\left(\hat{\hat{Y}}_{dt}\right) = (N_d)^{-1}\hat{\sigma}(\hat{Y}_{dt,B}) = (N_d)^{-1}\sqrt{\frac{\sum_{b=1}^{B}\left(\hat{Y}_{dt}(b)-\hat{Y}_{dt,B}\right)^2}{B-1}} \tag{23}$$

where $\hat{Y}_{dt,B}$ is the bootstrap small area estimate of $Y_{dt,B}$ and $\hat{Y}_{dt}(b)$ is the small area estimate of $Y_{dt}$ for bootstrap replicate $b$.

For both methods, the estimate of CV for $\hat{\hat{Y}}_{dt}$ is:

$$\widehat{CV}\left(\hat{\hat{Y}}_{dt}\right) = \frac{\hat{\sigma}(\hat{\hat{Y}}_{dt})}{\hat{\hat{Y}}_{dt}} \tag{24}$$

## 5. Simulation Studies

Simulation studies are designed to characterize proposed small area estimation in a spatiotemporal mixed model.

### *5.1 The data-generating process*

A total of 384 scenarios were used in the simulation. The scenarios vary according to size or the number of first-level units at each time point, the true values of the model parameters and the distribution of the random effects. The different sizes considered in the simulation are $N$ = 50, 150 and 250. All the simulation datasets are balanced which means that all unit $i$'s in each dataset have four observations representing the four time points in a quarterly survey. The spatial effect $\beta$ has two values, 0.2 and 2, representing the non-dominating and dominating effects of the spatial indicator, respectively. The temporal effect $\rho$ takes on values 0.2 (stationary) and 0.9 (near nonstationary). The random effects are assumed to be normal, however, to check robustness to model misspecification of the proposed model and estimation procedures, some scenarios were set to have random effects that follow an exponential distribution. The following are the assumed values of the parameters and the distribution of the random effects:

<u>First-level random effects</u>

$\mu_i \sim N\left(0,\sigma_u^2\right)$ with values of $\sigma_u^2$ = 0.25 and 4

$u_i \sim Exp(\lambda)$ with values of $\lambda$ = 0.5 and 2 which yield true values of $\sigma_u^2$ = 0.25 and 4

Second-level random effects

$v_{it} \sim N(0, \sigma_v^2)$ with values of $\sigma_v^2 = 0.25$ and 4

$v_{it} \sim Exp(\lambda)$ with values of $\lambda = 0.5$ and 2 which yield true values of $\sigma_v^2 = 0.25$ and 4

The residual $a_{it}$ is normal with $\sigma_a^2 = 0.1$ and 4. The number of bootstrap replicates B for the bootstrap estimation is 200.

***Data generation***

1. Create small area domains (e.g., municipality), Primary Sampling Units (PSUs) as the first-level sample units and rotation groups (e.g., groups of sample housing units) as the second-level sample units. Create a time variable that contains the indicator for the four quarters. The PSUs should be the same for all quarters and the rotation group within the PSU should vary by quarter to mimic the rotation of samples.
2. Create a spatial variable $X$ simulating the values of a municipality per capita income.
3. Draw random samples for the residual $a_{it}$ and random effects ($u_i$ and $v_{it}$) from their assumed distributions.
4. Generate $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + a_{it}$ where $\varepsilon_{i,0} \sim N\left(0, \sigma_a^2/(1-\rho^2)\right)$ where $\rho$ is the assumed temporal effect.
5. Generate $Y_{it} = \alpha + \beta X_{it} + u_i + v_{it} + a_{it}$ where $\alpha$ and $\beta$ are the assumed values of the parameters.

## 5.2 Evaluation of bias and predictive ability

To assess the performance of the proposed procedures in estimating the parameters of the proposed spatiotemporal mixed model in (3), the absolute bias of the estimators was obtained using the following formula:

$$\left|Bias(\hat{\theta})\right| = \left|(\hat{\theta} - \theta)\right| \tag{25}$$

where $\hat{\theta}$ is the estimate for the parameter $\theta$. The ratio of the absolute bias to the true value of the parameter was also used to compare the bias across parameters. It is given by:

$$\frac{\left|Bias(\hat{\theta})\right|}{\theta} \tag{26}$$

The Mean Absolute Percentage Error (MAPE) was used to measure the predictive performance of the fitted model.

$$MAPE\left(\hat{Y}_{it}\right) = N^{-1}\sum_{i=1}^{N}\left|\frac{\hat{Y}_{it} - Y_{it}}{Y_{it}}\right| \tag{27}$$

where $N$ is the total number of units.

### 5.3 Convergence of the backfitting algorithm

Using one of the simulation datasets used in this study, the convergence of the backfitting algorithm confirms the fast convergence of the backfitting algorithm observed by Landagan and Barrios (2006), except for the temporal effect. The estimate for $\rho$ converged at a much slower pace (between 10-25 iterations).

### 5.4 Bias of parameter estimates

Performance of the proposed estimation procedure for the parameters of the spatiotemporal mixed model was evaluated based on bias and the predictive ability of the fitted model through the Mean Absolute Percentage Error (MAPE) presented in Section 5.5.

***Effect of sample size***

Table 1 presents the average absolute bias of parameter estimates obtained from the 96 simulation datasets with different settings, 32 simulation datasets for each sample size. These datasets assume the random effects to be normal and do not account for model misspecification. The average absolute bias for $\beta$ tend to decrease as sample size increases. Similar is true for $\sigma_a^2$, only that the reduction rate of bias was relatively slower, larger sample size is required to neutralize the bias for the parameter. Increasing the sample size has also reduced the biases for both values of ρ and for $\sigma_u^2 = 0.25$ with slightly higher gain when increasing from 50 to 150 than from 50 to 250. The estimate for $\sigma_u^2 = 4$ also showed gain in increasing the sample size with the least bias at $n = 250$. For $\sigma_v^2$, the effect of increasing the sample size has increased the bias with a sharp increase at $n = 250$. Overall, there was an improvement in the accuracy of estimates in increasing the sample size from 50 to 250 but not for $\sigma_v^2$.

**Table 1. Absolute Bias of Parameter Estimates Using Backfitting Method**

| Parameter/Value | | Sample Size | | | |
|---|---|---|---|---|---|
| | | 50 | 150 | 250 | all sizes |
| β | 0.2 | 0.059 | 0.043 | 0.007 | 0.036 |
| | 2 | 0.059 | 0.042 | 0.008 | 0.036 |
| ρ | 0.2 | 0.715 | 0.532 | 0.560 | 0.602 |
| | 0.9 | 1.389 | 1.107 | 1.299 | 1.265 |
| $\sigma_u^2$ | 0.25 | 3.930 | 2.393 | 2.890 | 3.071 |
| | 4 | 4.465 | 4.775 | 2.841 | 4.027 |
| $\sigma_v^2$ | 0.25 | 1.155 | 1.237 | 3.358 | 1.916 |
| | 4 | 1.104 | 1.356 | 3.471 | 1.977 |
| $\sigma_a^2$ | 0.1 | 0.072 | 0.063 | 0.059 | 0.065 |
| | 4 | 3.895 | 3.884 | 3.815 | 3.865 |
| No. of Scenarios | | 32 | 32 | 32 | 96 |

***Effect of model misspecification***

Recall that the random effect components of the spatiotemporal model assumes the distribution of the random effects as $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_i \sim N\ \ 0,\sigma_v^2$ . To check for robustness to model misspecification, simulation was conducted with the first-level and second-level random effects following an exponential distribution (distribution of the random effects are positively skewed). This was done in three different ways: 1) the distribution of the first-level random effects is exponential and the second-level is normal; 2) the distribution of the first-level random effects is normal while the second-level is exponential, and; 3) both levels of random effects are exponential. The parameter $\lambda$ of the exponential distribution was assumed to have true values 0.5 (long-tailed) and 2 (short-tailed), which yield variances of 4 and 0.25, respectively since $Var(X)=\frac{1}{\lambda}$ for $X \sim Expo\ (\lambda)$.

