SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN

WORKING PAPER SERIES

Testing for the Presence of Structural Change and Spatial Heterogeneity in a Spatio-Temporal Model

by

Ruby Anne E. Lemence
and
Erniel B. Barrios

UPSS Working Paper No. 2009-05
June 2009

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

Testing for the Presence of Structural Change and Spatial Heterogeneity in a Spatio-Temporal Model

by

Ruby Anne E. Lemence[a] and Erniel B. Barrios[b1]
[a]*Graduate Student, School of Statistics, University of the Philippines Diliman, Quezon City, Philippines*
[b]*Professor, School of Statistics, University of the Philippines Diliman, Quezon City, Philippines*

Abstract

In a spatio-temporal model, structural change and/or spatial heterogeneity can easily affect estimation of parameters. Alternative estimation procedures that are relatively robust can be considered in cases where such deviations from model specifications are suspected to be present. Following the spatio-temporal model proposed by Landagan and Barrios (2007), we develop a nonparametric procedure for testing the presence of structural change and spatial heterogeneity using bootstrap techniques and the forward search algorithm. The time series bootstrap can filter the effect of temporary structural change in the construction and a confidence interval for the temporal parameter. The forward search will also facilitate the construction of a robust confidence interval for the spatial parameter. The confidence intervals are then used in deciding whether to reject the null hypothesis that there is no structural change/spatial heterogeneity or not. Using simulation studies, the proposed test procedure is capable of detecting presence of structural change and spatial heterogeneity under certain conditions.

Keywords: bootstrap, forward search, spatio-temporal model, nonparametric test
AMS Codes: 62F40, 11N36, 62F35, 62G10

---

[1] Correspondence: Erniel B. Barrios, School of Statistics Bldg., Magsaysay Ave., University of the Philippines, Diliman, Quezon City, 1101, Philippines, email: ernielb@yahoo.com, ebbarrios@up.edu.ph

1. Introduction

Technological advances have led to tremendous improvement in data collection techniques allowing for data referenced both in space and in time. This necessitates models that take into account the dependence structure of data across space and over time, this give rise to spatio-temporal models.

Growing interests on modeling spatio-temporal systems in recent years is attributed to its emerging applications in the environmental and health sciences, which include monitoring of regional ozone levels, disease mapping, and analysis of satellite data, to name a few (Stroud et al., 2001). Spatio-temporal models are useful in monitoring and evaluation of policies, projects, interventions, etc., and the information it generates are useful tools in increasing public awareness as well as facilitating decision making (Landagan and Barrios, 2007). According to Stroud et al. (2001), another reason for the recent surge of interest in this area is the increased computational power as “space-time data sets are often large and therefore require substantial computing resources to fit even simple models.”

In many field of applications, it is not unlikely that we encounter data with atypical observations or observations that seem to have a different structure compared with the rest. For time series data, unexpected events may lead to changes in the parameters and these are exhibited by breaks in the series, for example, a shift in the mean of the series. Meanwhile, for data collected across space, some spatial units may have varying characteristics from the rest even if they are exposed to the same environmental settings.

In modeling, it is important that we take into account the observed changes in structure since failure to do so may lead to biased estimates of parameters. It is imperative that we still are able to estimate the process in the presence of such aberrations in the data. Several procedures that enable robust estimation of the process even if there are atypical observations in the data have already been proposed and these include the forward search algorithm and bootstrap techniques.

The forward search algorithm is a powerful method used in detecting atypical observations by starting with an initial outlier-free subset of the data, then moving on to a larger subset of observations until all of the observations were utilized. Meanwhile, bootstrapping is a computationally intensive method for making inference about the population that involves estimating the empirical distribution function (EDF) thru resampling from the data.

In this paper, the forward search algorithm is proposed in detecting spatial heterogeneity in spatio-temporal data using the additive model proposed by Landagan and Barrios (2007). On the other hand, we propose to test for structural change using bootstrap procedures, particularly the AR-sieve bootstrap for time series and the nonparametric bootstrap for independent observations.

