Modelling Rice Yield in the Philippines using Dynamic Spatio-Temporal Models

Stephen Jun V. Villejo
School of Statistics
University of the Philippines Diliman

Recent unpredictable and extreme weather episodes and infestation are some of the realistic occurrences of structural change in the agricultural system which produce outliers and extreme values in our data, and consequently pose problems when building statistical models. An estimation procedure which is robust to structural change is therefore necessary. Three spatial-temporal models with varying dynamic characteristics of the parameters are postulated each with a different estimation procedure for the agricultural yield in irrigated areas of the Philippines. One of which is a robust estimation procedure using forward search algorithm with bootstrap in a backfitting algorithm. The other two algorithms also used the backfitting algorithm but infused with the Cochranne-Orcutt procedure. The robust estimation procedure and the other one which considers varying parameter across space gave competitive predictive abilities and are better than the ordinary linear model. Simulation studies show the superiority of the robust estimation procedure over the Cochranne-Orcutt procedure and ordinary linear model in the presence of structural change.

Keywords: Spatial-temporal model; Backfitting algorithm; robust estimation; additive model

1. Introduction

The main stimulus of the development of this spatial-temporal model is in modelling agricultural production which accounts for spatial and temporal dependencies. In an agricultural system, there are dependencies among spatial units and the error term will most likely contain autocorrelation. Classical methods that assume independent observations are unarguably violated in data sets indexed by space and time, hence, classical regression models may yield misleading results.

Spatial and time interaction is very evident in the context of agricultural production (Landagan and Barrios, 2007). Similar crops are planted in the same location due to geographical, climatic, and soil conditions. Adjacent locations are most likely similar or resembles almost the same characteristics in terms of soil quality, availability and quality of irrigation systems, amount of rainfall, terrain and climate or weather condition. In addition to the spatial dependence, there could also exist dependence among values adjacent in time due to seasonal weather patterns. Moreover, yield of crops is affected by physical and geographical conditions, indicator/quantitative variables that represent information within the area with similar conditions, and another term which accounts for production shocks. These components in the model will vary through space and time. Systems with such complexities which exhibit spatial and temporal dependence is best facilitated by dynamic spatio-temporal models. In this way, there can be a better understanding of the dynamic relationships in the system and to give accurate forecasts for policy-making purposes.

Moreover, recent unpredictable and extreme weather episodes and infestation are some of the realistic occurrences of structural change in the system. These episodes are coined as "structural change." According to Hackl and Westlund (1989), an essential element of "structural change" is nonconstancy of the parameters and could be evident in time-series or cross-sectional data. In other words, there is a change in the system and in the behavior of the random variables which occurs in a certain period of time or in particular neighborhoods. Structural change will be evident in many cases and these occurrences are just temporary since after a certain period of time, the series will go back to its original and regular behavior. However, they have adverse effects for instance in the production of agricultural crops and will be a problem in the estimation of the statistical model. An estimation procedure which is robust to structural change is therefore necessary.

Three estimation procedures are proposed in this study, which are based on three postulated models describing the dynamic structure of the system. One of the proposed estimation procedures is based on the backfitting algorithm imbedded with bootstrap and forward search algorithm for robust estimation. The other two algorithms are modifications of the Landagan and Barrios (2007) model through the assumption of varying covariate, spatial and temporal parameters. Results are compared to the Cochranne-Orcutt procedure and ordinary regression. Simulation studies are also conducted to assess the quality of the proposed model under specified conditions and assumptions.

2. Related Literature

Landagan and Barrios (2007) proposed an estimation procedure of a spatial-temporal model using the backfitting algorithm infused with the Cochrane-Orcutt procedure. The main motivation of the paper is in understanding agricultural

production in the Philippines particularly the yield of certain crops. Yield of crops is affected by geophysical conditions like the area harvested, by spatial parameters which gives information on characteristics within areas of similar conditions, and an error term which captures other information and shocks. The following model is postulated:

$$Y_{it} = \beta X_{it} + \delta W_{it} + \varepsilon_{it}, \text{ where } \varepsilon_{it} = \mu_i + \nu_{it} \quad (1)$$

where Y_{it} is the response variable from location i and time t, X_{it} the set of covariates from location i at time t, W_{it} the set of variables in the neighborhood system of location i at time t, and ε_{it} the error component that would take the disturbances. The error component is postulated to follow an auto-regressive behavior. An important assumption of the model is constant covariate, spatial and temporal effect across space and across time. The idea is to alternately and iteratively estimate the parameters with the covariate and spatial effects first then the temporal effect. The covariate considered in the paper is area harvested. Several neighborhood variables are considered, namely count of adjacent units with yield greater than national yield, average production of neighboring units, average yield of neighboring units, count of units within a region with production greater than the national median production, count of units within a region with yield greater than the national median yield, regional mean production and regional mean yield. The postulated spatial-temporal model is found to be superior to commonly used models like the ordinary linear model, linear model with auto-correlated errors and mixed linear model. Among the seven neighborhood variables, regional average yield gave the smallest mean absolute prediction error. The proposed method by Landagan and Barrios is generally capable of summarizing relationships among observations related in space and time as long as appropriate neighborhood variables are used. The model is also superior over the ordinary linear model, linear model with auto-correlated errors, and mixed linear model.

