SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

SMALL AREA ESTIMATION WITH A MULTIVARIATE SPATIAL-TEMPORAL MODEL

by

Arturo M. Martines, Jr
and
Erniel B. Barrios

UPSS Working Paper No. 2009-08
July 2009

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

Small Area Estimation with a Multivariate Spatial-Temporal Model

Arturo M. Martines, Jr.
Asian Development Bank

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

Abstract

A multivariate generalization of a spatial-temporal is postulated. This is used in model-based small area estimation by allowing small area information to be borrowed from other units through the spatial and temporal correlations. An estimation procedure that takes advantage of the backfitting algorithm, AR-sieve bootstrap and Lorenz curve parameterization is proposed.

The model and the estimation procedure is illustrated using data on mean per capita income quintiles of households in the Philippines with provinces as unit of analysis. There is a considerable speed in the convergence of the iterative estimation of the parameters. Furthermore, generation of unit-record synthetic household income is feasible even if statistical modeling is done at the provincial level. Estimates of poverty indices based on the synthetic unit-record data generated from the multivariate spatial-temporal model are generally more reliable than the direct survey estimates. There are only small deviations between the model-based and direct survey estimates of poverty indices at the domain level validating the accuracy of the model-based small area estimates generated from the multivariate spatial-temporal model.

Keywords: multivariate spatial temporal model, backfitting algorithm, AR-sieve, small area estimation, Lorenz curve parameterization

I. Introduction

Classical models that assume independence of observations can loss large amount of information from data that inherently contains temporal or spatial relationships or their interactions. Spatial correlations manifest when localized interventions are implemented in a neighborhood where the units from which measurements are collected from. Temporal correlations can manifest when random shocks persist to affect future observations of the time series. Monitoring data usually exhibit the temporal and spatial dependencies that are optimally accounted by spatial-temporal models.

Small area estimation suffers from the difficulty of finding adequate and reliable information (exogenous variables) at the small area level that can support estimation of the parameter of interest. While existing methods usually benefits from borrowing information from other units within the survey or borrows information from other sources, it is imperative whether borrowing of information from neighboring small areas and from the same small areas from another time point is viable. The interaction of spatial and temporal dependence can yield large amount of information than when one of the sources is ignored.

There is a growing literature that establishes spatial-temporal modeling as a small area estimation technique, see for example [15]. Estimation in a spatial-temporal model was complicated because of the complex structure of the error variance-covariance matrix associated with the model. However, there an increasing amount of work trying to find simple ways to estimate parameters in a spatial-temporal model, for example, [10] uses the backfitting algorithm embedded with the Cochrane-Orcutt procedure.

Poverty modeling benefits a lot from the assumptions that a spatial-temporal model considers. [3] postulated some spatial-temporal models and found evidence of geographic clustering of provinces in the Philippines on the basis of different poverty indicators. A number of model-based approaches have been proposed to strengthen direct survey poverty estimates. Most of these procedures involve modeling either at the small area level (e.g., model with provincial poverty rate as the dependent variable) or at the unit level (e.g., model with household per capita income as the dependent variable). The latter approach usually requires vast amount of data such as unit-record survey and administrative/census data.

Poverty is a multidisciplinary issue and for purposes of monitoring, many indicators are used to account for various aspects of the problem. While these indicators represent different aspects of the problem, they are inherently interrelated. A univariate analysis may fail to take advantage of such relationship resulting to loss in information. This study proposes to model at the small area level in a multivariate context some indicators of poverty. The resulting model is used to simulate unit-level income or consumption data that can be used to compute different income based-poverty indicators at the small area level. In particular, we use the grouped distribution data on income at the small area level as the dependent vector. This study proposes an iterative computing procedure to estimate parameters of a multivariate spatial-temporal model using a combination of backfitting algorithm in a multivariate regression and vector autoregression (VAR) estimation.

2. Spatial-temporal Models and Small Area Estimation

Small area estimation entails generation of estimates at lower levels of disaggregation (e.g., geography, anthropological group, vulnerable groups, etc.) than what the survey domains are capable of generating reliable estimates. The demand for small area statistics has been growing continuously due its operational significance as information support to economic planning, delivery of social service and decision making at the local level. There is also a growing pattern of moving towards decentralization from centralized planning specially among developing countries.

Small area estimates can still be produced by applying design-unbiased methods on data from sampled units per small area. However, small sample sizes or in some cases, absence of a sample from that small area can pull down efficiency of direct survey estimators. Small area estimation builds on the idea that other smaller units or possibly other data sources will lend information to enrich the estimation at the local area of interest, increasing accuracy of the estimates.

There is a rich literature on small area estimation techniques ranging from simple ratio estimation to the more complex modeling strategies. The choice of a technique should mainly be geared towards addressing the specific problem of interest as well as the characteristics of the data available. As an example, ratio estimators are typically used when there are broad areas that are large enough to permit reliable direct survey estimates but small enough such that all small areas within a given broad area are homogenous with respect to the characteristic being measured. The estimates are computed by applying the rate (e.g., poverty rates) for the broad area as indicated by its corresponding direct estimator to the small area population that may be available from census or other administrative data. On the other hand, ratio estimators with auxiliary data can be used in association with broad area ratio estimators utilizing information given by a variable that is correlated with the characteristic of interest. The main limitation of this estimator is the fact that it does not allow for other effects and focuses only on a single variable correlated with the target indicator.