Deviations from the model usually occurs quite frequently whether in time or in space. However, most of the existing tests for structural change are based on asymptotic results. In a spatio-temporal setting, however, there is difficulty in coming up with large datasets that satisfy the requirements of the asymptotic procedures. In some cases, we may encounter data collected over a long period of time but only for a few spatial units or data collected for a short period of time but for a relatively large number of spatial units, thus, a need for tests that can address short-term data or small spatial samples. The bootstrap methods and the forward search algorithm are locally efficient methods; hence, they are expected to work well in testing the presence of structural change and spatial heterogeneity in a spatio-temporal data.

A structural change is defined as the change in the value of the temporal parameter. A structural change occurs if the value of the temporal parameter changed at some time points, then return to the previous values afterwards. On the other hand, spatial heterogeneity is present if for some spatial units, the value of the spatial parameter is different than the rest. These changes in the parameter values are expected to be temporary, after which, it will reverse to its original values.

For structural change, we consider a situation where the temporal effect, ρ, is different for some time periods compared with the rest, while assuming constant covariate effects across locations and time. Furthermore, we assume that the spatial effect is constant over time. Meanwhile, for spatial heterogeneity, we take into account a scenario where the spatial effect is different for some spatial units compared with the rest assuming no change in the covariate and temporal effects across space and time.

2. Spatio-temporal Models

Spatio-temporal models take into account space and time dependence. The need for models that take into account the dependence across time and space was fueled by the enormous amount of data gathered as an outcome of improved data collection process spurred by recent advances in science and technology. Hering et al. (2006) and Waller et al. (2007) provided examples of the applications of spatio-temporal modeling in their studies of the spatio-temporal epidemiology of foot-and-mouth disease, spatio-temporal wildfire ignition point patterns, and country-level incidence and reporting of Lyme disease, respectively.

Landagan and Barrios (2007) proposed an iterative procedure for estimating spatio-temporal models that incorporates the Cochranne-Orcutt procedure into the backfitting algorithm. They considered an additive model for yield of cereals, given by $Y_{it} = X_{it}\beta + W_{it}\gamma + \varepsilon_{it} \quad i = 1, 2, ..., N \quad t = 1, 2, ..., T$ as a function of X_{it}, the set of covariates from location i at time t, w_{it}, the set of spatial indicators for location i at time t, and an error term, $\varepsilon_{it} = \mu_i + \upsilon_{it}$. The individual effects, μ_i, are independently and identically distributed with zero mean and constant variance, while the remainder of the error, υ_{it}, is an autoregressive process of order p. It was shown that the proposed hybrid of the backfitting algorithm and the Cochranne-Orcutt procedure is efficient since the iterative parameter estimation usually converges after second or third iterations. Dumanjug et al. (2009) proposed an alternative estimation procedure for the spatio-temporal model considered in Landagan and Barrios (2007), incorporating the block bootstrap into the hybrid of the backfitting algorithm and the Cochranne-Orcutt procedure. Dumanjug et al. (2009)

introduced two bootstrap procedures, one, using "independent" block bootstraps and another, using overlapping blocks of consecutive vectors.

3. Bootstrap in Time Series Data

Bootstrapping, as proposed by Efron in his 1979 paper, was originally applied to models that assume independent observations. The application of the bootstrap has evolved to include models of dependent data such as in time series. Politis (2003) noted that in the 1990's, the framework of the bootstrap methods moved from the independent observations towards dependent data (e.g., time series, random fields). Resampling in time series data is different from the procedure for the independent data. Bühlmann (2002) observed that the construction of the empirical distribution in time series data is more complicated than in the independent observations proposed by Efron in 1979. According to Bühlmann (1997), the bootstrap has indeed become a powerful nonparametric method for estimation; however, the procedure usually fails in the case of dependent observations when the order of the observations is disregarded. When resampling time series data, it is very important that the dependence structure of the data is preserved.