Dumanjug, Barrios, Lansangan (2010) introduced bootstrap methods into the algorithm of Landagan and Barrios (2007). Two bootstraps were introduced: the first method using the ordinary bootstrap and the second method using time blocks of consecutive observations. The confidence intervals of the covariate and temporal effects were narrower from using the second method. The second method gave more efficient and reliable bootstrap estimates. Furthermore, the simulation study showed that the model gave very low prediction errors.

Generalized additive models

Hastie and Tibshirani (1986) introduced the class of *generalized additive models* which is similar to the normal linear regression model and the linear logistic model where the covariates assume a linear or other parametric form. They also proposed a method for its estimation. The linear form $\Sigma\beta_j X_j$ is replaced by a sum of smooth functions $\Sigma s_j X_j$, hence the term *generalized additive models*.

The additive model generalizes the linear regression model and it is therefore assumed that

$$E(Y \mid X_1, X_2, ..., X_p) = s_0 + \sum_{j=1}^{p} s_j(X_j). \quad (2)$$

The s_j (.)'s are smooth functions where E [s_j (X_j)]= 0.

The s_j (.)'s are estimated in an iterative manner using the procedure called the *local scoring algorithm* which uses scatter plot smoothers to generalize the usual Fisher scoring procedure for computing maximum likelihood estimates. The smooth functions produced by the local scoring can be used as a data description, for prediction, or to suggest covariate transformations. Generalized additive models provide a flexible method for identifying nonlinear covariate effects in exponential family models and other likelihood-based regression models.

Backfitting algorithm

Hastie and Tibshirani (1986) introduced the class of generalized additive models which is similar to the normal linear regression model and the linear logistic model where the covariates assume a linear or other parametric form. The linear form $\Sigma\beta_j X_j$ is replaced by a sum of smooth functions $\Sigma f_j X_j$, hence the term generalized additive models. The iterative process used by Hastie and Tibshirani to estimate the additive regression model is the backfitting algorithm (Friedman and Stuetzle, 1981). The backfitting algorithm enables the fitting of an additive model where the predictor effects can be examined separately. The estimation can be done using any regression-type mechanism and can be parametric, nonparametric, or semi-parametric in nature. The additive model is defined and characterized as follows:

$$Y = \alpha + \sum_{i=1}^{p} s_j(X_j) + \varepsilon \quad (3)$$

where the X's are the predictors, the error term ε is independent of the X's, $E(\varepsilon) = 0$ and $Var\ (\varepsilon) = \sigma^2$. The s_j's are arbitrary univariate functions,

i. Initialize: $\alpha = ave\ (y_i), f_j = f_j^0, j = 1, ..., p$

ii. Cycle: $j = 1, \ldots, p$

$$f_j = S_j\left[\left(y - \alpha - \sum_{k \neq j} f_k\right)\middle| x_j\right]$$

iii. Continue (ii) until the individual functions do not change where S_j denotes a smoothing of the response y against the predictor X_j.

Another way of viewing the process is given by Hastie and Tibshirani (1986). Suppose the model is given by $Y = s_0 + \sum_{i=1}^{p} s_j(X_j) + \varepsilon$ and the partial

residual is defined by $R_j = Y - s_0 - \sum_{k \neq j} s_k(X_k)$, then $E\left[R_j \mid X_j\right] = s_j(X_j)$ and minimizes $E\left[Y - s_0 - \sum_{k=1}^{p} s_k(X_k)\right]^2$. This provides a way for estimating each $\hat{s}_i(\cdot)$ given estimates $\{\hat{s}_i(\cdot), i \neq j\}$.

The algorithm starts with the initialization:

$$s_0 = E(Y), s_1^1(\cdot) = s_2^2(\cdot) = \cdots = s_p^p(\cdot) = 0, m = 0.$$

Then iterate: $m = m + 1$ for $j = 1$ to p and do:

$$R_j = Y - s_0 - \sum_{k=1}^{j-1} s_k^m(X_k) - \sum_{k=j+1}^{p} s_k^{n-1}(X_k)$$

$$s_j(X_j) \quad E(R_j \mid X_j).$$

The iteration will continue until:

$RSS = E\left[Y - s_0 - \sum_{j=1}^{p} s_j^m(X_j)\right]^2$ fails to decrease.

The RSS does not increase at any step of the algorithm which assures convergence.

Buja et al. (1989) proved convergence and consistency of the backfitting in an additive model with a general smoother. Chen and Tsay (1993) has shown that the backfitting procedure will work in the time-series context if the dependence structure is not quite strong. The convergence and asymptotic property of the estimators using backfitting algorithm are studied by Opsomer (2000).

Forward search algorithm

The forward search algorithm is a powerful method to detect multiple masked outliers, to determine their effects on models fitted to the data, and to detect systematic model inadequacy. (Atkinson and Riani, 2007). The algorithm starts by choosing an initial outlier-free subset of m observations from the *n* observations and obtaining robust estimates from the subset. The algorithm will move forward to a larger subset by getting the squared residuals from the *n* observations using the least squares fit to the subset of *m* observations and using the *m* +1 observations as the new larger subset. Usually, only one observation is added at each step, but sometimes two or more is added in instances when at least one leave which is often and indication of the introduction of some of a cluster of outliers. In the algorithm, a series of parameter estimates will be obtained starting from very robust estimates to least squares at the end. The observations which

are far from the fitted model which are usually outliers, unidentified subset, or an indication of systematic failure of the model, will enter at the end of the search. Quantities indicative of model quality and inadequacy will be monitored.

Bootstrap

The bootstrap method is a general method used to estimate the sampling distribution of some random variable on the basis of some observed data (Efron, 1979). Let $X = (X_1, X_2,..., X_n)$ denote the random sample from a completely unspecified probability distribution F and $x = (x_1, x_2,..., x_n)$ be the realization of the random variable. The bootstrap method is accomplished by drawing a random sample with replacement of size n wherein all the $x_1, x_2,..., x_n$ have the same chances of selection. The bootstrap distribution can be accomplished by direct theoretical calculation, Monte Carlo approximation, or Taylor Series expansion which has similar form as the Jackknife method.

3. The Spatial-Temporal Models

Similar to Landagan and Barrios (2007), this paper is motivated to characterize and understand the dynamics of agricultural production. Adjacent locations or spatial units in the same neighborhood are homogenous with respect to soil quality, availability of irrigation systems, amount of rainfall, terrain, and climate or weather conditions. Yield for certain crops can be affected by factors such as physical and geophysical conditions, indicator variables which represent information within the area with similar conditions, and an error term that can account for production shocks such as irregular weather, infestation, etc.

Three models are postulated as follows:

$$\text{Model 1: } Y_{it} = \beta_t X_{it} + \delta_t W_{it} + \epsilon_{it} \tag{4}$$

$$\text{Model 2: } Y_{it} = \beta_i X_{it} + \delta_i W_{it} + \epsilon_{it} \tag{5}$$

$$\text{Model 3 : } Y_{it} = \beta_i X_{it} + \delta_t W_{it} + \epsilon_{it} \tag{6}$$

where Y_{it} is the response variable from location i at time t, X_{it} the set of covariates from location i at time t, Wit the set of variables in the neighborhood system of location i at time t, and ϵ_{it} the error component that would take the disturbances. In the spatial-temporal model explored in this paper, the response variable Y_{it} is the yield of irrigated rice at province i and time t, the covariate X_{it} is the area harvest in province i at time t, the only spatial variable explored is the regional mean yield which gave a competitive predictive performance for lag 1 models according to Landagan and Barrios (2007). The error term is assumed to follow some autoregressive behavior and is postulated as

$$\epsilon_{it} = \rho_i \epsilon_{i,\, t\text{-}1} + u_i \tag{7}$$

The disturbances follow a one-way error component with $u_i \sim IID\ (0, \sigma_u^2)$ and the remaining disturbances following AR(1).

For the first postulated model in (4), the covariate (β_t) effect and spatial effect (δ_t) varies across time. For the second postulated model in (5), both the covariate and spatial effect varies across locations. Lastly, for the third postulated model in (6), the covariate effect varies across space while the spatial effect varies across time. For all the three postulated models, the temporal effect varies across space.

4. Estimation of the Model

Algorithm 1: Estimation of model 1

The backfitting algorithm is used to estimate the parameters of the spatial-temporal model where the parameters of the model will be estimated iteratively. Forward search algorithm and bootstrap strategy is incorporated in the backfitting procedure to give robust estimates. The algorithm is given in the following steps:

Step 1. Using the realizations $\{y_{it}\}$, $i=1, \ldots, N$, $t=1, \ldots, T$, where i is the index for location and t the index for time, β_t and δ_t will be simultaneously estimated using ordinary least squares regression for each of the T time points. Forward search algorithm will be incorporated in each of the ordinary least squares estimation through the following steps:

i. A subset of size n, $n < N$ from N observations for each time point will be chosen. The n observations are ideal and outlier-free. The ordinary least squares will be fitted on the full data set. The choice of the n observations corresponds to the smallest n residuals.

ii. Fit the model $Y_{it} = \beta_t X_{it} + \delta_t W_{it} + \epsilon_{it}$ to the selected n observations and generate the parameter estimates $\hat{\beta}_t$ and $\hat{\delta}_t$.

iii. The fitted values will be computed to the $N–n$ left-out observations and the residuals will be obtained.

iv. The observation corresponding to the smallest residual from the $N–n$ residuals will be included in the subset of observations from (i).

v. Fit the model $Y_{it} = \beta_t X_{it} + \delta_t W_{it} + \epsilon_{it}$ on the $n + 1$ observations in (iv).

vi. The previous procedure will be repeated iteratively adding one observation at a time until all N locations have been included in the model or until the model is behaving wildly based on the Cook's D. The Cook's D is said to be influential if its value exceeds $\frac{4}{n}$ where n is the number of observations. The algorithm then stops if the Cook's D is no longer influential to the model based on this threshold.

The forward search is used to obtain robust estimates of the spatial and covariate parameters. Observations which behaves wildly based on the postulated model will not be included in the estimation of the parameters.

Step 2. After obtaining the final subset from the forward search algorithm for each time point, the following is performed for each time point:

i. Draw a simple random sample of size m with replacement.

ii. Fit the model $Y_{it} = \beta_t X_{it} + \delta_t W_{it} + \in_{it}$. This gives a value of the statistic computed using the bootstrap sample in (i).

iii. Repeat (i) and (ii) B times, yielding B $\hat{\beta}_t$ and $\hat{\delta}_t$. The Monte Carlo estimates for the mean and standard deviation of δ_t will be obtained.

$$\widehat{\delta_t^*} = \frac{1}{B}\sum_{K=1}^{B} \widehat{\delta_{tk}} \tag{8}$$

$$S.E.(\widehat{\delta_t^*}) = \frac{1}{B-1}\sum_{K=1}^{B} (\widehat{\delta_{tk}} - \widehat{\delta_t^*})^2 \tag{9}$$

$$\widehat{\beta_t^*} = \frac{1}{B}\sum_{k=1}^{B} \widehat{\beta_{tk}} \tag{10}$$

$$S.E.(\widehat{\beta_t^*}) = \frac{1}{B-1}\sum_{k=1}^{B} (\widehat{\beta_{tk}} - \widehat{\beta_t^*})^2 \tag{11}$$

From the notation, $\widehat{\delta_{tk}}$ is the kth bootstrap estimate of δ_t while $\widehat{\beta_{tk}}$ is the k[th] bootstrap estimate of β_t.

From this step, the set of parameter estimates $\left\{\widehat{\delta_1^*}, \widehat{\delta_2^*}, ..., \widehat{\delta_T^*}, \widehat{\beta_1^*}, \widehat{\beta_2^*}, ..., \widehat{\beta_T^*}\right\}$ with their standard errors is obtained. Optimality is achieved since the covariate and spatial parameter are simultaneously estimated for each time point.

Step 3. For each Y_{it}, generate the residuals $e_{it} = Y_{it} - \widehat{\beta_t^*} X_{it} - \widehat{\delta_t^*} W_{it}$. The computed residuals will contain information on the true error and the temporal parameter. For each location, AR(1) estimation will be performed to get the temporal parameter estimates.

Step 4. In this step, we have the set of parameter estimates $\left\{\widehat{\delta_1^*}, \widehat{\delta_2^*}, ..., \widehat{\delta_T^*}, \widehat{\beta_1^*}, \widehat{\beta_2^*}, ..., \widehat{\beta_T^*}, \hat{\rho}_1^*, \hat{\rho}_2^*, ..., \hat{\rho}_N^*\right\}$. A new dependent variable will then be computed adjusting for the temporal component, given by $Y_{it}^{new} = Y_{it} - \hat{\rho}_i \varepsilon_{i,t-1}$.

After Step 4, the algorithm will go back to *Step 1* using the new values of the dependent variable. After removing the temporal effect, the focus will be on the covariate-spatial effect when the ordinary least squares is once again fitted in *Step 1*. In updating the estimates of the error term in *Step 3*, the original values

of Y will be used. After updating the error terms and performing *Step 3*, the new dependent variable will once again be computed using the original values of the dependent variable and the updated estimates of the error terms from *Step 3*. The iteration converges when there are minimal changes in the values of the parameter estimates.

An alternative model postulate is assuming constant covariate effect across time points. As an example, if the X_{it} is the area harvested at time t, this means that if area harvested is high at $t = 0$, the area harvested will remain to be high at $t = 1, 2, \ldots, T$. In this framework, the estimation procedure will change particularly in the estimation of the constant covariate effect β. The Monte Carlo Estimate of β will be computed as follows:

$$\hat{\beta}^* = \frac{1}{BT}\sum_{i=1}^{T}\sum_{k=1}^{B}\widehat{\beta_{lk}} \tag{12}$$

The index i is for the time point while the index k is for the bootstrap sample. Thus, $\widehat{\beta_{lk}}$ is the k[th] bootstrap estimate of β from time point i. The standard error of the estimate is then computed as follows:

$$S.E.(\hat{\beta}^*) = \frac{1}{BT-1}\sum_{i=1}^{T}\sum_{k=1}^{B}\left(\widehat{\beta_{lk}} - \hat{\beta}^*\right)^2 \tag{13}$$

Algorithm 2: Estimation of model 2

For the postulated model (2), a modification of Landagan and Barrios (2007) is implemented. The estimation is as follows:

Step 1: For each spatial unit, the parameters β_i and ρ_i are estimated simultaneously using the Cocchranne-Orcutt procedure. The residuals, which contains information on the spatial component, will be generated.

Step 2: Another regression will be performed on the residuals from Step 1. Since the spatial effect is assumed to be varying across locations, the estimation of the spatial effect δ_i will also use the regression model with autocorrelated errors using the Cochranne-orcutt procedure. The estimates for the ρ will be disregarded in this step. This will be performed per spatial unit.

Step 3: A new dependent variable will be computed adjusting for the spatial component, given by $Y_{it}^{new} = Y_{it} - \hat{\delta}_i W_{it}$. Another regression model with autocorrelated errors will be fitted in Step 1. The iteration converges when there are minimal changes in the parameter estimates. When the backfitting algorithm have converged, another regression model will be fitted but this time all the parameter will be estimated simultaneously.

Algorithm 3: Estimation of model 3

The estimation of model (3) is basically similar to the estimation procedure proposed by Landagan and Barrios (2007) but with varying parameters. That is, in every step of the estimation procedure the estimate for the parameters will not be averaged.

5. Results and Discussion

In modelling yield of rice in irrigated areas in the Philippines, the covariate considered is the total area harvested, while the neighborhood variable is the regional mean yield. The total area harvested is shown to be a good proximate indicator of agricultural efficiency under specified conditions (Landagan and Barrios, 2007). The spatial variable considered is the regional mean yield which showed to have a competitive mean absolute percentage error from the model using the proposed estimation procedure from the same paper. Logarithmic transformations were made on the variables for modelling purposes. Autoregressive structure of order 1 was explored. The data used is a quarterly data from 1994 to 2014 second quarter with 76 provinces and 16 regions. Thus, the length of the time series is 82 with 76 spatial units. In such a long series, autocorrelation and spatial correlation would most likely be evident.

Other than the 3 proposed algorithms, other estimation procedures were run for comparison purposes, wherein most of them are just variants and modifications of the proposed algorithm in the estimation of model 1. The modifications are the following: constant covariate effect; without bootstrap; forward search algorithm only, no backfitting, without bootstrap, but all parameters will be estimated; and forward search only, no bootstrap, no backfitting, no auto-regression. The other models are ordinary linear model; and regression with auto-correlated errors estimated using the Cochranne-Orcutt procedure.

In the Landagan and Barrios (2007) proposed algorithm, changes in the parameter estimates for the backfitting algorithm usually becomes negligible after second of third iteration. In this paper's proposed estimation procedure on the postulated model 1 with varying parameters, the convergence would take relatively longer. For the case when the covariate effect is assumed constant, changes in the parameter estimates would stabilize after 4 or 5 iterations. For the cases when all the parameters are assumed to be varying, the backfitting algorithm would converge longer than usual. The changes in the estimates of the spatial effect and temporal effect would be negligible after 4 iterations and 5 iterations respectively, but not for the estimate of the covariate effect whose parameter changes in the iterations are unstable. The change in the estimate of the covariate effect would be acceptable after 15 iterations. For the estimation of model 2, the estimation converged after three iterations. The third proposed algorithms took 8 iterations before convergence.

Parameter estimates

Model 1: β and δ varies across time

Since total area harvested is a proximate indicator of agricultural efficiency, the parameter β_t accounting for the covariate effect is expected to be positive. From the models estimated, the models using the proposed algorithm with varying parameters had 2 negative β_t estimates while the model using the proposed algorithm without bootstrap had 1 negative β_t estimate. For all the other models which are modifications of the proposed algorithm, the covariate effect estimates for all time points are positive. Despite some anomalies discovered for 2 out of 82 time points, it can still be concluded that area is a proximate indicator of agricultural efficiency. As harvested area increases, efficiency in the use of farm technologies and other inputs can be maximized leading to higher productivity.

Even when the spatial and temporal parameters are introduced, the proposed estimation procedure did not lead to major adjustments in the signs of the β_t estimates. In the proposed estimation procedure of Landagan and Barrios (2007), the estimate of the constant covariate effect β is negative. It is hypothesized that the simultaneous estimation of β and ρ (the constant temporal effect) caused the automatic adjustments in the parameter resulting to the negative sign of β. In this paper's proposed algorithm, the varying covariate effect and spatial effect are simultaneously estimated first before estimating the temporal effect. The different order in the estimation may have avoided the shift in the signs of the covariate and spatial parameter estimates.

The spatial parameter γ_t estimate is also expected to be positive. In all the models estimated, the sign of the effect of γ_t for all time points is consistently positive. Also, from Landagan and Barrios (2007), this is consistently estimated to be positive for both cases when only the spatial component is considered and when both spatial and temporal components are considered.

Model 2: β and δ varies across space

The parameter estimates obtained from the second proposed estimation procedure is almost similar to those obtained from the usual Cochranne-Orcutt procedure. The effect of the spatial variables are positive for all locations. The effect of the covariate is not consistently positive for all spatial units. Nonetheless, majority of the spatial units has a positive estimate of the covariate effect.

Model 3: β varies across space and δ varies across time

The estimates of the covariate effects obtained from implementing the third proposed estimation procedure are positive for all spatial units. The estimates for the spatial parameters for all time points are also positive. The result obtained from here is similar with model (1).

Predictive ability of models

The proposed algorithm 1 will be superior over the other models when the assumption of varying covariate and spatial effect across time and varying temporal effect across locations is satisfied. Moreover, the advantage of the algorithm which uses robust estimation procedures will be evident under structural change. Below are some historical plots of the covariate and neighborhood variable for four sample provinces. Figure 1 shows the historical plot of regional yield for four provinces. For the earlier time points, there is a possible occurrence of structural change in which province _32 behaved differently from the other 3 provinces and hence could be a possible outlier in the dataset. There is also a possible occurrence of structural change in the middle of the series in which the same province has the different behavior. In Figure 2, a historical plot on area harvested for another set of 4 provinces is shown. It is very noticeable that province _30 has a different behavior from the rest somewhere in the earlier time point. This outlying behavior is most likely influential and has a serious effect in the estimation procedure and modelling stage. This possible structural change is crucial and are best facilitated by robust estimation procedures such as the proposed algorithm.

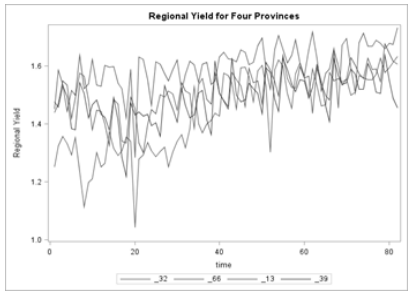

Figure 1. Historical Time Plot of Regional Mean Yield for Four Provinces in the Philippines

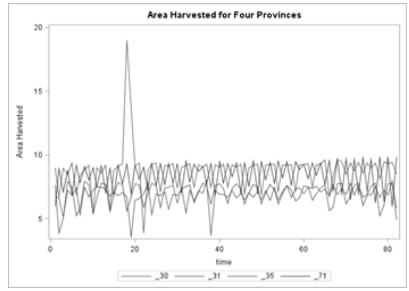

Figure 2. Historical Time Plot of Area Harvested for Four Provinces in the Philippines

The predictive power of all the explored estimation procedures are assessed using the in-sample mean absolute prediction error (MAPE). The MAPE for the different models are presented in Table 1. The proposed algorithm assuming constant covariate effect gave a MAPE of 7% which is relatively higher than the other estimation procedures. This is an evidence that the covariate effect is not constant, and would vary across time or across space. There is an improvement in the predictive ability of the model from the ordinary linear model to the proposed algorithm assuming varying parameters. Ordinary linear model gave a MAPE of 5.8789%. After introducing the forward search algorithm, there is 0.23% reduction

in the error over ordinary regression model. Since forward search algorithm is a robust procedure, it is expected to give better prediction ability especially when there are outliers in the dataset. A slight improvement in the MAPE could be an evidence that there are no extremely influential values in the yield for all spatial units. This could also be an effect of the logarithmic transformation done on the variables. Furthermore, estimating the temporal effects through auto-regression also lead to a reduction in the predictive error from 5.87% to 5.32%. Temporal components are important for the agricultural system and including them in the model will make the predictions more accurate. For the procedure using the proposed algorithm but without bootstrap, the MAPE is 5.29%. There is a slight improvement from the previous model which performs forward search and autoregression only. The backfitting algorithm eliminates some of the biases in the estimation which lead to more accurate estimation of the parameters. This could be the reason for the improvement of the predictive ability of the model. The proposed algorithm 1 gave the most competitive predictive ability among all the procedures from the modifications of the proposed algorithm. This is contributed by the robust procedure through the forward search, the bootstrap, and backfitting algorithm.

Table 1. Mean Absolute Percentage Errors (MAPE) for different models

Estimation Procedure	MAPE
Proposed Algorithm 1: β and δ varies across time	5.2914
Proposed Algorithm 1 but β is constant	6.9995
Proposed Algorithm 1 but without Bootstrap	5.2926
Forward Search only, No Bootstrap, No Backfitting	5.3155
Forward Search only, No Autoregression, No Backfitting	5.6422
Ordinary Least Squares: β and δ varies across time	5.8769
Cochranne-Orcutt Procedure: β and δ varies across space	3.8158
Proposed Algorithm 2: β and δ varies across space	3.8597
Proposed Algorithm 3: β varies across space and δ varies across time	9.8148

The MAPE obtained from the third proposed algorithm gave the poorest MAPE of 9.82%. The assumptions from the postulated model 3 therefore does not hold on the data resulting to inferior predictive ability. The MAPE obtained from the second proposed algorithm gave the smallest from the 3 proposed estimation procedures. Among the three postulated models, the assumptions for the second model best characterizes the spatio-temporal relationship of interest. The obtained MAPE 3.86% is slightly larger by 0.04% than the one obtained from the fitted regression model with autocorrelated errors. The MAPE of the models are comparable. This could be an evidence that the spatio-temporal model of yield

in the Philippines is more represented by varying parameters across locations and space.

Occurrence of structural change is very likely to be present in the dataset. However, algorithm 1 which was expected to give the best model fit and smallest prediction error did not give the highest predictive ability and is inferior over algorithm 2 and the Cochranne-Orcutt procedure. A strong reason for this is that the assumptions of the proposed algorithm do not hold for the data. Nonetheless, the fit of algorithm 1 still gives a competitive predictive performance. For datasets which will follow the postulated model 1 and in the presence of structural change or contamination, algorithm 1 will be superior and will give the best predictive ability. algorithm 2 will give the smallest MAPE for cases when the covariate effect and spatial effect varies across locations while algorithm 3 will have a comparative advantage when the covariate effect varies across locations and the spatial effect varies across time. However, it should be noted that a consequence of doing robust estimation through forward search algorithm is that not all points will be included in the computation of the estimates which might have negative consequences on the predictions especially on points which are considered outlier or extreme values.

6. Simulation Study

The first postulated model along with the estimated procedure will be evaluated using a simulated data in the balanced ($N = T$) and unbalanced ($N > T$) scenarios. The true model is given below:

$$Y_{it} = \beta X_{it} + \delta_t W_{it} + \in_{it} \tag{13}$$

$$\in_{it} = \rho_i \in_i, t - 1 + a_{it}, \ |\rho_i| < 1, a_{it} \sim IID\ (0, \sigma^2 a) \tag{14}$$

The covariate X_{it} was generated from Normal (100, 10) while the spatial variable was generated from the Poisson distribution with varying parameters. There are two scenarios on the number of clusters or neighborhoods: 5 and 10. In the case where there are 5 clusters, the parameters of the Poisson distribution are 10, 20, 30, 40 and 50. In the case where there are 10 clusters, the parameters are 10, 20, 30,…, 80, 90, 100. For the balanced data case, the number of time points and spatial units is 20. For the unbalanced case, there will be 30 time points and 50 spatial units. The simulation study aims to recreate the reality of the agricultural behavior with structural change. The structural change is nested on the following features: contamination in the recent period of the time series vs. contamination in the middle of the time series, structural change in all clusters in the time points affected by the contamination versus structural change in 20% of the clusters (if 5 clusters, only 1 is affected; if 10 clusters, 2 clusters are affected), and the change in the parameter in the contaminated spatial units to be 10% and 20%. The number of clusters will show the performance of the estimation procedure when the

population is divided to small number of neighborhood systems versus otherwise. The scope of contamination will be facilitated by making all clusters affected by the contamination versus having only a subset of clusters contaminated. The setting where only 20% of the clusters is contaminated may be viewed as the case wherein the structural change occurred within specific locations. The presence of the contamination is either at the middle time points or at the recent time points of the series. The percentage of time points affected by the contamination is set to be 10%.

Simulation results

Table 2 shows the in-sample MAPE comparing the proposed algorithm, Cochranne-Orcutt procedure, and ordinary linear model under different scenarios.

Table 2. Table on the MAPEs for the different simulation scenarios

					Contamination in Recent Periods of the Time Series			**Contamination in Middle Periods of the Time Series**		
					MAPE					
	T	**N**	**Cluster**	**Cont**	**Proposed Algo-rithm**	**Co-chranne Orcutt**	**OLS**	**Proposed Algo-rithm**	**Co-chranne Orcutt**	**OLS**
Contamination in all clusters	20	20	5	0.1	4.91	26.35	26.61	4.95	26.66	26.98
	20	20	5	0.2	5.18	26.67	26.9	5.18	27.09	27.45
	20	20	10	0.1	4.93	26.54	26.54	4.98	26.85	27.19
	20	20	10	0.2	5.21	26.88	26.88	5.22	27.3	27.67
	30	50	5	0.1	4.72	29.21	29.21	4.75	29.23	29.66
	30	50	5	0.2	4.99	29.75	29.87	5.04	29.48	29.41
	30	50	10	0.1	4.75	29.25	29.41	4.78	29.27	29.73
	30	50	10	0.2	5.02	29.8	29.95	5.07	29.54	29.47
Contamination in 20% of clusters	20	20	5	0.1	4.94	26.38	26.74	4.93	26.41	26.77
	20	20	5	0.2	5.17	26.4	26.77	5.19	26.47	26.73
	20	20	10	0.1	4.93	26.57	26.94	4.95	26.6	26.83
	20	20	10	0.2	5.19	26.6	26.98	5.15	26.66	26.97
	30	50	5	0.1	4.74	28.98	29.20	4.78	29.01	27.04
	30	50	5	0.2	4.99	29.06	29.27	5.01	29.08	29.20
	30	50	10	0.1	4.74	29.02	29.27	4.78	29.05	29.27
	30	50	10	0.2	4.98	29.1	29.35	5	29.12	29.35

The proposed algorithm shows superior predictive ability under the different scenarios considered. Therefore, if the assumption of varying (or constant) covariate effect and spatial effects across locations is satisfied, the proposed algorithm is superior over the commonly used estimation procedures. Another advantage of the estimation procedure is its robustness in the presence of structural change.

7. Conclusions

Modelling yield of rice in irrigated areas of the Philippines is best represented by a model with varying covariate, spatial, and temporal effects across locations and with an estimation procedure which will consider that dynamic aspect of the system. That is the reason why the proposed algorithm 2 gave the best predictive performance among the 3 proposed algorithms. Moreover, the advantage of optimality from the simultaneous estimation of the covariate and spatial parameter also led to better forecasts. From the modifications of the proposed algorithm 1, the contribution of the robust estimation through forward search and bootstrap is evident in the reduction of the prediction error. Even if the reduction in the prediction error seems to be small, this could be an effect of the logarithmic transformations done on the dataset. The proposed algorithms 1 and 2 are generally capable of summarizing relationships in a spatio-temporal model. Since the neighborhood variable used in the models is regional mean yield which is selected based on the results from Landagan and Barrios (2007), the consistency of the quality of results from the estimation procedures could be affected by the appropriateness of the neighborhood indicators used.

The prediction performance of the estimation procedure will depend on the assumptions that will hold in the dataset. Based on the simulation study, the results of estimating the postulated spatio-temporal model A using the proposed algorithm 1, is superior compared to the ordinary linear model estimated using ordinary least squares and regression model with autocorrelated errors using Cochranne-Orcutt procedure. Extending the simulation studies to more settings and scenarios can be explored. Also, a more extensive simulation study of the two proposed algorithms will be the next step.

References

LANDAGAN, O. and BARRIOS, E.,2007, An estimation procedure for spatio-temporal model, *Statistics and Probability Letters* 77: 401-406.

DUMANJUG, C., BARRIOS, E. and LANSANGAN, J., 2010, Bootstrap procedures in spatiotemporal model, *Journal of Statistical Computation and Simulation* 80: 809-822.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models, *Statistical Science* 1(3): 297-318.

FRIEDMAN, J. H. and STUELTZE, W., 1981, Project pursuit regression, *Journal of the American Statistical Association* 76: 817-823.

ATKINSON, A. and RIANI, M., 2002, Forward search added-variable t-tests and the effect of masked outliers on model selection, *Biometrika* 89: 939-946.

EFRON, B., 1979, Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* 7(1):1-26.

HACKL, P. and WESTLUND A. H., 1989, Statistical analysis of "structural change": An annotated bibliography, *Empirical Economics* 14(2): 167-192