One class of estimation method used in small area estimation is to construct a model that will establish relationship between the response (target) variable and the covariates following for instance the linear model $Y = X\beta + \varepsilon$. While surveys produced myriad of information on different indicators, it will be an arduous task to develop different models for each indicators. [9] argued that the development and thorough testing of a model-dependent estimator may be justified by the increased precision that it brings when there is a characteristic or set of variables that are of utmost importance. In such cases, using typical regression techniques, a synthetic estimator of Y, say $\hat{Y}$ is equal to $X\hat{\beta}$. [13] cautioned that random area effects are not taken into account in developing an estimate of Y under this framework. The extension of fixed effects model to the random effects model or their combination (mixed model) can fill the gap of regression techniques. The error term of the linear model is divided into two components, contribution of the small area and pure error. The former type of error usually arises from information peculiar to each observation that is not captured by the covariates. In poverty analysis, [6] postulated a model of per capita household expenditure which allows cluster correlation in the disturbance term. Residual location effects may potentially diminish the precision of welfare estimates, thus, it is important to explain the variations in consumption due to location as far as possible with the choice and construction of the covariates X's.

The error ε in the model $Y=X\beta+\varepsilon$ may contain spatial diffusion effects and other neighborhood characteristics. It can also contain temporal effects that are not accounted by the X's. [7] argued that this corresponds to a spatial-temporal database that deals with geometries changing over time. One of the major challenges in small area estimation methodology is the difficulty of finding adequate and reliable exogenous variables at the small area level that can capture the variability of the parameter of interest.[15] suggested the exploitation of spatial autocorrelation amongst the small area units in the form of a spatial model in a class of model-based estimation methods. For time series data, the temporal dependencies can be included into a generalization of a spatial-temporal model, see for example [12]. To estimate per capita consumption expenditure, [15] postulated a spatial-temporal model as a Kalman filter to improve the direct survey estimators.

Estimation of spatial-temporal models is done using various approaches. [13] and [14] used Bayesian approach to estimate the parameters. Using Markov Chain Monte Carlo (MCMC) techniques based on Metropolis Hastings algorithm on cancer mortality data in Germany, [14] simultaneously estimated the spatial and temporal autocorrelation, dispersion and temporal trend in the small area estimation model. A weakness to this method is that due to the large number of parameters being estimated from a relatively small data, the resulting parameters are strongly dependent. The Metropolis-Hastings algorithm takes longer burn-in periods to achieve stationarity of the posterior distribution. [11] investigated the asymptotic properties of the maximum likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE) for spatial autoregressive (SAR) models. [10] proposed an estimation procedure for the spatial-temporal model that imbeds the Cochranne-Orcutt procedure into the backfitting algorithm. The general idea of the procedure is to alternately estimate the parameters β, ϱ and γ in an iterative framework.

3. Multivariate Spatial-Temporal Model

Consider the model

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{u}_i \quad (1)$$

$$\mathbf{u}_i = W_i\varrho + \mathbf{v}_i \quad (2)$$

where $\mathbf{Y}_i$ is the 1 x r response vector from i[th] unit, $\mathbf{X}_i$ the 1 x p vector of correlates from i[th] unit, and $\mathbf{u}_i$ is the error component, postulated as equation (2). W_i is a neighborhood variable accounting for spatial dependencies and ϱ is its corresponding spatial effect on $\mathbf{u}_i$. The remainder disturbances $\mathbf{v}_i$ is distributed with mean zero and constant variance. The spatial parameters along with those of the covariate effects are estimated separately from the temporal parameters in the subsequent estimation procedures.

3.1 Estimation of the Covariate Effects and Spatial Parameters

Assuming additivity, equations (1) and (2) can be combined into

$$\mathbf{Y}_i = \mathbf{X}_i\beta + W_i\varrho + \mathbf{v}_i = X_i^*\beta^* + \mathbf{v}_i \quad (3)$$

The ordinary least squares estimator of β^* is $\mathbf{(X^{*\prime}X^*)^{-1}X^{*\prime}Y}$. Estimation of β and ϱ using the backfitting algorithm can yield some advantages over the simultaneous least squares estimation.

[2] noted that estimation of parameters of an additive model using a modified backfitting algorithm may resolve the potential problem in least squares estimation where the design matrix can become ill-conditioned. In backfitting, parameters are estimated sequentially, thus, the design matrix does not suffer from the ill-conditioning that large dimensions could potentially cause. The estimation procedure based on the backfitting algorithm is outlined as follows:

Step 1: In (3), the spatial component is ignored and the resulting model is treated like an ordinary multivariate regression and the parameters β are estimated. Compute the residual vector ***u*** per ith unit. The errors represented by these residuals contain information on the spatial component that is ignored in the initial estimation of β.

Step 2: Another multivariate regression is performed on the residual vector ***u*** in Step1 with the neighborhood variable to estimate the spatial component ϱ. Compute the predicted values $\hat{u}$ from the estimated model.

Step 3: A new response vector is then recomputed adjusting for the spatial component, i.e., subtract the predicted values $\hat{u}$ computed from Step 2 from the original dependent vector **Y**. This will set aside the spatial effect to focus on the correlate effect when a multivariate regression on **Y*** = **Y** - $\hat{u}$ is again fitted in Step 1. The inputs in Step 2 are revised and estimation implemented, and iteration continues. The iteration converges when there are minimal changes in the values of β and ϱ (i.e, $\hat{\beta}$ and $\hat{\rho}$ are taken to the parameter estimates at the final iteration).

A bootstrap procedure is used to assess the efficiency of estimators for β and ϱ. Resampling can generate the empirical distribution of $\hat{\beta}$ $[\hat{\beta} \sim (\mu_{\hat{\beta}}, \Sigma_{\hat{\beta}})]$ and $\hat{\rho}$ $[\hat{\rho} \sim (\mu_{\hat{\rho}}, \Sigma_{\hat{\rho}})]$. A loss function such as the Mahalanobis distance between $\mu_{\hat{\rho}}$ and $\hat{\rho}$ can be used to assess the performance of $\hat{\rho}$, this is given by

$$D(\hat{\rho}, (\mu_{\hat{\rho}}, \Sigma_{\hat{\rho}})) = (\hat{\rho} - \mu_{\hat{\rho}})' \, \Sigma_{\hat{\rho}}^{-1} (\hat{\rho} - \mu_{\hat{\rho}})$$

(4)

An analogous formula can be used to assess the estimator of each of the correlate effects in $\hat{\beta}$, one at a time.

3.2 Estimation of the Temporal Parameters

Equation (3) is extended to account for the temporal dependencies into the following:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{W}\varrho + \mathbf{v} \text{ where } \mathbf{v}_t = \mathbf{v}_{t-1}\gamma + \eta_t$$

(5)

The procedure proposed by [10] where the covariate parameters β and temporal effect γ are simultaneously estimated using regression model with autocorrelated errors through Cochrane-Orcutt procedure may not be optimal for short time series data of length T < r. For time T, we would also like to borrow strength (in the context of small area estimation) from time T-1, T-2,…., 1. using the concepts of vector autoregression (VAR) and the AR-sieve boostrap to propagate the data points. This is then used to estimate β and ϱ as in Section 3.1.

Following [10], assume constant covariate and spatial effects across time periods. We then use the same $\hat{\beta}$ and $\hat{\rho}$ for all time points to compute the vector of disturbances $v_1, v_2, \ldots, v_T$ for each location. To estimate the temporal parameter, the following algorithm is implemented.

Step 1: In each of the T time points, we have r vector of disturbances $\mathbf{v}_t$. For T < r, AR-sieve bootstrap is used to lengthen the time series. In particular, for i[th] location, the composition of the vector of disturbances $\mathbf{v}_t$ is denoted by:

$$
\begin{gathered}
(v_{i,1}^{(1)}, v_{i,1}^{(2)}, \ldots, v_{i,1}^{(r)}) \\
(v_{i,2}^{(1)}, v_{i,2}^{(2)}, \ldots, v_{i,2}^{(r)}) \\
\vdots \\
(v_{i,T}^{(1)}, v_{i,T}^{(2)}, \ldots, v_{i,T}^{(r)})
\end{gathered}
\tag{6}
$$

To implement AR-sieve bootstrap, consider the Seemingly Unrelated Regression (SUR) representation for a VAR(1) model given by:

$$
\begin{gathered}
v_{i,t}^{(1)} = \gamma_{11} v_{i,t-1}^{(1)} + \gamma_{12} v_{i,t-1}^{(2)} + \ldots + \gamma_{1r} v_{i,t-1}^{(r)} + \eta_{i.t}^{(1)} \\
v_{i,t}^{(2)} = \gamma_{21} v_{i,t-1}^{(1)} + \gamma_{22} v_{i,t-1}^{(2)} + \ldots + \gamma_{2r} v_{i,t-1}^{(r)} + \eta_{i.t}^{(2)} \\
\vdots \\
v_{i,t}^{(r)} = \gamma_{r1} v_{i,t-1}^{(1)} + \gamma_{r2} v_{i,t-1}^{(2)} + \ldots + \gamma_{rr} v_{i,t-1}^{(r)} + \eta_{i.t}^{(r)}
\end{gathered}
\tag{7}
$$

Estimate γ_{jj} by fitting AR(1) model separately for each of the $v^{(j)}$ and γ_{jk} , $j \neq k$ by the correlation of $v^{(j)}$ with $v^{(k)}$. Substitute these preliminary estimates in the VAR model to compute for the residual vector $\eta_{i,t}$ per time period. From the T vectors $\eta_{i,1}$, $\eta_{i,2}, \ldots,$ $\eta_{i,T}$, choose at random one vector say, $\eta_{i,t}$. Given these preliminary estimates γ_{ik}, and the sampled vector $\eta_{i,t}$, the T time period data can now be lengthened into T + 1 using the VAR model. We continue the process of lengthening the time series data until there is 'enough' time periods and there are minimal changes in the values of γ_{ik}'s.

Step 2: Once the vector of temporal effects $\hat{\gamma}$ are estimated for all n locations, there will be n sets of γ_{ik} , $j = 1,2,..,r$; $k = 1,2,\ldots, r$. The final estimates of γ_{ik} are obtained by an ordinary bootstrap from the n sets. By construction, this vector $\hat{\gamma}$ is unbiased for the mean of its empirical distribution.

The two algorithms above help mitigate the curse of dimensionality in a multivariate spatial-temporal model when there are few time points available, commonly experienced in small area estimation. We provide a strategy for artificially propagating the data so that the multivariate model can be estimated even with small data sets. Although the method is computationally intensive, it provides a simple alternative to the more complicated estimation procedures. Furthermore, the method takes advantage of the desirable small sample performance of the bootstrap for time series that is known to be locally efficient, see for example, [4].

4. Estimation of Small Area Poverty

[5] discussed the use of grouped data on income or expenditures to estimate poverty measures by simulating synthetic unit-level income or expenditures. The parametric Lorenz curves are fitted on the grouped data. Grouped data on income (or expenditures) are correlated (i.e., the mean of 1st quintile is likely to be correlated with the mean of 2nd quintile, etc) and a multivariate approach is useful. From surveys, grouped data can be easily computed at the small area level. A spatial-temporal model is proposed to improve the reliability of these estimates. The following algorithm is proposed:

Step 1: Per small area, compute direct survey estimates of the grouped data. This serves as the response vector for the spatial-temporal model. Following Section 3, estimate $\hat{\beta}$, $\hat{\rho}$ and $\hat{\gamma}$.

Step 2: Using Lorenz GQ and Beta parameterization, simulate synthetic data of income (or expenditure) per small area. This is done by constructing (k, L) data points from $\hat{\mathbf{Y}}$ to estimate the curvature parameters of the parametric Lorenz function L(k). Synthetic data are produced by evaluating the product of the mean per capita income and derivative of the Lorenz function, at different values of k.

Step 3. Using the synthetic data, estimate the FGT measures per small area.

[5] proposed a parametric approach in estimating poverty measures without needing to simulate unit-level data on income or expenditure. The poverty measures computed from the parametric approach are expected to be approximately equal to poverty measures computed from the synthetic unit-level data. The synthetic unit-level data is useful in estimating the accuracy (i.e., bias, variability) of the poverty estimates through bootstrap procedures.

5. Application

The multivariate spatial-temporal model for small area estimation of poverty measures is applied to the Philippines' Family Income and Expenditures Survey (FIES) data, a survey conducted triennially by the Philippine National Statistics Office. Since the survey considers the 17 regions as the domains, the small areas are the 83 provinces that are smaller than the regions. The response vector **Y** is composed of the following variables: the average household per capita income; and average household per capita income for each of the five quintiles. From a long list of independent variables considered to form **X,** (both available from the survey and the 2000 Census), the following three correlates are used in the modeling: (1) proportion of households whose head has college education; (2) proportion of households whose head is employed in agriculture, hunting or forestry, and; (3) proportion of households with roofs and walls made of weak materials. To define the neighborhood systems, average per capita income of nearby provinces, average per capita internal revenue allotment of nearby provinces, and average per capita expenditure on social services of nearby provinces are considered as spatial distance measures. These three indicators can potentially account for the spatial spillover among the provinces that can possibly affect poverty among the neighboring provinces. A spatial model is constructed using these indicators and the 2003 FIES data. Information from 1994, 1997 and 2000 FIES are also used to estimate the temporal effect. The per capita income for 1994, 1997 and 2000 were all expressed at 2003 prices using consumers' price index to induce comparability. Moreover, out of the 83 provinces, only 70 provinces are used in modeling the vector of per capita income quintiles since some provinces were created only during some rounds of the FIES. Furthermore, provinces whose direct survey estimates of per capita income quintiles have larger sampling errors were not included in model building, along with those that have very small sample sizes.

5.1 Estimates

With the estimation algorithm in Section 3, point estimates of β and ϱ using backfitting algorithm are converging after the fourth iteration. This is in agreement with the observation of [10] on the speed of convergence using backfitting algorithm. Minimal difference of $\hat{\beta}$ across different models (i.e., models using different neighborhood variable) are noted, this is somehow an evidence of robustness of the method to the choice of a spatial neighborhood system. This may also be explained by the appropriateness of the three neighborhood systems to capture the spatial spillover of the effects of poverty (or poverty alleviation). This is also an advantage from estimating the parameters using backfitting algorithm over ordinary multivariate regression where we include the neighborhood variable as one of the columns in **X**. By estimating the spatial effect ϱ through regression of the neighborhood variable on the partial residual vector **u**, it is easier to differentiate the spatial effect from the correlates of the dependent variable.

The Mahalanobis distance loss function is used to evaluate the estimators for β and ϱ. The mean vector and the covariance matrix of $\hat{\beta}$ and $\hat{\rho}$ are computed from its empirical distribution. Using bootstrap procedures, 1000 replicates of the parameter estimates were generated using different resample sizes. The values of the loss functions tend to be more stable over the iterations for larger sample sizes. Varying the resample size produces minimal effect on the behavior of the parameter estimates with respect to the loss function. Although the estimators are biased, the values of the loss function are almost negligible. This further validates the usefulness of the backfitting algorithm that works satisfactorily especially when the additivity of the model assumption holds.

The predictive ability of the model is evaluated using mean absolute percentage error (MAPE). The high values for MAPE can be explained by the fact that the direct estimates at the small area level are volatile, expecting MAPE to be not too low, see Table 1 for details.

[Table 1 Here]

In the course if simulation[1], the dependent vector is expressed into its natural logarithmic form to minimize the volatility of the dependent variables. However, the transformation can possibly invalidate the additivity assumption of the model. In addition, we only considered a limited number of independent variables and used them in modeling the response vector. Additional independent variables may be needed to capture the peculiarities of each quintile, but still taking into account the correlation of these quintiles with each other.

It can also be noted that the MAPEs for the dependent variable provincial mean per capita income for the three spatial models are generally lower than the MAPEs for the other random variables. This is because the provincial mean per capita incomes for each quintile also contribute to the overall mean. Thus, in the multivariate framework, information on the mean per capita income for all quintiles helps in predicting the mean per capita income.

---

[1] When the dependent vector is expressed into its natural logarithmic form, approximately 20% reduction in MAPE has been noted.

Some problems are expected in estimating γ either through AR-sieve or backfitting. Since there are only very few data points, in this example, only 4 time points per province, after a few iterations, a set of seemingly-convergent parameter estimates is achieved. After the first few iterations of AR-sieve, the estimates seem to converge because there is an "overfitting" lurking in the procedure. The parameter estimates of VAR(1) model seem not to differ significantly between two consecutive iterations. However, when two or three additional iterations of lengthening the data through AR-sieve are continued, the values can become totally different, see Table 2 for details.

[Table 2 Here]

Thus, it may be misleading to conclude that parameters estimates converge after the first m iterations when the reason of minimal differences between two consecutive iterations can only be attributed to overfitting. Suppose the iteration is continued even when there is already an evidence of convergence, during several iterations after the abrupt convergence stage, the parameter estimates may change significantly between consecutive iterations. "True" convergence is likely to be achieved when there are considerably many time points. However, even if the "true" convergence is achieved, one problem that has to be taken into consideration is the possibility that the pseudo-data dominated the time series data such that the parameter estimates are influenced more by the patterns from the pseudo-data rather than from the actual data. Since we have started with only four time points, it is possible that the underlying stochastic process that gave rise to the realization is not embodied in only four observations. In such case, the resulting $\hat{\gamma}$ as an estimate of the temporal effect may be hard to interpret. One way to resolve this problem is to limit the number of parameters estimated in the VAR(1) model to minimize the effect of the simulated data. It is also interesting to note that while doing AR-sieve to lengthen the time series data, γ_{ii} are more stable than γ_{ik}, $i \neq k$. This leads to the idea of possibly restricting $\gamma_{ik} = 0$, $i \neq k$ to reduce the VAR(1) model into a model with only AR(1) parameters. Such model may also be easier to interpret because this implies that the error term in Y_i (i.e., the random variable v_i) at time T, i = 1,2,…, 6 is influenced by the error term of Y_i at time T-1. If γ_{ik} are not restricted, the VAR(1) model implies that the error terms of Y_i are also influenced by the error terms of Y_j j = 1,2,…,6, $j \neq i$ at T-1, which again, adds complexity to the estimation problem. It should be noted that while AR-sieve is a plausible tool to lengthen time series data it can work more efficiently if the data already contains the time series patterns for all the VAR(1) parameters.

The spatial-temporal models with the restriction $\gamma_{ik} = 0$ for $i \neq k$ are evaluated using MAPE in Table 3.

[Table 3 Here]

Recall that using the spatial model, the MAPE for the mean per capita income are generally lower than the MAPE using the other dependent variables because the latter help in predicting the mean per capita income in the multivariate context. In the spatial-temporal models, the MAPE for the first dependent variable is within the range of MAPEs for the remaining dependent variables. The possible reason is that up to a certain extent, the multivariate properties are ignored because we restricted $\gamma_{ik} = 0$ for $i \neq k$. Such restriction reduced the VAR(1) form for the vector representing the temporal effect to univariate AR(1) models.

The temporal effect is also estimated through backfitting that works well in estimating β and ϱ. To borrow strength from the immediate past data, and since we have a panel of provinces, we can merge $\mathbf{v}_{2000}$ with the cross-section data in 2003 to proceed with the estimation. The only difference now is instead of doing the backfitting approach in two stages, we do it in three.

Further, recall that when we estimated the parameters of the multivariate spatial model, five iterations were used. Since more parameters are being estimated in a multivariate-spatial temporal model, it is ideal to add few more iterations after convergence.

At a first glance, it seems that $\hat{\beta}$ and $\hat{\rho}$ are converging after few iterations of the backfitting algorithm. But if we take a closer look, the estimates at the j^{th} iteration is closer with the estimates at $(j+2)^{th}$ iteration, $(j+4)^{th}$ iteration and so on. Similarly, the estimates at $(j+1)^{th}$ iteration are closer with the values noted at $(j+3)^{th}$ iteration, $(j+5)^{th}$ iteration, etc., these are not observed when models used ignores temporal effect. Hence, the problem may have been caused by the addition of a temporal component in the model. The major setback that we observed from adding suspected insignificant components in the model is that the convergence of the parameter estimates for other components can also be affected. Thus, as we did in the AR-sieve approach, restricting γ_{ik}, $i \neq k$ to be null can be considered.

5.2 Evaluation of the Multivariate Spatial-Temporal Model in Small Area Estimation

The unit-level synthetic income generated using parametric Lorenz functions for each province are aggregated at the regional level since the regions form the survey domains. The poverty incidence was computed using this set of synthetic income for each province and aggregated to the regions. Table 4 summarizes the estimates based on design-unbiased methods (direct survey estimate) and the spatial model-based estimates with three different neighborhood systems.

[Table 4 Here]

The regions seem to exhibit significantly different spatial model-based estimates from the direct survey estimates of poverty incidence. The provinces that caused the deviations at the regional level are those excluded in modeling due to inadequate time points. Except for Zamboanga del Sur and Misamis Oriental, the provinces in Table 5 have been excluded in the estimation during the model-building stage. The spatial model-based poverty incidence estimates without these provinces are now comparable with the direct survey estimates in Table 6.

[Table 5 Here]

[Table 6 Here]

Without these "peculiar" provinces in Table 5, the resulting spatial model-based poverty estimates are at least comparable with the direct survey estimates, at the domain level. As indicated by the standard errors depicted in Table 6, the spatial model-based estimates are more reliable than the corresponding survey estimates. Expectedly, the spatial model-based estimates at the small area level (i.e., provincial level) also manifest lower coefficients of variation.

Addition of a temporal component in the spatial model increased the forecast ability of the model. However, there are some provinces whose spatial-temporal model-fitted grouped data yielded invalid Lorenz curves using either Beta or GQ parameterizations.

7. Conclusions

Poverty reduction has been the overarching goal of most nations around the globe. This universal objective is strengthened through the adoption of the Millennium Development Goals outlined by the United Nations. Recognizing that poverty alleviation programs may be more efficient if these are targeted at the local level, small area poverty estimation has been one of the forefront subjects for research. [1] even pointed out that there are often large variations in the growth-poverty performance across subnational units (e.g., regions, states, provinces). Researches dealing with cross-country comparisons suggest that incomes of the poor move one-for-one with overall average incomes. In poverty estimation, units such as small areas usually cluster together since the socio-cultural dynamics can easily cause the so-called spatial spillover. The usual independent observations assumption will no longer hold. It becomes operationally useful to integrate a spatial-temporal component in modeling such type of systems.

The results of this study support the observation of [3], substantiating the evidence of spatial clustering among provinces with reference to selected poverty indicators. The use of a single model for the Philippines causing significant deviations between survey-based and model-based estimates of poverty indices for some provinces indicate that implementing a poverty alleviation strategy for a group of provinces may be more efficient than tailor-fitting an alleviation program for all geographic units.

Although AR-sieve can be a potential tool to lengthen the time series so that VAR(1) parameters becomes estimable, the resulting estimates may not be adequately representative of the temporal system where the 4 time point realizations really came from. Estimation of all the parameters of the VAR(1) model resulted to estimates that seem to be more influenced by the pseudo-data generated by using AR-sieve. A restriction on γ_{ik} for $i \neq k$ is necessary if there are very few time points available.

The use of multivariate approach in this study is anchored on two reasons. Clearly, a deeper understanding of the poverty situation requires probing further than separate univariate assessment of different poverty indicators. For instance, it is useful to simultaneously look at the entire income distribution and not only the bottom tail. Although the concept of poverty is essentially synonymous to being at the bottom part of the income curve, information about the middle and upper tail can give a better understanding on the dynamics of poverty situation. The second reason for using a multivariate approach is that the year of interest is a non-census year. [6] outlined how welfare estimates can be computed through a combination of sample survey household information with unit-record population census information, for census years. [8] pointed that small area welfare estimates for non-census years are less reliable and for that reason, are not usually generated.

References

[1] Balisacan, A. and Pernia, E., 2002. Probing beneath cross-national averages: poverty, inequality, and growth in the Philippines. ERD Working Paper Series No. 7, Asian Development Bank, Manila, Philippines.

[2] Barrios, E., 2007. Spatial effect in the efficient access of rural development. ADBI Discussion Paper No. 65, Asian Development Bank Institute, Tokyo, Japan.

[3] Barrios, E. and Landagan, O., Geographic distribution of the poor: Is poverty contaminating?. Proceedings of the 9th National Convention on Statistics (National Statistical Coordination Board, Manila, Philippines, 2004).

[4] Buhlmann, P., 2002. Bootstraps for time series. Statistical Science, 17(1) 52–72.

[5] Datt, G., 1998. Computational tools for poverty measurement and analysis. FCND Discussion Paper No. 50, Food Consumption and Nutrition Division, International Food Policy Research Institute, Washington, DC, USA.

[6] Elbers, C., Lanjouw, J. and Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica, 71(1) 355-364.

[7] Erwig, M., Güting, R., Schneider, M. and Vazirgiannis, M., 1998. Abstract and discrete modeling of spatio-temporal data types. In Proceedings of the 6th ACM International Symposium on Advances in Geographic Information Systems, pp. 131-136, ACM: New York, USA.

[8] Hoogeveen, J., Emwanu, T. and Okwi, P., 2003. Updating small area welfare indicators in the absence of a new census. Available at: http://siteresources.worldbank.org/INTPGI/Resources/Hoogeveen.UpdatingSmallAreaWelfareIndicators.pdf, downloaded: May 7, 2009.

[9] Kalton, G., 2002. Models in the practice of survey sampling (revisited). J. of Official Stat., 18(2) 129-154.

[10] Landagan, O. and Barrios, E., 2007. An estimation procedure for a spatial-temporal model. Stat. and Prob. Letters, 77 401-406.

[11] Lee, L., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial-autoregressive models. Econometrica, 72(6) 1899-1925.

[12] Melendreras, P. and Porcu, E. , Some recent contributions to spatio-temporal modeling. Departamento de Matemáticas, Universitat Jaume I de Castellón, 2006. Available at: http://www.seio2006.ull.es/premio_melendreras/premio_Melendreras_Porcu.pdf, downloaded: May 7, 2009.

[13] Rubio, V. and Richardson, S. , 2006. Bayesian methods for small area estimation using spatio-temporal models. Imperial College London, U.K., Available at: http://www.unavarra.es/metma3/Papers/PDFS_POSTER/gomezrubiovirgilio1.pdf, downloaded: May 7, 2009.

[14] Schach, U., 2001. Modeling approaches to a spatio-temporal small area estimation. Department of Statistics, University of Dortmund, Germany, Available at:

http://www.statistik.tu-dortmund.de/fileadmin/user_upload/Lehrstuehle/MSind/SFB_475/2001/tr19-01.ps, downloaded:May 7, 2009

[15] Singh, B., Shukla, G. and Kundu, D., 2005. Spatio-temporal models in small area estimation. Survey Methodology, 31(2) 183-195.

Table 1. Mean Absolute Percentage Error for Different Neighborhood Systems

Dependent variable	Neighborhood System		
	Spatial Model (Household Income)	Spatial Model (IRA)	Spatial Model (Social Services)
Mean	17.13	17.32	17.36
Quintile1	21.17	21.49	21.66
Quintile2	20.25	20.55	20.70
Quintile3	19.65	19.82	19.96
Quintile4	18.00	18.15	18.24
Quintile5	21.20	21.30	21.35

Table 2. Illustration of possible problem on convergence of $\hat{\gamma}$ estimated using VAR(1) model from pseudo-data generated by AR-sieve

Dependent Variable	Independent Variable	mth iteration Estimate	(m+1)th iteration Estimate	(m+2)th iteration Estimate
Mean	mean at t-1 (residual v)	-0.74	-0.66	----
	quintile1 at t-1 (residual v)	----	----	0.01
	quintile2 at t-1 (residual v)	0.62	0.57	-0.06
	quintile3 at t-1 (residual v)	----	----	----
	quintile4 at t-1 (residual v)	0.03	0.02	0.06
	quintile5 at t-1 (residual v)	----	-----	0.02
Quintile1	mean at t-1 (residual v)	-0.35	-0.35	-0.01
	quintile1 at t-1 (residual v)	0.17	0.18	-0.01
	quintile2 at t-1 (residual v)	----	----	----
	quintile3 at t-1 (residual v)	----	----	-0.01
	quintile4 at t-1 (residual v)	-0.08	-0.09	0.03
	quintile5 at t-1 (residual v)	----	----	----
Quintile2	mean at t-1 (residual v)	-0.95	-0.83	-0.04
	quintile1 at t-1 (residual v)	0. 76	0.70	0.02
	quintile2 at t-1 (residual v)	----	----	----

	quintile3 at t-1 (residual v)	----	----	----
	quintile4 at t-1 (residual v)	-0.84	-0.82	0.05
	quintile5 at t-1 (residual v)	----	----	0.03
Quintile3	mean at t-1 (residual v)	-0.23	-0.24	-0.17
	quintile1 at t-1 (residual v)	0.25	0.27	0.02
	quintile2 at t-1 (residual v)	----	----	----
	quintile3 at t-1 (residual v)	----	----	----
	quintile4 at t-1 (residual v)	----	----	----
	quintile5 at t-1 (residual v)	0.98	0.96	0.03
Quintile4	mean at t-1 (residual v)	-0.87	-0.86	-0.02
	quintile1 at t-1 (residual v)	----	----	----
	quintile2 at t-1 (residual v)	----	----	0.09
	quintile3 at t-1 (residual v)	-0.57	-0.60	0.01
	quintile4 at t-1 (residual v)	----	----	0.08
	quintile5 at t-1 (residual v)	----	----	0.04
Quintile5	mean at t-1 (residual v)	0.23	0.20	0.98
	quintile1 at t-1 (residual v)	0.81	0.80	-0.02
	quintile2 at t-1 (residual v)	----	----	0.06
	quintile3 at t-1 (residual v)	----	----	----
	quintile4 at t-1 (residual v)	----	----	----
	quintile5 at t-1 (residual v)	----	----	----

Table 3. Mean Absolute Percentage Error for Different Neighborhood Systems With Model Restrictions

Dependent Variable	Neighborhood Systems		
	Spatial-Temporal Model (Household Income)	Spatial-Temporal Model (IRA)	Spatial-Temporal Model (Social Services)
Mean	9.95	10.13	10.27
Quintile1	8.55	8.38	8.39
Quintile2	9.33	9.31	9.37
Quintile3	10.05	10.06	10.06
Quintile4	10.07	10.04	10.11
Quintile5	13.53	13.55	13.62

Table 4 Comparison of Poverty Incidence, Model-based and Direct Survey Estimates

Region	Direct Survey Estimates		Neighborhood System					
			Spatial-Temporal Model (Household Income)		Spatial-Temporal Model (IRA)		Spatial-Temporal Model (Social Services)	
	Poverty Incidence	Standard Error	Poverty Incidence	Standard Error	Poverty Incidence	Standard Error	Poverty Incidence	Standard Error
1	20.47	1.46	19.89	0.42	19.94	0.42	19.70	0.41
2*	23.25	1.54	40.61	0.49	39.50	0.45	41.34	0.52
3*	8.50	0.74	14.89	0.35	15.02	0.29	14.94	0.36
4a	9.89	0.94	10.83	0.27	11.17	0.31	11.14	0.34
4b*	37.19	2.19	24.89	0.36	24.56	0.42	24.82	0.42
5	40.85	1.83	36.23	0.47	36.39	0.48	36.40	0.47
6	31.67	1.74	27.48	0.42	27.43	0.37	27.40	0.40
7	33.54	2.18	25.51	0.40	25.14	0.39	25.31	0.39
8	43.30	1.94	42.11	0.50	41.99	0.50	42.00	0.49
9*	53.81	2.82	36.65	0.46	36.94	0.50	36.56	0.44
10*	39.98	2.35	25.04	0.45	24.89	0.48	24.85	0.43
11	31.84	2.20	27.07	0.52	27.18	0.47	26.85	0.42
12	37.73	2.35	42.94	0.55	43.09	0.47	42.81	0.50

NCR	1.27	0.21	1.04	0.11	1.12	0.10	1.19	0.12
CAR*	20.65	2.07	27.24	0.50	27.00	0.45	27.15	0.48
ARMM	45.70	2.77	41.29	0.45	40.49	0.44	40.86	0.51
Caraga*	48.00	2.40	35.17	0.45	34.93	0.46	34.82	0.45

** regions whose model-based estimates of poverty incidence are not within 3 standard deviations of direct survey estimates*

Table 5 Comparison of Poverty Incidence for Selected Provinces, Model-based and Direct Survey Estimates

Region	Province	Direct Survey Estimate	Neighborhood System		
			Spatial-Temporal Model (Household Income)	Spatial-Temporal Model (Household Income)	Spatial-Temporal Model (Household Income)
2	Nueva Vizcaya	9.50	28.67	27.69	29.07
	Quirino	24.39	46.51	45.28	47.20
3	Aurora	24.59	34.75	35.78	35.37
4b	Occidental Mindoro	36.59	45.33	44.96	45.23
9	Zamboanga del Sur	43.46	21.87	21.60	21.25
10		35.29	22.86	22.94	22.45

	Camiguin				
	Misamis Occidental	48.22	20.18	19.90	20.05
	Misamis Oriental	29.26	9.60	9.22	9.28
CAR	Ifugao	26.53	52.86	53.02	52.96
Caraga	Agusan del Norte	36.05	13.82	13.28	13.43

Table 6 Comparison of Poverty Incidence for Model-based and Direct Survey Estimates (Using average per capita income of nearby provinces as neighborhood variable)

Region	Direct Survey Estimates		Model-based Estimates		Provinces Excluded
	Poverty Incidence	Standard Error	Poverty Incidence	Standard Error	
1*	20.47	1.35	25.75	0.52	
2	23.18	1.16	26.51	0.43	Quirino
3	8.50	0.65	9.25	0.21	
4a	11.06	1.06	9.74	0.25	Batangas, Rizal
4b	40.65	2.79	38.70	0.70	Oriental Mindoro, Romblon
5	40.85	1.46	41.04	0.55	
6*	30.09	1.31	34.91	0.48	Antique, Guimaras
7*	24.38	1.67	33.74	0.54	Bohol, Negros Oriental

8	43.30	1.79	44.28	0.53	
9	53.48	1.88	47.12	0.45	
10	38.30	1.79	41.09	0.54	Bukidnon
11	32.36	1.81	30.31	0.51	Davao del Norte
12	37.45	1.74	39.00	0.52	
NCR	1.21	0.24	1.01	0.11	2nd district
CAR	19.07	1.59	23.66	0.44	Abra
ARMM*	39.29	2.96	54.53	0.56	Maguindanao
Caraga	47.99	1.93	43.24	0.49	

* *Regions whose model-based estimates of poverty incidence are not within 3 standard deviations of direct survey estimates*