Model misspecification was first analyzed as to its effects to the bias of the estimates of the variances of the random effects as the model violation pertains to the model assumption of the random effects. Table 2 contains the average estimates and the corresponding average absolute bias for $\sigma_u^2$ with true values 0.25 and 4. Parameter estimates were obtained through backfitting iteration using 288 simulation datasets with and without model misspecification. Table 3 also contains the average estimates and average absolute bias for $\sigma_v^2$ with true values 0.25 and 4.

First-level random effects

Results of the simulation revealed that when there is a violation to model assumption of the first-level random effects, the overall average absolute bias for $\sigma_u^2$ was higher compared to the case where there is no violation to model assumption. It was slightly higher when only the model assumption for the first-level random effects was violated than when both levels of random effects were violated (Table 2).

While it was noted that model misspecification could increase the bias of the estimates for $\sigma_u^2$ obtained through backfitting iteration, the bias, however, could be reduced by increasing the sample size. At $n$ = 250, the biases for $\sigma_u^2 = 0.25$ or $\lambda = 2$ under the misspecified models were reduced to levels that were almost the same as the normal model. However, for $\sigma_u^2 = 4$ or $\lambda = 0.5$, the reduction rate of bias was slower and needed larger sample size in order to neutralize the bias caused by model misspecification.

**Table 2. Estimates and Absolute Bias for the Variance of the First-Level Random effects ($\sigma_u^2$) With and Without Model Misspecification**

| a. No Violation to Model Assumptions, $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_{it} \sim N\ \ 0,\sigma_v^2$ | | | | |
|---|---|---|---|---|
| Sample Size | $\sigma_u^2 = 0.25$ ($\lambda = 2$) | | $\sigma_u^2 = 4$ ($\lambda = 0.5$) | |
| | $\sigma_u^2$ | \|Bias\| | | \|Bias\| |
| 50 | 4.095 | 3.930 | 4.095 | 4.465 |
| 150 | 2.610 | 2.393 | 2.610 | 4.775 |
| 250 | 3.136 | 2.890 | 3.136 | 2.841 |
| All Sample Sizes | 3.280 | 3.071 | 3.280 | 4.027 |
| b. First-level Random Effects are Exponential, $u_i \sim$ Expo ($\lambda$) and $v_{it} \sim N\left(0,\sigma_v^2\right)$ | | | | |
| Sample Size | $\sigma_u^2 = 0.25$ ($\lambda = 2$) | | $\sigma_u^2 = 4$ ($\lambda = 0.5$) | |
| | $\hat{\sigma}_u^2$ | \|Bias\| | $\hat{\sigma}_u^2$ | \|Bias\| |
| 50 | 4.978 | 4.737 | 10.674 | 6.674 |
| 150 | 4.919 | 4.669 | 8.844 | 4.844 |
| 250 | 3.201 | 2.954 | 7.319 | 3.319 |
| All Sample Sizes | 4.366 | 4.120 | 8.945 | 4.945 |
| c. Both Levels of Random Effects are Exponential, $u_i \sim$ *Expo* ($\lambda$) and $v_{it} \sim$ *Expo* ($\lambda$) | | | | |
| Sample Size | $\sigma_u^2 = 0.25$ ($\lambda = 2$) | | $\sigma_u^2 = 4$ ($\lambda = 0.5$) | |
| | $\hat{\sigma}_u^2$ | \|Bias\| | $\hat{\sigma}_u^2$ | \|Bias\| |
| 50 | 4.120 | 3.936 | 5.405 | 4.465 |
| 150 | 5.326 | 5.076 | 9.486 | 5.486 |
| 250 | 3.269 | 3.019 | 7.518 | 3.518 |
| All Sample Sizes | 4.238 | 4.010 | 7.470 | 4.490 |

Second-level random fffects

There was no indication that the estimates for $\sigma_v^2$ was affected by the departure from normality of the second-level random effects. Overall, the measures of the average absolute bias for $\sigma_v^2$ were generally lower when there is model misspecification in the random effects. (Table 3).

**Table 3. Estimates and Absolute Bias for the Variance of the Second-Level Random effects ( $\sigma_v^2$ ) With and Without Model Misspecification**

| a. No Violation to Model Assumptions, $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_{it} \sim N\left(0,\sigma_v^2\right)$ | | | | |
|---|---|---|---|---|
| | $\sigma_v^2$ = 0.25 (λ = 2) | | $\sigma_v^2$ = 4 (λ = 0.5) | |
| | $\hat{\sigma}_v^2$ | \|Bias\| | $\hat{\sigma}_v^2$ | \|Bias\| |
| All Sample Sizes | 2.147 | 1.916 | 5.288 | 1.977 |
| b. Second-level Random Effects are Exponential, $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_{it}$~*Exp (λ)* | | | | |
| | $\sigma_v^2$ = 0.25 (λ = 2) | | $\sigma_v^2$ = 4 (λ = 0.5) | |
| | $\hat{\sigma}_v^2$ | \|Bias\| | $\hat{\sigma}_v^2$ | \|Bias\| |
| All Sample Sizes | 1.980 | 1.740 | 5.420 | 1.666 |
| c. Both Levels of Random Effects are Exponential, $u_i$~ Expo (λ) and $v_{it}$~ Expo (λ) | | | | |
| | $\sigma_v^2$ = 0.25 (λ = 2) | | $\sigma_v^2$ = 4 (λ = 0.5) | |
| | $\hat{\sigma}_v^2$ | \|Bias\| | $\hat{\sigma}_v^2$ | \|Bias\| |
| All Sample Sizes | 2.094 | 1.867 | 5.097 | 1.728 |

***Other model parameters***

The average absolute bias for the spatial and temporal effects and the variance of the white noise by estimator and sample size did not exhibit sensitivity to model misspecification as measures of bias for these parameters under the misspecified models did not have large discrepancy from the model with no violation to model assumption. With or without model misspecification, increasing the sample size can reduce the bias of the estimates for the parameters. The robustness of the spatial effect to model misspecification is also supported by the property of least squares estimators for the fixed effect parameters in an ordinary linear regression model. Least squares estimators are the Best Linear Unbiased Estimators (BLUE) for the regression coefficients for as long as the three Gauss-Markov conditions are satisfied which do not include assumption on normality. For this reason, the basic estimates for $\beta$, obtained through backfitting procedure, is considered robust from model misspecification.

### *5.5 Predictive ability of the model*

The predictive performance of the spatiotemporal mixed model was investigated with respect to changes in sample size, changes in the true value of the parameters and misspecification error.

***Effect of sample size***

Tables 4 presents the average MAPE by parameter and size obtained from the 96 simulation datasets. MAPE generally increases with increasing sample size, regardless of the assumed parameter values. This is a salient feature of the bootstrap method embedded into the backfitting algorithm that do not necessarily require a large sample size.

There was an upward trend in the average MAPE as the true values of the parameters were increased with the exception of $\beta$ whose average MAPE dropped when its true value was increased from 0.2 to 2, see Table 4 for details.

**Table 4. Mean Absolute Percentage Error (MAPE) at Varying Sample Sizes**

| Parameter/Value | | Sample Size | | | |
|---|---|---|---|---|---|
| | | 50 | 150 | 250 | all sizes |
| β | 0.2 | 1.643 | 1.868 | 2.442 | 1.985 |
| | 2 | 0.619 | 0.644 | 0.791 | 0.685 |
| $\rho$ | 0.2 | 1.096 | 1.225 | 1.478 | 1.266 |
| | 0.9 | 1.166 | 1.288 | 1.755 | 1.403 |
| $\sigma_u^2$ | 0.25 | 1.118 | 1.181 | 1.520 | 1.273 |
| | 4 | 1.144 | 1.332 | 1.712 | 1.396 |
| $\sigma_v^2$ | 0.25 | 1.041 | 1.222 | 1.522 | 1.262 |
| | 4 | 1.222 | 1.291 | 1.711 | 1.408 |
| $\sigma_a^2$ | 0.1 | 0.891 | 0.970 | 1.110 | 0.990 |
| | 4 | 1.372 | 1.543 | 2.123 | 1.679 |
| Overall | | 0.862 | 0.973 | 1.064 | 0.862 |
| No. of Scenarios | | 32 | 32 | 32 | 96 |

***Effect of model misspecification***

We present in Table 5 the average MAPE by size obtained using different model assumptions for the random effects. Violation to model assumptions for the random effects did not largely affect the predictive ability of the proposed model and the estimation procedure since obtained average MAPE were practically unchanged. This suggests that predictions are still reliable even for situations where the random effects deviate from normality assumption.

**Table 5. Mean Absolute Percentage Error (MAPE) With and Without Model Misspecification**

| Size | MAPE |
|---|---|
| a. No Violation to Model Assumptions, $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_{it} \sim N\left(0,\sigma_v^2\right)$ | |
| Size | MAPE |
| 50 | 0.860 |
| 150 | 0.973 |
| 250 | 1.064 |
| all sizes | 0.862 |
| b. First-level Random Effects are Exponential, $u_i \sim Expo\ (\lambda)$ and $v_{it} \sim N\left(0,\sigma_v^2\right)$ | |
| Size | MAPE |
| 50 | 1.000 |
| 150 | 0.973 |
| 250 | 0.997 |
| all sizes | 1.000 |
| c. Second-level Random Effects is Exponential, $u_i \sim N\left(0,\sigma_u^2\right)$ and $v_{it} \sim Expo\ (\lambda)$ | |
| Size | MAPE |
| 50 | 0.921 |
| 150 | 0.848 |
| 250 | 0.903 |
| all sizes | 0.921 |
| d. Both levels of random effects are exponential, $u_i \sim Expo\ (\lambda)$ and $v_{it} \sim Expo\ (\lambda)$ | |
| Size | MAPE |
| 50 | 0.960 |
| 150 | 0.847 |
| 250 | 0.863 |
| all sizes | 0.960 |

### 5.6 Comparison of the spatiotemporal mixed model with other models

The predictive ability of the proposed spatiotemporal model was compared with the predictive performance of the spatiotemporal model without the fixed effects in Landagan and Barrios (2006), nested linear mixed model (LM), non-nested LM and ordinary least squares (OLS). Comparison was done using the simulation datasets or scenarios with the lowest (best scenarios) and highest MAPE (worst scenarios) in the proposed spatiotemporal mixed model using the hybrid backfitting estimation procedure. Table 6 shows the estimates of MAPE for the different models.

The parameters for the Landagan and Barrios (2006) spatio-temporal model contain the parameters in the proposed spatiotemporal mixed model (excluding the random effects). For OLS, the parameters are the constant α and spatial effect $\beta$. The two linear mixed models contain the model parameters in the proposed spatio-temporal mixed model, except for the temporal effect $\rho$.

Consistently, the OLS had the poorest performance as expected because of its inability to capture the random effects and temporal effect inherent in the data. Between the two linear mixed models, the nested random effects yielded better prediction than the non-nested random effects indicating that in a nested panel data, the nesting of the random effects should not be ignored in the model. The Landagan and Barrios (2006) spatio-temporal model had almost the same prediction performance with the OLS and was outperformed by the two linear mixed models. This indicates that spatial and temporal effects are not sufficient to model the characteristics of the data generated from a rotated panel survey with nested effects.

The spatiotemporal mixed model is superior over the four models considered, both for the best and worst scenarios. Having the best prediction performance even in the worst scenarios, the proposed spatiotemporal mixed model will likely have the best predictive ability in the rest of the scenarios used in this study.

**Table 6. Comparison of Mean Absolute Percentage Error (MAPE) of the Proposed Model with Other Models**

| Case No. | Sample Size | Spatio-Temporal Model | | Linear Mixed non-nested | Linear Mixed nested | OLS |
|---|---|---|---|---|---|---|
| | | Proposed | Landagan & Barrios (2006) | | | |
| A. Scenarios with the Lowest MAPE for the Proposed Model (best) | | | | | | |
| 12 | 50 | 0.241 | 2.356 | 1.950 | 0.256 | 2.296 |
| 34 | 150 | 0.246 | 2.242 | 1.739 | 0.249 | 2.229 |
| 66 | 250 | 0.276 | 2.595 | 1.872 | 0.281 | 2.633 |
| B. Scenarios with the Highest MAPE for the Proposed Model (worst) | | | | | | |
| 17 | 50 | 2.758 | 28.532 | 31.27 | 2.975 | 32.356 |
| 49 | 150 | 3.418 | 33.305 | 29.134 | 3.438 | 33.146 |
| 93 | 250 | 4.342 | 45.653 | 31.943 | 4.508 | 42.528 |

### *5.7 Small area estimation*

The proposed spatio-temporal mixed model is used in small area estimation of statistics that are continuous in nature such as ratio, rates and averages where the sample data are generated from a rotated panel survey with nested effects.

We present first a brief description of the rotated panel survey that is simulated, the spatiotemporal model in relation to actual data, the pseudo population, the sample selection, and the simulation results.

***The simulated rotated panel survey***

The panel survey simulated in this study is one that is repeated quarterly and adopts a sampling design with two levels of selection, the units in the second level being rotated and nested within the first level. An example of this is the selection of sample barangays or PSUs in the domain of interest for the first level and the selection of households or group of households in the second level within the sampled barangays or PSUs. The sample units in the first level are fixed over time while the second-level sample units are rotated quarterly with no overlapping second-level sample units in the survey rounds which means that a second-level sample at time t will no longer be at time $t + 1$. Hence, a rotation scheme with overlapping second-level sample units in the survey rounds is not considered in this study. The rotation of the second-level sample units are incorporated in the assignment of identification code for each of the second-level units.

The assumption of a balanced data where the number of responding second-level sample units is the same in each survey rounds is quite difficult to fulfill when the second-level units refer to individuals (say, households or persons) since there is a higher possibility of nonresponse at this level. Because of the missing responses for the sample units in some survey rounds, the dataset becomes unbalanced. In some surveys, instead of selecting samples at the individual level, cluster of individuals or rotation groups are selected which is the scenario that is being simulated in this study. With clusters or rotation groups as samples, nonresponse can occur if all the individuals in the sample cluster or rotation group have no responses, e.g., when there is flooding, demolition of informal settlers, peace and order problem and other unforeseen circumstances in the area that made it isolated and impenetrable. Also, simulated in this undertaking is a design where the number of sample clusters or rotation groups is one for each first-level sample unit.

***The pseudo population***

A pseudo population serves as the sampling frame in selecting samples for small area estimation. It contains 250 first-level units in each time point or $N = 250$. Instead of generating separate pseudo population for the simulation of the small area estimation procedure, the pseudo population were selected from among the 96 simulation datasets. These were the scenarios with the lowest and highest MAPE and with $N = 250$. These scenarios are scenario 66 and 93 whose settings and MAPE are presented in Table 7. The parameter values in Scenario 66 were also found to have the lowest average absolute bias and those of scenario 93 were the ones with the highest average bias, except for the temporal component where $\rho = 0.9$ had the highest average absolute bias.

**Table 7. Settings and Mean Absolute Percentage Error (MAPE) of the Best and Worst Scenarios**

| Parameter | Scenario 66 (Best) | Scenario 93 (Worst) |
|---|---|---|
| $\alpha$ | 4 | 11 |
| $\beta$ | 2 | 0.2 |
| $\rho$ | 0.2 | 0.2 |
| $\sigma_u^2$ | 0.25 | 4 |
| $\sigma_v^2$ | 0.25 | 4 |
| $\sigma_a^2$ | 0.1 | 4 |
| MAPE | 0.276 | 4.342 |

***Sample selection***

To simplify the selection process, as this is not the main concern in this undertaking, but rather on small area estimation involving data from panel survey, the sample selection was done in the first level only. In the pseudo population, there is only one second-level unit in every first-level unit and at each time point. Hence, this second-level unit was automatically the sample for the first-level unit. There were three sample sizes $n_d$ used: 20, 50 and 100. The selection was systematic sampling with an interval of $k = \frac{N_d}{n_d}$. This means that there are $k$ possible groups of samples from the population and the selected group depends on the random start. This $k$ is referred to as the number of rotation groups in the Philippine Labor Force Survey. In this paper, the proposed small area estimation procedures were applied and evaluated using all possible groups of samples from each of the two pseudo population. Prior to sample selection, the spatial variable used in the simulation was sorted to ensure that samples were more or less distributed equally in terms of municipality per capita income.

***Small area estimates***

The small area estimation was done following the procedures described in Section 4. There are two proposed small area estimation procedures, both methods adopt the bootstrap technique but in different ways. Method 1 applied the bootstrap technique in estimating the model parameters while Method 2 applied the bootstrap procedure directly to the estimation of the unknown sum $Y_{dt,n}$. Method 1 was implemented using Algorithm (3), and Method 2 using Algorithms (4).

One interesting property of bootstrap estimators is their being asymptotically normal. Guarte (2009) has shown the asymptotic normality of the bootstrap estimators of the spatial and temporal parameters the spatiotemporal model.

Figure 1 shows the histogram of the bootstrap estimates for varying sample sizes. The settings are the same as in Scenario 66 in Table 7. The bootstrap estimates for the spatial effect $\beta$ exhibited normality even at small sample size. The estimates for the temporal effect $\rho$ and the first-level random effects $\sigma_u^2$ were not normal at small sample size but they tend to normality as sample size increased. This does not follow for the estimates of the variance of the random effects $\sigma_v^2$ . The distribution of the bootstrap estimates for the parameter departed from normality when sample size was increased. For instance, at $n = 50$, it can be seen that the estimates for follow a normal distribution. Formal test using skewness and kurtosis tests ($p < 0.970$) also showed normality of the estimates. However, at $n = 150$, the estimates started to depart from normality as shown in the formal test and even farther from normality at $n = 250$ which is evident in the histogram and normal density plot in Figure 1. The variance of the error term $\sigma_a^2$ did not exhibit normality in any of the sample sizes used. A larger sample and/or larger number of bootstrap replicates might be needed to normalize the estimates for the variance of the error term.

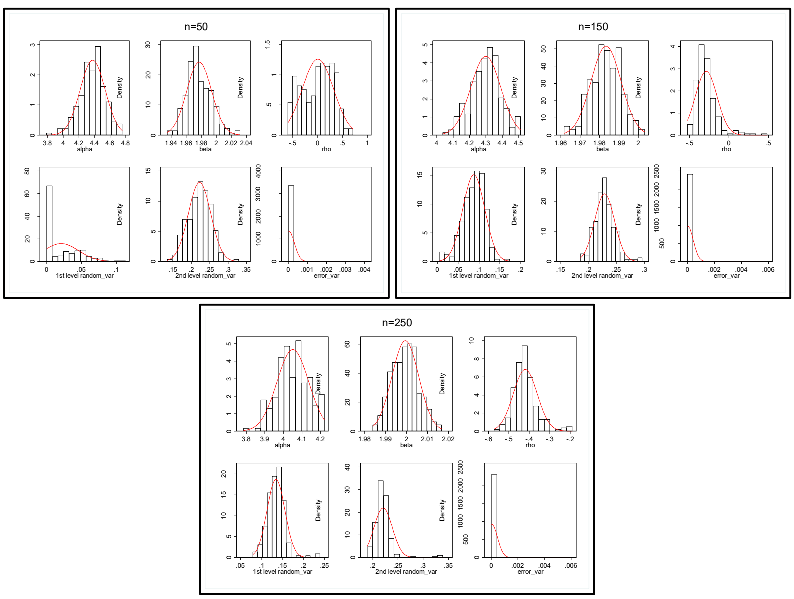


**Figure 1. Histogram and Normal Density Plot of Bootstrap Estimates Based on 200 Bootstrap Replicates at Varying Sizes**

**($\beta = 2, \rho = 0.2, \sigma_u^2 = 0.25, \sigma_v^2 = 0.25$,and $\sigma_a^2 = 0.1$ )**

Tables 8 and 9 show the estimates of and the absolute bias and measures of precision using two pseudo populations. These pseudo populations were the best and worst scenarios based on MAPE as shown in Table 8. For the best scenario (Scenario 66), the true value of the parameter of interest at time $t = 4$ was $\overline{Y}_{dt}$ = 26.942. The estimates for this parameter using different possible systematic samples of different sizes did not vary. The lowest for Method 1 was 26.837 and the highest was 27.192. For Method 2, the lowest was 26.805 while the highest was 27.242. The range of possible estimates was wider for larger sample size. Between the two methods, Method 1 had slightly lower estimates, on the average.

In the worst scenario (Scenario 93) where $\overline{Y}_{dt}$ =13.524 at $t = 4$, the possible estimates for the parameter did not also differ largely where values ranged from 12.921 to 14.555 for Method 1 and 12.880 to 14.714 for Method 2. As in the best scenario, Method 1 also obtained slightly lower estimate, on the average, for all sample sizes.

The measures of precision of the small area estimates were evaluated through their SEs and CVs. These measures were analyzed on how they were affected by the pseudo population, sample selection and sample size.

In both scenarios, the SEs and CVs were lower for Method 1 than for Method 2 and were also lower for the best scenario than for the worst scenario.

For both methods, the SEs and CVs were stable with the change of the sets of samples within the sample size. For as long as the selection of samples is not biased with respect to the spatial variable used in the model, the estimates are stable which means that any set of samples used in the estimation will produce estimate with the same precision as the other possible sets of samples with the same sample size. Note that in the sample selection, the pseudo population was sorted by value of the spatial variable before systematic sample selection to implicitly stratify the units by the spatial variable.

Across sample size, Method 1 showed stability on the reliability of estimates indicating that large sample size is not needed to achieve reliable estimates using this method unlike the design-based. Precision of estimates for Method 2 can be improved by increasing the sample size but even for small sample size drawn from a population that poorly fits into the proposed model, the small area estimation procedure still yield reliable estimates. The CVs for the worst scenario ranged from 2.820 percent to 4.615 percent for $n = 20$.

Average bias for $\overline{Y}_{dt}$ at $t = 4$ was also low for both methods, with relatively lower average absolute bias for Method 1 in all sample sizes and in both scenarios. As expected, the worst scenario produced higher absolute bias than the best scenario. On the average, both methods and scenarios were slightly decreasing as sample size decreased. This can be justified by the consistency of bootstrap estimators as supported by the known theories and/or empirical investigations.

**Table 8. Small Area Estimates, Measures of Precision, and Bias by Random Start and Method: Population – Best Scenario (Scenario 66) ($\overline{Y}_{dt}$ =26.942, $t$ = 4)**

| $n$ = 20 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 26.883 | 26.869 | 0.059 | 0.073 | 0.003 | 0.012 | 0.010 | 0.044 |
| 2 | 27.091 | 27.160 | 0.149 | 0.218 | 0.002 | 0.143 | 0.009 | 0.528 |
| 3 | 27.018 | 26.971 | 0.076 | 0.029 | 0.002 | 0.115 | 0.009 | 0.425 |
| 4 | 26.852 | 26.805 | 0.090 | 0.137 | 0.002 | 0.121 | 0.008 | 0.452 |
| 5 | 27.127 | 27.038 | 0.185 | 0.096 | 0.003 | 0.149 | 0.010 | 0.550 |
| 6 | 26.837 | 26.883 | 0.105 | 0.059 | 0.003 | 0.140 | 0.010 | 0.520 |
| 7 | 27.062 | 27.242 | 0.119 | 0.300 | 0.002 | 0.129 | 0.008 | 0.473 |
| 8 | 27.192 | 27.179 | 0.250 | 0.237 | 0.002 | 0.124 | 0.009 | 0.456 |
| 9 | 26.924 | 27.051 | 0.018 | 0.108 | 0.002 | 0.091 | 0.008 | 0.335 |
| 10 | 27.001 | 27.046 | 0.059 | 0.104 | 0.002 | 0.095 | 0.008 | 0.350 |
| 11 | 26.965 | 26.904 | 0.023 | 0.038 | 0.002 | 0.121 | 0.009 | 0.449 |
| 12 | 27.035 | 27.150 | 0.093 | 0.208 | 0.002 | 0.100 | 0.008 | 0.369 |
| Average | 27.009 | 27.025 | 0.112 | 0.134 | 0.003 | 0.117 | 0.012 | 0.431 |
| $n$ = 50 | | | | | | | | |
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 26.986 | 26.982 | 0.044 | 0.040 | 0.002 | 0.054 | 0.008 | 0.201 |
| 2 | 27.026 | 27.135 | 0.084 | 0.193 | 0.002 | 0.066 | 0.009 | 0.242 |
| 3 | 26.999 | 26.975 | 0.057 | 0.033 | 0.002 | 0.057 | 0.008 | 0.210 |
| 4 | 27.132 | 27.152 | 0.190 | 0.210 | 0.003 | 0.076 | 0.009 | 0.281 |
| 5 | 26.916 | 26.916 | 0.026 | 0.026 | 0.002 | 0.072 | 0.009 | 0.267 |
| Average | 27.012 | 27.032 | 0.080 | 0.100 | 0.002 | 0.065 | 0.009 | 0.240 |
| $n$ = 100 | | | | | | | | |
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 26.958 | 26.959 | 0.016 | 0.016 | 0.002 | 0.040 | 0.008 | 0.150 |
| 2 | 27.007 | 27.091 | 0.065 | 0.148 | 0.002 | 0.049 | 0.009 | 0.180 |
| Average | 26.983 | 27.025 | 0.040 | 0.082 | 0.002 | 0.045 | 0.009 | 0.165 |

**Table 9. Small Area Estimates, Measures of Precision, and Bias by Random Start and Method: Population – Worst Scenario (Scenario 93)**
**($\overline{Y}_{dt}$ =13.524, $t$ = 4)**

| $n$ = 20 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 13.248 | 13.230 | 0.276 | 0.295 | 0.011 | 0.430 | 0.080 | 3.247 |
| 2 | 14.089 | 14.474 | 0.565 | 0.949 | 0.011 | 0.619 | 0.078 | 4.276 |
| 3 | 13.809 | 13.744 | 0.285 | 0.219 | 0.011 | 0.485 | 0.077 | 3.531 |
| 4 | 13.060 | 12.880 | 0.464 | 0.645 | 0.010 | 0.553 | 0.078 | 4.297 |
| 5 | 14.377 | 13.974 | 0.853 | 0.450 | 0.012 | 0.645 | 0.087 | 4.615 |
| 6 | 12.921 | 13.146 | 0.604 | 0.378 | 0.012 | 0.596 | 0.094 | 4.531 |
| 7 | 14.074 | 14.714 | 0.549 | 1.189 | 0.009 | 0.515 | 0.067 | 3.499 |
| 8 | 14.555 | 14.508 | 1.031 | 0.984 | 0.010 | 0.484 | 0.072 | 3.339 |
| 9 | 13.469 | 13.944 | 0.055 | 0.420 | 0.010 | 0.367 | 0.073 | 2.630 |
| 10 | 13.786 | 13.976 | 0.262 | 0.452 | 0.010 | 0.427 | 0.072 | 3.059 |
| 11 | 13.674 | 13.346 | 0.150 | 0.178 | 0.011 | 0.494 | 0.079 | 3.703 |
| 12 | 13.839 | 14.461 | 0.315 | 0.937 | 0.010 | 0.408 | 0.069 | 2.820 |
| Average | 13.742 | 13.866 | 0.451 | 0.591 | 0.011 | 0.502 | 0.077 | 3.629 |
| $n$ = 50 | | | | | | | | |
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 13.722 | 13.656 | 0.197 | 0.131 | 0.010 | 0.241 | 0.074 | 1.767 |
| 2 | 13.842 | 14.420 | 0.317 | 0.895 | 0.011 | 0.286 | 0.081 | 1.986 |
| 3 | 13.730 | 13.610 | 0.206 | 0.085 | 0.010 | 0.238 | 0.072 | 1.747 |
| 4 | 14.350 | 14.461 | 0.826 | 0.936 | 0.011 | 0.319 | 0.077 | 2.204 |
| 5 | 13.410 | 13.382 | 0.115 | 0.143 | 0.011 | 0.305 | 0.081 | 2.279 |
| Average | 13.811 | 13.906 | 0.332 | 0.438 | 0.011 | 0.278 | 0.077 | 1.997 |
| $n$ = 100 | | | | | | | | |
| Random start | $\hat{\hat{Y}}_d$ | | \|Bias\| | | SE | | CV (%) | |
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 | Method 2 |
| 1 | 13.732 | 13.762 | 0.208 | 0.238 | 0.011 | 0.139 | 0.077 | 1.008 |
| 2 | 13.828 | 13.844 | 0.303 | 0.319 | 0.010 | 0.130 | 0.072 | 0.938 |
| Average | 13.780 | 13.803 | 0.256 | 0.279 | 0.010 | 0.134 | 0.075 | 0.973 |

## *5.8 Advantage of the proposed model-based method over the design-based method*

The large sampling variability of design-based estimators when sample size is small makes them unreliable as bases for policy formulation and program development in small areas. Increasing the sample size can increase the precision of estimates, however, the drawback is the higher budgetary requirement for data generation. Because of limited resources, often, surveys are designed to provide reliable estimates at higher geographic level.

The proposed model-based methods provide a solution to the issues confronting the design-based estimators. With the proposed model-based estimation procedures, estimates are precise even with small samples while prediction error is low. Note that variance computation for Method 1 is independent of the sample size, hence, a large sample size is not required to obtain precise estimates for as long as the estimates for the variance components and the error variance are low. The variance computation using bootstrap procedure, which was used for Method 2, is also independent of sample size. This procedure is also known to reduce sampling variability. The design-based variance estimators are more complicated than the proposed variance estimators, especially for surveys with complex design.

Since the small area estimation procedure works as an imputation technique for non-sampled units in the small area, then errors in imputing the values for the non-sampled units are also low, resulting to accurate small area estimate for the parameter of interest.

## 6. Illustration: Estimation of Unemployment Rate

The proposed model-based small area estimation procedures presented in Section 4 are illustrated using actual data on unemployment rates from the Labor Force Survey (LFS). The small area estimates are for the unemployment rates of the cities/municipality of the National Capital Region (NCR) for the fourth quarter of 2008. The LFS is designed to generate reliable estimates on labor and employment with region as domain starting 2006. Despite region as the domain of the survey, estimates of employment rate at the city/municipality level of NCR are still reliable based on their coefficient of variations (CVs) which are all below 10 percent. In the case of unemployment rate, precision is very low at the city/municipality level of NCR and even for majority of the provinces because of insufficient sample size and problem on attrition. The limitation of design-based estimation to generate reliable estimates for areas lower than the domain is expected to be resolved with the proposed small area estimation procedures.

### *6.1 Brief description of the sampling design of the 2008 LFS*

The 2008 LFS and other household-based surveys of the PSA share the same sample areas to minimize cost in conducting the surveys. These sample areas are contained in the 2003 Master Sample which was designed to generate reliable statistics at the regional level. Hence, sample size cannot assure reliability of statistics below regional level. The 2003 Master Sample has a two-stage stratified design with the selection of Primary Sampling Units (PSUs) in the first stage and EA for the second stage. Only one EA is selected for each sample PSU and both

PSUs and EAs are selected with probability proportional to size where measure of size is the number of households in the stratum. A PSU refers to a barangay or a group of adjacent and contiguous barangays that, together, have at least 500 households. An EA refers to a barangay or a portion of a barangay with about 350 households. The same sample PSUs are used quarterly in the LFS but the sample EAs may change depending on whether the EA is exhausted or not. An EA is exhausted if all the housing units in the EA have been used as samples in the LFS.

Sample housing units are not included in the Master Sample. They are being selected quarterly within the sample EA before the start of the survey round. This constitutes the third stage of selection for the LFS. Sample housing units are selected systematically within the sample EA and at most three households are interviewed in the sample housing units. With systematic sampling, *k* groups of sample housing units can be formed where *k* is the sampling interval. These groups are referred to as the rotation groups in the LFS since they are being rotated quarterly. In each quarter, one rotation group is selected for every sample EA.

### 6.2 Fitting of unemployment data in the spatio-temporal mixed model

The spatiotemporal mixed model is further described particularly on how the unemployment data are fitted in the model. As an initial step in small area estimation of unemployment rate for each city/municipality of the NCR, the computed unemployment rates for the sample PSUs at time t are fitted in the spatiotemporal model as:

$$unemp\ rate_{it} = \alpha + \beta income_i + u_i + v_{it} + \varepsilon_{it}, \qquad (18)$$

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + a_{it}$$

where *unemp* refers to unemployment rate, *income* refers to city/municipality per capita income, $i = 1, 2, \ldots, n$ PSUs and $t = 1, 2, 3, 4$ represents the quarters.

Since the PSUs and rotation groups (groups of sample housing units) are samples from their respective population, they are treated as random effects to model the between-PSU and between-rotation group variability, respectively. In the spatio-temporal mixed model, the PSU random effects are denoted by $u_i$'s which are the same for all time points to reflect the design of the LFS that have the same sample PSUs over time. The rotation group random effects are denoted by $v_{it}$'s in which the rotation quarterly is indicated in the dataset by assigning a unique identification code for the sample rotation groups within the sample PSU. Note that there is only one sample rotation group for the sample PSU. So, in sample PSU 1, the random effect of the sample rotation group at time 1 is indicated as 1, rotation group 2 at time 2 as 2 and so on. For convenience, the numeric codes 1; 2; 3 and 4, similar to the codes for the four time points, were used to uniquely identify the rotation groups for the sample PSUs. This is the reason why the second-level random effects are denoted in the model as $v_{it}$ instead of $v_{ijt}$. Other codes can be used for the rotation groups for as long as they indicate uniqueness of the quarterly sample rotation groups within the PSU to describe the quarterly rotation of sample housing units in the PSU.

The quarterly collection of unemployment data in the LFS introduces correlation of the residuals over time, hence, they are modeled as an autoregressive process (AR). For a short time series with only four time points, an AR(1) was deemed reasonable to model the temporal component.

### 6.3 Estimates of unemployment rate using the proposed spatio-temporal mixed model

The target data for estimation is the unemployment rate for the fourth quarter of 2008 using the LFS results for the four quarters of 2008. Though Method 1 was more superior as to results and computational concerns based on the simulation results in Chapter 4, Method 2 was also applied in the actual data estimation to assess its performance in comparison with the design-based method. The backfitting and bootstrap techniques used are the same algorithms used in the simulations which are presented in Algorithms 1 and 2, respectively, in Chapter 3. Algorithm 3 was applied for small area estimation procedure using Method 1 and Algorithm 4 for Method 2 which are also in Chapter 3.

In the estimation of the parameters of the spatio-temporal model, the pooled observations for the whole NCR were utilized. The estimation of parameters cannot be done by city/municipality because of the constant value of the city/municipality per capita income in all PSUs in the city/municipality. This will create collinearity issue when estimating the spatial parameter. In regions outside the NCR, small area estimation for the cities/municipalities can be done by pooling the observations in the province instead of the region.

In applying the backfitting procedure using Algorithm 1, the residuals, resulting from the elimination of the spatial and random effects from the realizations of the response variable, were tested for stationarity before proceeding to the computation of the initial values of $\rho$ and $\sigma_a^2$ in Step 3 of the algorithm. The test used was the Harris-Tsavalis (HT), one of the panel data unit root tests available in STATA for balanced data. The HT unit root test assumes that all panels (the i's or PSUs) share the same autoregressive parameter ρ which is the assumption for the proposed spatiotemporal mixed model. The test is based on the regression model $y_{it} = \rho y_{i,t-1} + Z'_{it}\gamma_i + \epsilon_{it}$ where $Z'_{it}\gamma_i$ refers to panel-specific mean and trend. The tested null hypothesis is Ho: $\rho = 1$ (the panels contain unit root) versus the alternative Ha: $\rho \neq 1$ (panels are stationary). The test was implemented with no constant (no panel-specific mean and trend). Results suggest for stationarity of the residuals with computed $z$ statistcs of -43.54 and p-value of 0.000). Before estimating ρ from the AR(1) process, the residuals $\varepsilon_{it}$'s were centered in order to have a mean of zero which is the assumption for the $\varepsilon_{it}$'s.

The Landagan and Barrios (2006) spatiotemporal model was compared with the proposed model in terms of reliability in estimating unemployment rates for the NCR. This model, without the fixed effects, was fitted to the unemployment rates of the sample PSUs in NCR where fitted equation was used to impute the employment rates for the non-sampled PSUs for the region. The estimation procedure used for the Landagan and Barrios (2006) spatiotemporal model was backfitting embedded with Cochranne–Orcutt procedure.

The 2008 LFS for the NCR had a total of 4,415 PSUs with a sample size of 559 (13%) quarterly, bulk of which was in Quezon City (137 PSUs quarterly), followed by Caloocan city (68 PSUs quarterly) and Manila (63 PSUs quarterly). The lowest sample size was in San Juan and Pateros, each with 6 PSUs quarterly. On the average, the number of sample households was eight per PSU. (Table 10)

**Table 10. Sample Size of the LFS by City/Municipality of the NCR**

| City/ Municipality | Total No. of PSUs | No. of Sample PSUs | Total No. of Households | No. of Sample Households |
|---|---|---|---|---|
| | N | n | N' | n' |
| Manila | 698 | 63 | 326,869 | 683 |
| Quezon City | 982 | 137 | 459,989 | 1,033 |
| Mandaluyong | 124 | 15 | 57,871 | 127 |
| Marikina | 163 | 19 | 76,272 | 153 |
| Pasig | 231 | 34 | 107,960 | 237 |
| San Juan | 50 | 6 | 23,422 | 60 |
| Caloocan | 518 | 68 | 242,436 | 489 |
| Malabon | 155 | 20 | 72,607 | 139 |
| Navotas | 103 | 16 | 48,085 | 93 |
| Valenzuela | 225 | 34 | 105,444 | 259 |
| Makati | 215 | 37 | 100,678 | 178 |
| Pateros | 26 | 6 | 12,098 | 50 |
| Taguig | 215 | 20 | 100,756 | 196 |
| Las Pinas | 197 | 31 | 92,203 | 225 |
| Muntinlupa | 159 | 15 | 74,235 | 235 |
| Paranaque | 198 | 22 | 92,589 | 265 |
| Pasay | 156 | 16 | 72,878 | 197 |
| Total | 4,415 | 559 | 2,066,392 | 4,619 |

The results of the estimation were compared with the design-based approach and the Landagan and Barrios (2006) spatio-temporal model, without the fixed effects. Presented in Table 11 are the small area estimates using the model-based and design-based methods. With the proposed model-based approach, 8 out of the 17 cities/municipality of NCR had higher estimates of unemployment rate than the design-based. These are a mixture of large and small areas, namely, Quezon

City, Manila, Pasig, San Juan, Navotas, Valenzuela, Pateros and Taguig. The small area estimates using the proposed model were consistent with the design-based method in terms of the pattern of estimates across areas. This means that cities/municipality with high/low estimates of unemployment rates in the design-based were also the cities/municipalities with high/low estimates in the proposed model. In Landagan and Barrios (2006) spatiotemporal model, estimates of unemployment rate for the cities/municipality were nearly the same for most areas especially for Method 1.

The weighted means for the competing models were obtained with weights equal to the ratio of the total number of household (N') to the sample households (n') in 2008. The weighted mean of the estimates of the proposed model-based approach using Method 1 was equal to the average of the design-based estimates. Method 2 of the proposed model had a higher weighted mean than the design-based method. Both methods under the Landagan and Barrios (2006) model had weighted means lower than the design-based approach.

**Table 11. Small Area Estimates of Unemployment Rate (%) for the Cities/ Municipality of the National Capital Region: Fourth Quarter, 2008**

| City/ Municipality | Weight (N'/n') | Model-Based | | | | Design-Based |
|---|---|---|---|---|---|---|
| | | Proposed Model | | Landagan & Barrios Model (2006) | | |
| | | Method 1 | Method 2 | Method 1 | Method 2 | |
| Manila | 683 | 13.2 | 13.3 | 12.4 | 12.4 | 12.7 |
| Quezon City | 1,033 | 10.1 | 10.2 | 12.0 | 12.0 | 9.9 |
| Mandaluyong | 127 | 14.5 | 15.2 | 12.5 | 12.5 | 16.9 |
| Marikina | 153 | 14.7 | 15.2 | 12.6 | 15.2 | 16.2 |
| Pasig | 237 | 10.1 | 10.6 | 12.0 | 12.0 | 9.2 |
| San Juan | 60 | 12.3 | 12.7 | 12.7 | 12.7 | 8.8 |
| Caloocan | 489 | 15.8 | 15.9 | 12.8 | 12.9 | 17.9 |
| Malabon | 139 | 13.3 | 13.8 | 12.4 | 12.4 | 13.9 |
| Navotas | 93 | 22.8 | 23.5 | 13.9 | 13.9 | 22.3 |
| Valenzuela | 259 | 15.3 | 15.7 | 12.7 | 12.7 | 14.3 |
| Makati | 178 | 16.0 | 16.5 | 12.8 | 16.5 | 17.1 |
| Pateros | 50 | 14.9 | 15.0 | 13.8 | 15.0 | 13.3 |
| Taguig | 196 | 12.4 | 12.9 | 12.3 | 12.9 | 11.9 |
| Las Pinas | 225 | 10.7 | 11.1 | 12.0 | 12.0 | 12.4 |
| Muntinlupa | 235 | 14.8 | 15.4 | 12.5 | 15.4 | 16.0 |
| Paranaque | 265 | 7.3 | 7.8 | 11.7 | 11.7 | 7.9 |
| Pasay | 197 | 7.6 | 8.2 | 11.8 | 11.8 | 9.3 |
| Total | 7,433 | | | | | |
| Simple Mean | | 13.3 | 13.7 | 12.5 | 13.2 | 13.5 |
| Weighted Mean | 13.5 | 13.9 | 12.5 | 13.2 | | |

*Source of Basic Data: 2008 Labor Force Survey Datafile, Philippine Statistics Authority*

We present in Table 12 the CVs for the estimates of unemployment rates for the cities/municipality of the NCR. For the design-based estimate of unemployment rate, only Quezon City and Caloocan City had CVs below 10 percent. These two cities had the largest sample size. The highest CV was noted in San Juan (57.68 percent), one of the two cities with the lowest sample size. CVs of all the cities/municipality were relatively lower for the proposed spatiotemporal mixed model, both in Method 1 and Method 2. The Landagan and Barrios (2006) model has also obtained improved estimates of unemployment rates with CVs that were far lower than the design-based approach.

For both model-based procedures, Method 2 yielded better estimates than Method 1, contrary to the findings during the simulation where Method 1 was superior over Method 2 in terms of precision. For small areas such as San Juan and Pateros with only 50 and 26 total PSUs, respectively, the unemployment rates were more precise using Method 2. This is because the bootstrap estimator, which was used for Method 2 as variance computation, can reduce sampling variability. It outperforms Method 1 when the estimates for variance components of the random effects and error variance are large. Below are the bootstrap estimates for the parameters using Model 1 for the two model-based procedures:

*Proposed Spatio-temporal Mixed Model Estimates:*

$$\hat{\alpha} = 12.094, \hat{\beta} = -0.004, \rho = 0.244, \hat{\sigma}_u^2 = 0.427, \hat{\sigma}_v^2 = 99.415, \hat{\sigma}_a^2 = 0.008$$

*Landagan and Barrios (2006) Spatio-Temporal Model Estimates:*

$$\hat{\alpha} = 12.331, \hat{\beta} = -0.007, \hat{\rho} = 0.124, \hat{\sigma}_a^2 = 149.660$$

Note that $\hat{\sigma}_a^2$ is large without the random components which is the case in Landagan and Barrios (2006). This made the SEs of the estimates of unemployment rates for the Landagan and Barrios (2006) model larger than that of the proposed model under Method 1.

With these results, model-based approach has further proven its worth as a solution to the existing limitations of the design-based approach for small area estimation.

**Table 12. Coefficient of Variations (%) of Unemployment Rates Using the Model-Based and Design-Based Methods**

| City/ Municipality | Model-Based | | | | Design-Based |
|---|---|---|---|---|---|
| | Proposed Model | | Landagan & Barrios Model (2006) | | |
| | Method 1 | Method 2 | Method 1 | Method 2 | |
| Manila | 2.87 | 1.72 | 3.78 | 2.42 | 10.66 |
| Quezon City | 3.17 | 0.21 | 3.28 | 2.38 | 8.40 |
| Mandaluyong | 6.21 | 0.89 | 8.87 | 2.31 | 20.97 |
| Marikina | 5.34 | 0.66 | 7.68 | 0.68 | 19.33 |
| Pasig | 6.48 | 0.77 | 6.79 | 2.37 | 16.50 |
| San Juan | 11.53 | 0.64 | 13.74 | 2.45 | 57.68 |
| Caloocan | 2.79 | 0.15 | 4.22 | 2.26 | 8.16 |
| Malabon | 6.04 | 0.76 | 7.99 | 2.33 | 12.88 |
| Navotas | 4.32 | 0.60 | 8.74 | 2.03 | 15.53 |
| Valenzuela | 4.37 | 0.54 | 6.46 | 2.23 | 13.61 |
| Makati | 4.26 | 0.57 | 6.56 | 0.61 | 10.75 |
| Pateros | 13.17 | 0.43 | 17.59 | 0.24 | 21.78 |
| Taguig | 5.50 | 0.70 | 6.82 | 0.74 | 13.44 |
| Las Pinas | 6.64 | 0.73 | 7.32 | 2.33 | 18.54 |
| Muntinlupa | 5.35 | 0.67 | 7.80 | 0.70 | 15.30 |
| Paranaque | 9.75 | 1.19 | 7.47 | 2.52 | 23.87 |
| Pasay | 10.53 | 1.34 | 8.35 | 2.51 | 18.76 |

*Source of Basic Data:2008 Labor Force Survey, Philippine Statistics Authority*

## 7. Concluding Notes

The postulated spatiotemporal mixed model has the capability to correctly describe the structure inherent in a rotating panel survey with nested random effects. The backfitting algorithm has simplified the estimation of model parameters while the bootstrap has reduced the bias in the estimates of the parameters.

The proposed model-based estimation procedures are capable of generating reliable small area statistics using sample data generated from a rotating panel survey. The procedures have performed well even at the boundary conditions. Application of the procedures to actual data on unemployment rates in the cities and municipality of the NCR has confirmed advantages over the design-based approach in terms of precision.

## REFERENCES


ANSLEY, C.F., KOHN, R., 1992, Convergence of the Backfitting Algorithm for Additive Models, *Journal of Australian Mathematical Society* 57: 316-329.

BRAKEL, J. and KRIEG, S., Modelling Rotation Group Bias and Survey Errors in the Dutch Labour Force Survey, Section on Survey Research Methods, pp. 2675-2682

BUJA, A., HASTIE, T., TIBSHIRANI, R., 1989, Linear Smoothers and Additive Models, *The Annals of Statistics*, 17(2): 453-510

COELHO, P.S, PEREIRA, L.N., 2011, A Spatial Unit Level Model for Small Area Estimation, *Revstat – Statistical Journal*, 9(2), 155-180.

DUMANJUG, C.F., BARRIOS, E.B., and LANSANGAN, J.R.G., 2010, Bootstrap Procedures in a Spatial-temporal Model, *Journal of Statistical Computation and Simulation*, 80(7), 809-822.

EFRON, B., TIBSHIRANI, R., 1986, Bootstrap Methods for Standard Errors and other Sources of Statistical Accuracy, *Statistical Science*, 1: 54-75.

GUARTE, J. M. ,2009, Nonparameteric Bootstrap Inference in a Spatial-Temporal Model, Unpublished Ph. D. Dissertation, School of Statistics, University of the Philippines Diliman.

HASTIE, T.J., TIBSHIRANI, R.J., 1990, *Generalized Additive Model*. St. Edmundsbury Press, Great Britain.

LANDAGAN, O. Z., and BARRIOS, E.B., 2006, An Estimation Procedure for a Spatial-Temporal Model. *Statistics and Probability Letters*, 77: 401–406

MARKER, D.A., 2001, Producing Small Area Estimates from National Surveys: Methods for Minimizing Use of Indirect Estimators, *Survey Methodology*, 27: 183-188

OPSOMER, J.D., 1999, Asymptotic Properties of Backfitting Estimators, *Journal of Multivariate Analysis*, 73: 166-179

PARK G. J., WU H., 2004, Backfitting and Local Likelihood Methods for Nonparametric Mixed-Effects Model with Longitudinal Data. Frontier Science and Technology Research Foundation.

PEREIRA, L.N., COELHO, P.S., 2007, Small Area Estimation using Linear Mixed Models with Heterogeneous Covariance Structures of Random Effects.

RAO, J.N.K., 2004, Small Area Estimation with Applications to Agriculture, *Journal of the Indian SocietShaoSahoy of Agricultural Statistics*, 57 (Special Volume), pp. 159-170

RAO, J.N.K., YU, M., 1994, Small-Area Estimation by Combining Time Series and Cross Sectional Data, *The Canadian Journal of Statistics*, 22(4):511-528

SALVATI, N., 2004, Small Area Estimation by Spatial Models: the Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP). Working Paper

SINGH, B.B., SHUKLA, G. K., and KUNDU, D., 2005, Spatio-temporal Models in Small Area Estimation. *Survey Methodology*, 31(2): 183-195.

SHAO, J., TU, D., 1995, *The Jacknife and Bootstrap*. Springer, New York.

2003 Master Sample Documentation, National Statistics Office, Philippines