Several bootstrap procedures for time series are available. Bühlmann (2002) provided a review of the different bootstrap methods proposed for time series. For stationary time series, the common approaches are the model-based approach and the blockwise bootstrap. The model-based approach resamples from i.i.d. residuals. Since the resampling involves an assumed model, these procedures are prone to model misspecification, aside from losing the nonparametric property of the bootstrap. In contrast, the blockwise bootstrap, so called by construction, is nonparametric and purely model-free. In the blockwise or moving blocks bootstrap approach, the series is divided into blocks and then blocks of observations, instead of individual observations are resampled. In general, the blockwise bootstrap is robust against misspecified models but by joining randomly selected blocks, it was found that the resampled series exhibit artifacts. Bühlmann (2002) cautioned that the bootstrap samples may loss conditional stationarity once the block dependence is ignored. To resolve this problem, Politis and Romano (1994) modified the blockwise bootstrap that yields a (conditionally) stationary bootstrap sample.

Bühlmann (2002) proposed another bootstrap procedure for time series called the sieve bootstrap. The procedure involves fitting parametric models first and then resampling the residuals of the estimated model. Bühlmann (2002) outlines the steps in constructing the sieve bootstrap sample $X_1^*,...,X_n^*$ as follows:

1. Choose starting values, e.g., equal to zero.
2. Generate an AR*(p(n))* process according to $\sum_{j=0}^{p(n)} \hat{\phi}_{j,n}(X_{t-j}^* - \bar{X}) = \varepsilon_t^*$ until 'stationarity' is reached and then throw the first generated values away.

The bootstrapped statistics T_n^* are defined by $T_n^* = T_n(X_1^*,...,X_n^*)$, where $T_n = T_n(X_1,...,X_n)$ is any statistic and T_n is a measurable function of *n* observations.

The application of bootstrap procedures is not limited to construction of confidence intervals alone it is also useful in testing hypotheses. Westerlund and Edgerton (2007) proposed a bootstrap procedure for testing cointegration in panel data while Paparoditis and Politis (2005) discussed bootstrap hypothesis testing in regression models. The use of bootstrap procedures were also explored in unit root testing in Paparoditis and Politis (2003), Park (2003), and Ioannidis (2005). Meanwhile, Singh and Berk (1994) introduced and developed a type-2 *p*-value for the bootstrap.

4. Forward Search Algorithm

The forward search algorithm was proposed to detect atypical observations in the data. (Atkinson and Riani, 2007) observed that the method is powerful in detecting multiple outliers and for detecting their effect on inferences about models fitted to the data. Initially applicable to models that assume independent observations, the applications of the method has been extended to models that assume dependent observations, for example, time series data.

The forward search algorithm is composed of three steps: (1) choice of initial subset; (2) progressing in the search, and; (3) diagnostic monitoring. The forward search algorithm starts with fitting the model to an ideal or outlier-free subset of the data then the search progresses by fitting the model to a larger subset by successively adding one observation in the estimation. During the search, the quantities that indicate the quality of the model and its adequacy, aside from the values of parameter estimates, are monitored for stability. If there are no outliers, the plot of the parameter estimates and the residuals remain stable as the size of the subset grows (Atkinson and Riani, 2002).

The forward search algorithm is a general method and it can be applied in many statistical models but it is more popularly used in regression models. The use of forward search in building a regression model is illustrated in Atkinson and Riani (2007). Meanwhile, Atkinson and Riani (2002) proposed a *t*-test suitable for the forward search, which they called the "added-variable *t*-test". Mavridis and Moustaki (2007) implemented the forward search algorithm for identifying outliers and extreme response patterns in latent variable models.

When implementing the forward search in regression models, the fitting of least squares is performed each time the subset is increased until the number of observations used is equal to the total number of observations while storing the coefficients, residuals and other diagnostic statistics. To avoid refitting the least squares linear model to all the units in each step of the search, Konis and Laurini (2007) proposed an algorithm where the fit is updated by considering only the change in units between the mth and (m+1)th subsets.

5. Nonparametric Testing for Structural Change and Spatial Heterogeneity

The spatio-temporal model proposed by Landagan and Barrios (2007) is given by the following:

$$Y_{it} = X_{it}\beta + \omega_{it}\delta + \varepsilon_{it}\,, \quad i = 1, 2, ..., N \qquad t = 1, 2, ..., T \qquad (1)$$

where Y_{it} is the response variable from location *i* at time t , X_{it} is a set of covariates from location *i* at time *t*, ω_{it} is a set of variables in the neighborhood system from location *i* at time t , and ε_{it} is the error component. The error component is assumed to be autocorrelated over time (temporal dependence), specifically, autoregressive process of order *1*, given by,

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + a_{it}\,,\; |\,\rho\,| < 1, \;\; a_{it} \sim IID(0, \sigma_a^2)\;.$$

Further, the above model assumes the following:

(i) constant covariate effect *(β)* across locations and time

(ii) constant temporal effect *(ρ)* across locations

(iii) constant spatial effect *(δ)* across time

This study takes into account the following cases of deviations from the spatio-temporal model above: the temporal effect, *ρ,* changed at specific time points and is referred to as the occurrence of structural change; and the spatial effect, δ, is different for some spatial units and is referred to as the presence of spatial heterogeneity.

Modifying the model in (1), the spatio-temporal model with structural change and spatial heterogeneity is given by the following:

$$Y_{it} = X_{it}\beta + \omega_{it}\delta_h + \varepsilon_{it}\,,\ \varepsilon_{it} = \rho_c \varepsilon_{it-1} + a_{it}\,,\ |\rho|<1,\ \ a_{it} \sim IID(0,\sigma_a^2) \qquad (2)$$

$$where\, \delta_h = \delta I(i)_{\{i \notin N^h\}} + \delta^{'} I(i)_{\{i \in N^h\}}$$

$$\rho_c = \rho I(t)_{\{t \notin T^c\}} + \rho^{'} I(t)_{\{t \in T^c\}}$$

$i = 1,2,...,N\ and\ t = 1,2,...,T$ and where $\delta\ and\ \rho$ are the original parameter values and $\delta^{'}\ and\ \rho^{'}$ are the new, temporary parameters values due to spatial heterogeneity and structural change, respectively. In addition, N^h is the set of spatial units affected by heterogeneity, and T^c is the set of time points where the temporary structural change occurred.

For each case, the proposed test methodology is composed of two phases: (i) estimation of model parameters and (ii) testing the null hypothesis of no structural change or no spatial heterogeneity. The AR-sieve bootstrap is used to replicate the process that generated the time series for each spatial unit. The purpose of the replication is to obtain as many estimates of the temporal parameter for subsequent use in the test. Meanwhile, the forward search algorithm is used in obtaining robust estimates of the spatial parameters. In the presence of spatial heterogeneity, the spatial parameter can be easily contaminated hence, some robust estimates are necessary to filter the effect of such heterogeneity. The sampling distribution of the statistics to be tested will be estimated using the bootstrap procedure. The estimated sampling distribution is then used to construct the 100*(1-α)% confidence interval for the parameter of interest.

5.1. Testing for Structural Change

A structural change occurs in time if for some points, there is a considerable change in the values of the model parameters. The null hypothesis of no structural change will be tested against the alternative that a temporary structural change occurs at some unknown time point using bootstrap methods.

Phase I. Estimation Algorithm using the AR-sieve Bootstrap

For each spatial unit, *i*, *i = 1, …, N*,

T.1. Estimate the parameters of the model, $Y_t = X_t\beta + \omega_t\delta + \varepsilon_t$ using ordinary least squares. Given the estimates, $\hat{\beta}$ *and* $\hat{\delta}$, compute the residuals $\{e_t\}$.

T.2. Fit an ARIMA model on $\{e_t\}$, the resulting parameter estimate is denoted by $\hat{\rho}$.

T.3. Given the $\hat{\rho}$ obtained in (T.2), simulate $\{e_t\}$ with $\{e_t^*\}$ by generating $\{a_t\}$,

where $a_t \sim N(0,\sigma_a^2)$ and σ_a^2 is equal to the MSE of the residuals in (T.2). Do this *m* times, yielding *m* sets of $\{e_t^*\}$, $e = \{\{e_t^{*k}, t = 1,..,T\}, k = 1,...,m\}$.

T.4. For each of the *m* simulated residuals in (T.3), generate pseudo-observations, $\{\{Y_t^{*1}\},\{Y_t^{*2}\},...,\{Y_t^{*m}\}\}$, where $Y_t^{*k} = X_t\hat{\beta} + \omega_t\hat{\delta} + e_t^{*k}$, $k = 1,2,...,m$.

T.5. For each of $\{Y_t^{*k}\}, k = 1,2,...,m$, estimate the model $Y_t^{*k} = X_t\beta^{(k)} + \omega_t\delta^{(k)} + \varepsilon_t$, compute new residuals and fit an ARIMA model for the new residuals. Store the new parameter estimates.

Phase II. Bootstrap Test for Structural Change

After exhausting all the *N* spatial units, we now have *(N x m)* estimates for $\hat{\rho}$. Ordinary bootstrap will then be performed using these *(N x m)* estimates to construct a confidence interval for ρ . These estimates of ρ are free from the contamination that may be caused by the structural change.

The following are the steps of the bootstrap procedure to be carried out:

(i) From the *(N x m)* estimates for ρ above, draw a simple random sample of size *n* with replacement.

(ii) For the sample obtained in *(i)*, compute and store the mean $\hat{\rho}^{(b)} = \frac{1}{n}\sum_{k=1}^{n}\hat{\rho}_k$ and the median $\hat{\rho}^{(b)}_{med}$.

(iii) Repeat Steps *(i)* and *(ii)* *B* number of times, where *B* is a large number.

(iv) Compute the bootstrap estimate for both the mean and the median using the formula, $\hat{\rho}_{BS} = \frac{1}{B}\sum_{j=1}^{B}\hat{\rho}^{(j)}$, Monte Carlo variance $\hat{\sigma}^2_{BS} = \frac{1}{B-1}\sum_{j}^{B}(\hat{\rho}^{(j)} - \hat{\rho}_{BS})^2$, and 100(1-α)% bootstrap confidence interval (BCI).

Conclude that there is no structural change if less than 100α% of the *(N*m)* $\hat{\rho}$'s, fall outside the computed 100(1-α)% BCI for ρ. Otherwise, we reject the null hypothesis and conclude that a temporary structural change occurred.

The AR-sieve bootstrap will ensure that the confidence interval constructed for ρ is not contaminated by the structural change, if there's any. Hence, when the estimates of ρ for each spatial unit are compared to this interval, presence of structural change will force some values to be outside the interval.

5.2. Testing for Spatial Heterogeneity

When there is spatial heterogeneity, we will assume that the spatial parameter δ is different for some spatial units assuming the covariate effects remain the same. The spatial unit manifesting the change is also assumed to be unknown.

The proposed procedure for testing spatial heterogeneity is also divided into two phases: estimation and detection. In the estimation phase, two types of estimates will be obtained: using the forward search algorithm and; using all the observations at once. The forward search algorithm is expected to generate robust estimates of δ even when spatial heterogeneity is present. These "forward searched" estimates will then be bootstrapped to construct a sampling distribution for $\hat{\delta}$. Meanwhile, the estimates obtained using the full sample allows us to estimate the actual process that generated the data at hand.

The 100*(1-α)% confidence interval for δ will be obtained using the bootstrapped sampling distribution of $\hat{\delta}$. The estimates from the full sample will be compared with the constructed confidence interval. Estimates found to be outside the constructed confidence interval provide evidence that there is spatial heterogeneity.

Phase I. Estimation Using the Forward Search Algorithm

The following steps will be performed for each time point, *t*, *t = 1,…, T*.

S.1. From the *N* observations, choose a subset of size *n*, $n<N$, such that it is outlier-free, thus, "ideal enough" to represent the *N* locations. This is done by fitting the model $Y_{it} = X_{it}\beta + \omega_{it}\delta + \varepsilon_{it}$ to the full dataset and then choosing the *n* observations corresponding to the *n* smallest residuals.

S.2. Fit the model in (S.1) to the selected *n* observations.

S.3. Use the parameter estimates obtained in (S.2) to compute for the fitted Y_{it}, $\hat{Y}_{it}$, for all $i = 1, 2, ..., N$ and obtain the residuals $e_{it} = Y_{it} - X_{it}\hat{\beta} - \omega_{it}\hat{\delta}$.

S.4. Select the *(n+1)* observations that correspond to the *(n+1)* smallest residuals in (S.3).

S.5. Fit the model in (S.2) to the *(n+1)* observations selected in (S.4).

S.6. Repeat (S.2) to (S.5) adding one location at a time until all the N locations are included in the estimation.

In Atkinson and Riani (2007), the size of the initial subset was set to be equal to the number of parameters in the regression model considered. In the simulation studies considered here, however, the size of the initial subset was set to be equal to N/2. This is done because when using very few observations in the first few runs, it was observed that the values of the Cook's *D* statistic exhibit relatively high fluctuations which is not necessarily due to the addition of influential observations but may also be because of the adjustments in the values of the parameter estimates as the sample size is increased from only just a few observations.

The search progresses by moving on to a larger subset of observations, adding one observation at a time. At each run, the model in (S.1) is fitted to the subset and then the estimates obtained are used to compute for the residuals using the full sample to determine the candidate observations to be included in the next run. We stop the search when the estimates being obtained are already "behaving wildly", that is, when large changes are observed in the values of Cook's *D* statistic.

As the search progresses, the values of the Cook's *D* statistic are computed and observed for each run. A rule of thumb in using the Cook's *D* statistic is to declare an observation influential if its value exceeds *4/n*, where *n* is the number of observations. In this study, instead of comparing each of the computed Cook's *D* statistic to *4/n*, the change in the maximum value of the Cook's *D* statistic in the succeeding run is considered. This is done by getting the absolute difference of the maximum Cook's *D* statistics from the previous run and the present run. The tolerance value for the absolute difference of the maximum of the Cook's *D* statistics considered here is 0.03, determined after observing how the corresponding parameter estimates behave with changes in the maximum of the Cook's *D* statistics. The search stops when the maximum of the Cook's *D* statistic in the present run differs from that of the previous run by at most 0.03, then we get the estimates from the previous run as the robust or the "forward searched" estimate.

Phase II. Bootstrap Test for Spatial Heterogeneity

From Phase I, we collect *T* estimates of the spatial parameter δ, given by $\underline{\hat{\delta}} = \left(\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_T\right)$. To test the null hypothesis that there is no spatial heterogeneity, we follow similar bootstrap procedure as in testing for structural change. The following are the steps to be carried out.

(i) From the *T* estimates for δ obtained using the forward search algorithm, draw a simple random sample of size *n* with replacement.

(ii) For the sample in (i), compute and store the mean and median of $\hat{\delta}$.

(iii) Repeat (i) and (ii) *B* number of times, where *B* is a large number.

(iv) Compute the bootstrap estimate for both the mean and the median using the formula,

$\hat{\delta}_{BS} = \frac{1}{B}\sum_{j=1}^{B}\hat{\delta}^{(j)}$, Monte Carlo variance $\hat{\sigma}_{BS}^2 = \frac{1}{B-1}\sum_{j}^{B}(\hat{\delta}^{(j)} - \hat{\delta}_{BS})^2$, and 100(1-α)% bootstrap confidence interval (BCI) for δ.

Conclude that there is spatial heterogeneity if more than 100α% of the $T\hat{\delta}$'s, computed using the full sample, fall outside the 100(1-α)% BCI for δ. