SCHOOL OF STATISTICS

UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

Estimation of Isotonic Spatio-temporal Model with Clustering

Michael Daniel C. Lucagbo
School of Statistics
Magsaysay Ave., University of the Philippines Diliman
Diliman Quezon City, 1101 Philippines
mclucagbo@yahoo.com

Erniel B. Barrios
School of Statistics
Magsaysay Ave., University of the Philippines Diliman
Diliman Quezon City, 1101 Philippines
ebbarrios@up.edu.ph
(Correspondence)

UPSS Working Paper No. 2015-02
March 2015

Abstract

We propose a spatio-temporal model for contagion with possible clustering of spatial units. It further considers monotonicity of spatial dependence-related covariates in the estimation via the backfitting algorithm embedded with FD-GMM bootstrap and monotone regression using integrated *B*-splines. A simulation study was designed to evaluate predictive ability of the model. The procedure performs best when there are few clusters, the time series is long or the cluster sizes are much greater than the time series length. Furthermore, if the model is properly specified, predictive ability is fairly robust to temporal stationarity.

Keywords: Spatio-temporal model, isotonic regression, FD-GMM bootstrap, backfitting algorithm, clustering, contagion

1. INTRODUCTION

Statistical models are used to describe the transmission of a contagious entity (such as a disease or a rumor) from person to person, or from the source to a community, between communities, etc. These models, called contagion models, have also found analogous application in econometrics, stock markets, asset allocations, among others. Defined in a broader sense as the transmission of an influence from one individual to another, the concept of contagion "occupies an important place both in biology – specifically in mathematical epidemiology – and in the social sciences, where it is manifested in problems as diverse as the diffusion of innovations, the spread of cultural fads, and the outbreak of political or social unrest" (Dodds and Watts, 2004).

Spatio-temporal models have become popular in analyzing disease transmission as they account for both the incidence from the time of the onset of the disease to the present and the distance between the source and the susceptibles. The epidemic contagion model presented in this study allows for change in number of contaminated individuals over time, as well as the spatial dependency among units within a geographic cluster. It is derived from the assumption that the number of affected persons in a spatial unit is a function of the incidence of the disease in the recent past, the distance of the spatial unit from the source, and the population size. Units closer to the source are naturally more prone to infection than those farther. Working on these assumptions, this study constructs a model and tests it for variable magnitudes of exposure and heterogeneity in the susceptibility of spatial units.

The model is postulated assuming homogeneous time affect across the clusters, and a heterogeneous spatial effect. Estimation of the time effect is implemented using panel data estimation methods such as the First-Difference Generalized Method of Moments Bootstrap (FD-GMMb), (Santos and Barrios, 2011). Capitalizing on the advantages of bootstrap methods for small-sample cases, (Santos and Barrios, 2011) noted that the FD-GMMb is optimal for small samples.

The spatial effect is estimated as a monotone regression in view of the fact that the incidence of a disease in a spatial unit is roughly a monotone function of the distance of the spatial

unit from the nucleus. The variogram, a value that gives information about the correlation between points that are separated by spatial distances, is used in estimating the number of units which are contaminated due to spatial correlation of the spatial unit with the nucleus. The estimation procedure for the monotone regression follows the procedure by (Winsberg and Ramsay, 1980), which performs monotonic transformation by integrating linear combinations of B-splines. The backfitting algorithm is then performed until the estimates converge.

2. REVIEW OF RELATED LITERATURE

2.1 First-Difference Generalized Method of Moments

In contrast to the usual panel data model, the dynamic panel data (DPD) model uses the lagged dependent variable as an explanatory variable. As in univariate time series analysis, modeling the dependency of the time series on its past value/s gives valuable insights on the temporal behavior of the series.

$$Y_{it} = \emptyset Y_{i,t-1} + \sum_{k=1}^{K} \beta_k X_{itk} + u_{it} \quad i = 1, \cdots, N; \quad t = 1, \cdots, T$$

One desirable property of the FD-GMM estimator is its consistency as both $N \to \infty$ and $T \to \infty$ provided that $(\log T)^2/N \to 0$. Moreover, the number of the orthogonality conditions $q = T(T-1)/2$ tends to infinity as $T \to \infty$. The FD-GMM estimator is also asymptotically normal provided $T \to \infty$ and $lim(T/N) = 0$. The asymptotic bias of FD-GMM has order $(1 + \emptyset)/N$. While the literature is replete with the asymptotic optimality properties of the FD-GMM estimator, there are doubts on its performance for small samples. Many panel data are usually formed from small samples of time points and/or panel units because of the structural change or random shocks that may occur in larger datasets. For small samples, Santos and Barrios (2011) have proposed parametric bootstrap on the FD-GMM in mitigating the bias and inconsistency that these estimators are known to exhibit for small samples. The bootstrap procedure uses AR sieve in replication.

2.2 *B*-splines

Suppose that there are *n* sets of observations $\{(y_i, x_i);\ i = 1, \dots, n\}$ and suppose that the responses y_i are generated from an unknown distribution $G(y|x)$. It is assumed that the observations on the explanatory variable are sorted by magnitude $x_1 < x_2 < \cdots < x_n$. The regression model based on B-spline basis functions is given by:

$$y_i = \sum_{j=1}^{m} w_j b_j(x_i) + \varepsilon_i$$
$$= \boldsymbol{w}'\boldsymbol{b}(x_i) + \varepsilon_i,\ \ i = 1, \cdots, n$$

where $\boldsymbol{b}(x) = (b_1(x), \cdots, b_m(x))'$ is an *m*-dimensional vector of *B*-spline basis functions and $\boldsymbol{w} = (w_1, \cdots, w_m)'$ is an *m*-dimensional vector of unknown parameters. Following the discussion of Konishi and Kitagawa (2008) on *B*-splines, suppose that the *B*-splines, constructed from polynomial functions, are of degree *k*. The *B*-spline basis function $b_j(x)$ is composed of known piecewise polynomials that are smoothly connected at points t_i, called *knots*. Setting up the knots required to construct m basis functions,

$$t_1 < t_2 < \cdots < t_k < t_{k+1} = x_1 < \cdots < t_{m+1} = x_n < \cdots < t_{m+k+1}$$

By setting the knots in this way, the *n* observations are partitioned into *m* – *k* intervals $[t_{k+1}, t_{k+2}], \cdots, [t_m, t_{m+1}]$. Furthermore, each interval $[t_i, t_{i+1}],\ i = k+1, \cdots, m$ is covered by *k* + 1 *B*-spline basis functions. The algorithm developed by de Boor (1978) can be conveniently used in constructing the *B*-spline basis functions. Generally, a *B*-spline function of degree *k* is written as $b_j(x; k)$. First, a *B*-spline function of degree 0 is defined as:

$$b_j(x; 0) = \begin{cases} 1, & for\ t_j \le x \le t_{j+1} \\ 0, & otherwise \end{cases}$$

Starting from the *B*-spline of degree 0, a *B*-spline function of degree *k* can be obtained using the recursive formula:

$$b_j(x; k) = \frac{x - t_j}{t_{j+1} - t_j} b_j(x; k-1) + \frac{t_{j+r+1} - x}{t_{j+r+1} - t_{j+1}} b_{j+1}(x; k-1)$$

2.3 Backfitting algorithm

The backfitting algorithm is an iterative process of fitting an additive model which allows the use of any regression type mechanism to estimate each univariate function f_j. Hastie and Tibshirani (1990) define the iterative process by:

Step 1: Initialize $f_i = f_i^0, i = 1, \cdots, p$.

Step 2: Cycle: $j = 1,2, \cdots, p, 1,2, \ldots, p, \ldots,$

$$f_j \leftarrow S_j\left(Y - \sum\nolimits_{k \neq j} f_k\right)$$

Continue Step 2 until the individual functions don't change.

The function is fitted simultaneously. The effects of all other variables from Y are removed before smoothing the partial residual on the predictor X_j. This modularity feature of the algorithm allows the accommodation of both parametric and nonparametric procedures in estimating each function. Buja et al. (1989) shows that the backfitting is the Gauss-Seidel iterative method for solving a set of normal equations associated with the additive model. He also provides conditions for consistency and nondegeneracy and proves convergence for the backfitting and related algorithms for a class of smoothers that includes cubic spline smoothers.

3. METHODOLOGY

We postulate the following model:

$$Y_{it}^{(k)} = \emptyset Y_{i,t-1}^{(k)} + \sum_{l=1}^{K}[\lambda^l(h)] + \varepsilon_{it} \quad where \;\; \varepsilon_{it} \sim iid\; N(0, \sigma_\varepsilon^2), \quad i = 1, \ldots, n_k, t = 1, \ldots, T$$

$Y_{it}^{(k)}$ can be interpreted as the incidence of contamination in the i^{th} spatial unit of the k^{th} cluster at time t. There are n_k spatial units in the k^{th} cluster, and K clusters. The time series length corresponding to each spatial unit is T. ϕ is the temporal effect while ε_{it} is a normally distributed error term with mean 0 and constant variance.

This quantity $\lambda^k(h)$ is interpreted as the *additional contamination due to spatial correlation between the susceptibles and the carrier in the cluster*. For observations outside the k^{th} cluster, $\lambda^k(h) = 0$.

3.1 Estimation Procedure

Estimation of the parameter $\emptyset$ and the monotonic function $\lambda^k(h)$ is done iteratively through the backfitting algorithm.

For a fixed cluster *k*,

Initialization

1. Ignoring the terms in λ, regress $Y_{it}^{(k)}$ on $Y_{i,t-1}^{(k)}$.This is the dynamic panel model and we estimate this using FD-GMM Bootstrap (Santos and Barrios, 2011) for each cluster *k*, resulting to $\widehat{\emptyset}_{BS1}$.
2. Compute $e_{1,it}^{(k)} \coloneqq Y_{it}^{(k)} - \widehat{\emptyset}_{BS1} Y_{i,t-1}^{(k)}$.
3. Regress $e_{1,it}^{(k)}$ on *h* using the monotone regression procedure via integrated B-splines.
4. Compute for $d_{1,it}^{(k)} \coloneqq Y_{it}^{(k)} - \hat{e}_{1,it}^{(k)}$ where $\hat{e}_{1,it}^{(k)}$ is the fitted value from the regression in (3).

The estimation of $\emptyset$ in Step 1 is carried out through the FD-GMMb to address the problem of small samples. As Santos and Barrios (2011) have noted, the FD-GMMb mitigates the bias and inconsistency that the FD-GMM estimator is known to exhibit for small samples. On the other hand, monotone nonparameteric regression is used in estimating the function $\lambda^k(h)$ because of the assumption that the additional contamination in a spatial unit which is attributed to the unit's spatial correlation with the cluster nucleus is a monotone function of *h*, the distance between the spatial unit and the nucleus.

Iteration

1. Regress $d_{1,it}^{(k)}$ on $d_{1,i,t-1}^{(k)}$ using FD-GMM Bootstrap. This gives us $\widehat{\emptyset}_{BS2}$ an "improved" estimate of $\emptyset$.
   (Note that $\widehat{\emptyset}_{BS1}$ in the initialization stage is not optimal because of the presence of $\lambda^k(h)$ in the error term)
2. Compute for $e_{2,it}^{(k)} \coloneqq Y_{it}^{(k)} - \widehat{\emptyset}_{BS2} Y_{i,t-1}^{(k)}$
3. Regress $e_{2,it}^{(k)}$ on h using the monotone regression procedure via integrated *B*-splines.
4. Compute for $d_{2,it}^{(k)} \coloneqq Y_{it}^{(k)} - \hat{e}_{2,it}^{(k)}$ where $\hat{e}_{2,it}^{(k)}$ is the fitted value from (3).

The quantity $d_{2,it}^{(k)}$ then replaces $d_{1,it}^{(k)}$ in (1) and steps (1) to (4) are repeated iteratively until the estimate of $\emptyset$ converges. The procedure (initialization and iteration) is repeated for all *K* clusters. Buja et.al. (1989) proves the convergence of the backfitting algorithm. Since the estimation of $\emptyset$ is done in all of the *K* clusters, it will produce *K* estimates of $\emptyset$, say $\widehat{\emptyset}^{(1)}, \cdots, \widehat{\emptyset}^{(K)}$ corresponding to each of the *K* clusters. These *K* estimates are then bootstrapped to produce the final estimate of $\emptyset$ as the model assumes a homogeneous value of $\emptyset$ across clusters. On the other hand, the monotonic function $\lambda^k(h)$ is assumed to vary across clusters. The final estimates of $\lambda^k(h)$ are effectively the values $\hat{e}_{it}^{(k)}$ obtained in the final iteration of the algorithm.

3.2 Simulation Settings

The simulation settings include varying number of clusters, variogram type, perturbation effect, number of spatial units in a cluster, time series length and spatial and temporal effects. The quantity $\lambda^k(h)$ has the form

$$\lambda^k(h) = \beta \times \frac{1}{\left\{c + (S-c)\left(1 - exp\left(-\frac{3h}{a}\right)\right)\right\}} \times P_{i,t-1}^{(k)} + \delta_i^{(k)} \quad where \;\; \delta_i^{(k)} \sim iid\; N\left(0, \sigma_\delta^2\right),$$

$$i = 1, \cdots n_k, k = 1, \ldots, K$$

The expression $c + (s - n)\left(1 - exp\left\{-\frac{3h}{a}\right\}\right)$ is called the *variogram* (in particular, the Exponential variogram). Following the discussion of Manly (2001), the quantity *S* is called the sill (largest possible value), *c* is called the nugget (smallest possible value), *h* represents the distance between a spatial unit and the nucleus of the cluster, and *a* is called the range of influence (the value of *h* for which the units are practically independent). The variogram is a monotonically increasing function of the distance *h* between a spatial unit and the nucleus of the cluster. It describes the nature of spatial correlation. The factor $P_{i,t}^{(k)}$ represents the population of the i^{th} spatial unit of the k^{th} cluster at time *t*. The factor β represents the spatial effect. The term $\delta_i^{(k)}$ is an error term with mean 0 and constant variance, it represents the perturbation of $\lambda^k(h)$.

In evaluating the model's prediction performance, the mean absolute percentage error (MAPE), whose formula is given below, is computed. The values $\hat{Y}_{it}^{(k)}$ are the postulated model's predicted values of $Y_{it}^{(k)}$. The MAPE is interpreted as the percentage by which the forecasts differ from the actual values on the average.

$$MAPE = \frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{t=1}^{T}\left|\frac{Y_{it}^{(k)} - \hat{Y}_{it}^{(k)}}{Y_{it}^{(k)}}\right|}{T\sum_{k=1}^{K} n_k} \times 100\%$$

Each case is replicated and the MAPE is computed for each replication.

Table 1. Simulation Settings

Parameters	Settings
Cluster size	$n_k = 10$
	$n_k = 30$
	$n_k = 50$
Number of clusters	K = 3
	K = 10

Time series length	T = 10
	T = 20
	T = 30
Stationarity	Φ = 0.1 β = 0.9
	Φ = 0.5 β = 0.5
	Φ = 0.9 β = 0.1
Misspecification error	m = 1
	m = 4

The scenario where $K = 3$ pertains to the case of having few clusters, and $K = 10$ is the scenario of having many. The quantity m is a constant multiplied to the error terms ε_{it} and $\delta_i^{(k)}$. The perturbation scenario of $m = 1$ looks at the model's performance when the perturbation effects are almost negligible (that is, the model is correctly specified), and the scenario of $m = 4$ is the case where the perturbations are substantial. To examine the model's performance for varying cluster sizes, the number of spatial units within a cluster is set to range from 10 to 50. The time series length T is set to range from 10 to 30 to examine the effects of varying time series lengths. Lastly, to be able to study the effects of model stationarity and nonstationarity, ∅ is set at different levels. Moreover, when ∅ = 0.90, it is the temporal term which is dominating, while for ∅ = 0.10, the spatial term dominates.

4. RESULTS AND DISCUSSION

This section looks at the differences in prediction performance of the estimation procedure across the different scenarios.

4.1 Cluster Size

Three cluster sizes are used in this study: small clusters (having 10 spatial units in each cluster), medium-sized clusters (30 spatial units) and large clusters (50 spatial units). The median of all MAPEs for each of the three cluster sizes is presented in Table 2. There is little change in MAPE between cluster sizes of 10 and 30, consistent with (Vera, 2010) observing no change in

MAPE for cluster sizes ranging from 5 to 20. Increasing the cluster size further to 50 spatial units results in a big improvement in the model's predictive ability.

Table 2. Median MAPE for Different Cluster Sizes

Cluster Size	Median MAPE
$n_k = 10$	109.68
$n_k = 30$	112.35
$n_k = 50$	55.68

4.2 Number of Clusters

Differences in the model's prediction performance are also studied for datasets with varying numbers of clusters: small (having 3 clusters) and large (having 10 clusters). Predictive ability of the model is slightly better when there are fewer clusters.

4.3 Length of Time Series

Three time series lengths are considered in this study, comparing the model's prediction performance when there are 10, 20, and 30 time points for each spatial unit. The median MAPE for each of these time series lengths can be seen in Table 3 below. The table appears to show that the MAPE is best for longer time series, but as will be shown later, the MAPE corresponding to varying time series lengths will also depend largely on the levels of the other parameters.

Table 3. Median MAPE for Varying Time Series Length

Time Series Length	Median MAPE
T = 10	118.95
T = 20	97.51
T = 30	30.29

4.4 Stationarity

Three values of the autoregressive parameter are considered: two values reflecting a stationary time series ($\emptyset = 0.10$ and $\emptyset = 0.50$) and one for a near-nonstationary time series ($\emptyset = 0.90$). The case for $\emptyset = 0.90$ also indicates a dominating temporal effect, while $\emptyset = 0.10$ indicates a dominating spatial effect.

There is no difference in the model's predictive ability between $\emptyset = 0.10$ and $\emptyset = 0.50$. However, the MAPE increases to 120.50 for $\emptyset = 0.90$. The model thus produces accurate predictions for as long as the time series is stationary. This increase in MAPE for a backfitting estimator when $\emptyset$ is near nonstationarity has also been observed by (Veron Cruz, 2012).

Table 4. Median MAPE for Stationarity

Autoregressive Parameter	Median MAPE
$\emptyset = 0.10$	107.96
$\emptyset = 0.50$	107.14
$\emptyset = 0.90$	120.5

4.5 Misspecification

The study also compares the predictive ability of the model when the model is specified correctly and when there is misspecification error. Misspecification error is simulated by inflating the variance of the error terms of the model. In a correctly specified model, the value of m is set to 1, and when misspecification error is present, $m = 4$.

The comparison of MAPEs for correct and incorrect model specification is shown in Table 5. Obviously, when the model is correctly specified, the prediction performance of the estimation procedure is much better. This is illustrated by the enormous difference in median MAPE between the two cases.

Table 5. Median MAPE for Model Specification Settings

Specification	Median MAPE
m = 1	11.51
m = 4	170.59

4.6 Discussion

Cluster Size and Number of Clusters

The median MAPE for each combination of cluster size and number of clusters is given in Table 6. Increasing the number of clusters improves the MAPE only when the cluster size is large. For small to moderat

e cluster sizes, increasing the number of clusters leads to a loss in predictive accuracy, especially when $n_k = 30$, where increasing the number of clusters from 3 to 10 more than doubles the median MAPE. Fewer clusters works best for moderate cluster sizes. When there are many clusters, prediction only improves for large cluster sizes.

Table 6. Median MAPE by Cluster Size and Number of Clusters

		Number of Clusters	
		K = 3	K = 10
Cluster Size	$n_k = 10$	106.98	112.93
	$n_k = 30$	54.27	125.35
	$n_k = 50$	65.39	55.68

Cluster Size and Length of Time Series

The median MAPE for combinations of cluster sizes and time series lengths are given in Table 7. As the length of the time series is increased, the estimation procedure improves in forecast accuracy when the cluster size is small to moderate, but not when the cluster size is large. On the other hand, increasing the cluster size results in an improvement in MAPE, except when the time

series is too long. Later on, it will be shown that the MAPE for these cases depend on whether the model is correctly specified or not. The best predictive performance occurs when T is long or n_k is large for a correctly specified model.

Table 7. Median MAPE by Cluster Size and Time Series Length

		Time Series Length		
		T = 10	T = 20	T = 30
Cluster Size	$n_k = 10$	118.99	113.96	30.29
	$n_k = 30$	124.64	59.13	33.19
	$n_k = 50$	59.53	55.56	61.42

Cluster Size and Stationarity

We present in Table 8 the median MAPE for each combination of cluster size and stationarity conditions of the time series component of the model. We can see that if the size of a cluster is fixed, the prediction worsens as ∅ gets closer to 1 (i.e., as the model approaches nonstationarity). This result is consistent for all three cluster sizes: small, moderate and large. The result is much clearer, however, for large cluster sizes. The values of ∅ for the two stationary cases result in approximately the same prediction performance.

For a fixed value of ∅, predictive ability changes little for small and moderate cluster sizes (roughly the same median MAPE for $n_k = 10$ and $n_k = 30$, for all three values of ∅). The prediction rapidly improves, however, when the cluster size is large, except when we deal with near-nonstationary case, for which the improvement is not as large.

Table 8. Median MAPE by Cluster Size and Stationarity

		Autoregressive Parameter		
		∅ = 0.10	∅ = 0.50	∅ = 0.90
Cluster Size	$n_k = 10$	108.09	104.82	120.50
	$n_k = 30$	109.52	116.98	126.71
	$n_k = 50$	55.30	66.24	92.08

Cluster Size and Misspecification

Shown in Table 9 are the median MAPEs for the different cluster sizes for each of the two specification scenarios. When the cluster size is fixed, a misspecified model clearly results in poor prediction performance. The difference in prediction performance between a correctly and incorrectly specified model increases with the cluster size, so that if $n_k = 50$, the median MAPE for incorrect specification is extremely high at 238.46%.

When the model is specified correctly, increasing the cluster size improves the model's forecasting performance. However, even if the cluster size is small ($n_k = 10$), a correctly specified model still yields reasonable forecasts with 15.86% median MAPE. This suggests that the estimation procedure predicts quite well when there is no misspecification. If, on the other hand, the model is incorrectly specified, then increasing the cluster size only adds to the prediction error.

Table 9. Median MAPE by Cluster Size and Specification

		Specification	
		Correctly Specified	Misspecified
Cluster Size	$n_k = 10$	15.86	135.67
	$n_k = 30$	11.43	178.27
	$n_k = 50$	10.06	238.46

Number of Clusters and Length of Time Series

Table 10 shows the effect on forecasting performance for varying levels of time series length and number of clusters. For fixed time series length, increasing the number of clusters seems to lead to poor prediction performance (it will be shown later that this will depend on the cluster size).

If the number of clusters is fixed, Table 10 shows that the performance improves as the time series gets longer. The optimal condition thus occurs for few clusters and long time series.

Table 10. Median MAPE by Number of Clusters and Length of Time Series

		Number of Clusters	
		K = 3	K = 10
Length of Time Series	T = 10	109.45	125.04
	T = 20	68.97	97.51
	T = 30	27.88	35.22

Number of Clusters and Stationarity

Table 11 now gives the median MAPE for each combination of number of clusters and stationarity. When the number of clusters is fixed, we once again see that the estimation performs poorly in the case of nonstationarity. We also see that the prediction accuracies when $\emptyset = 0.10$ and $\emptyset = 0.50$ do not differ much when the number of clusters is fixed.

When the value of $\emptyset$ is fixed, increasing the number of clusters increases the median MAPE. Thus the prediction performance is best for fewer clusters and for stationary time series.

Table 11. Median MAPE by Number of Clusters and Stationarity

		Autoregressive Parameter		
		$\emptyset = 0.10$	$\emptyset = 0.50$	$\emptyset = 0.90$
Number of Clusters	K = 3	98.85	103.24	114.54
	K = 10	113.21	110.36	138.36

Number of Clusters and Misspecification

In Table 12 below, we compare the forecasting performance for varying numbers of clusters, under the scenarios of correct and incorrect model specification. As expected, a wrongly specified model gives larger MAPE, regardless of the number of clusters. The difference in median MAPE for the misspecified and correctly specified cases is very large.

Assuming that the model is correctly specified, the model's performance is slightly better when there are fewer clusters. In both cases however, the predictive ability is good. This once again affirms the result that if the model is correctly specified, the proposed estimation procedure can predict well.

Table 12. Median MAPE by Number of Clusters and Specification

		Specification	
		Correctly specified	Misspecified
Number of Clusters	K = 3	10.76	179.14
	K = 10	12.64	162.00

Length of Time Series and Stationarity

The median MAPE for combinations of time series length and stationarity settings are given in Table 13. If the number of time points is fixed, prediction performance is once again poorest when $\emptyset = 0.90$, the case of nonstationarity. This is true for all time series settings, but the differences in prediction performance among the three stationarity settings is small when *T* is short or moderate. On the other hand, if we fix the value of $\emptyset$, prediction improves as the time series is lengthened. Thus, the best condition occurs for long time series and autoregressive parameter $\emptyset$.

Table 13. Median MAPE by Time Series Length and Stationarity

		Autoregressive Parameter		
		$\emptyset = 0.10$	$\emptyset = 0.50$	$\emptyset = 0.90$
Time Series Length	T = 10	113.10	121.72	124.86
	T = 20	97.51	97.37	99.74
	T = 30	27.40	28.65	69.17

Length of Time Series and Misspecification

In Table 14, we compare the median MAPE for the different time series lengths, and for the specification scenarios. Once again, we see that a wrongly specified model gives a much higher

MAPE than a correctly specified one, regardless of the time series length. The difference in MAPE between the two specification settings is larger for moderate and long time series.

Assuming that the model is correctly specified, we see that a short time series generally gives poor predictions (median MAPE of 48.89%). Good predictions occur for moderate and long time series, with little difference between the two. For a wrongly specified model, increasing the time series length will only worsen the predictions. The ideal case thus occurs when the model is correctly specified and the time series is moderate to long.

Table 14. Median MAPE by Length of Time Series and Specification

		Specification	
		Correctly specified	Misspecified
Number of Clusters	T = 10	48.89	134.91
	T = 20	10.15	214.52
	T = 30	10.27	208.30

Stationarity and Misspecification

Below, Table 15 gives the prediction performance for the different stationarity settings, for both the correctly specified and wrongly specified scenarios. As has been seen several times, correctly specified models give much better predictive performance than wrongly specified ones, with the largest disparity occurring when the model is nearly nonstationary.

If the model is correctly specified, the forecasts are generally good. The least predictive ability is seen when $\emptyset = 0.90$, with a median MAPE of 13.80%, which is still reasonably good. On the other hand, if the model is wrongly specified, increasing the value of $\emptyset$ worsens the prediction, with a hopelessly high MAPE of 307.83% under the nonstationary case.

Table 15. Median MAPE by Stationarity and Specification

		Specification	
		Correctly specified	Misspecified
Autoregressive Parameter	$\emptyset = 0.10$	10.07	137.07
	$\emptyset = 0.50$	9.63	161.89
	$\emptyset = 0.90$	13.80	307.83

We have seen that a correctly specified model always does better in terms of prediction than a misspecified one. Moreover, a stationary value of $\emptyset$ (it does not seem to matter whether $\emptyset = 0.10$ or $\emptyset = 0.50$) results in better prediction. Third, the prediction generally improves as the time series gets longer. Fourth, the prediction is usually better when there are few clusters, except when the cluster size is large. At this point, it is hard to make a conclusion about whether the cluster size should be large or small, since its effect is inextricably intertwined with that of the other factors.

5. Conclusions

We proposed a model that captures the isotonic relationship between the response and the spatial distance between observations in a cluster. The proposed estimation procedure has high predictive accuracy when the model is correctly specified and when the time series component is exhibit a stationary behavior. Higher predictive ability can be observed when there are few clusters and when the observed time series data is fairly long. When the time series length is at least moderate, predictive ability is high when number of elements in a cluster is larger or equal to the number of time points.

REFERENCES

Buja, A., T. Hastie, R. Tibshirani (1989). "Linear Smoothers and Additive Models." The Annals of Statistics, 17(2):453-510.

De Boor, C. (1978). A Practical Guide to Splines. New York: Springer-Verlag.

De Vera, E., (2014). "Semiparametric Poisson Regression Model for Clustered Data", The Philippine Statistician, 63(1): 33-42.

Dodds, P.S. and D.J. Watts (2004*). "*Universal Behavior in a Generalized Model of Contagion." Physical Review Letters, 92(21): 218701-1 to 218701-4.

Hastie, T.J. and R. J. Tibshirani (1990). Generalized Additive Models. London: Chapman & Hall.

Konishi, S. and Kitagawa, G. (2008). Information Criteria and Statistical Modeling. New York: Springer Science+Business Media.

Manly, B.F.J. (2001). Statistics for Environmental Science and Management. Boca Raton: Chapman & Hall, CRC.

Santos, L. and Barrios, E. (2011). "Small Sample Estimation in Dynamic Panel Data Models: A Simulation Study." American Open Journal of Statistics, **1**, 58-73.

Veron Cruz, R., and Barrios, E. (2014). "Estimation Procedure for a Multiple Time Series Model." Communication is Statistics, Simulation and Computing, 43(10): 2415-2431.

Winsberg, S. and Ramsay, J. O. (1980). "Monotonic Transformations to Additivity Using Splines." Biometrika, **67**, 3, pp. 669-74.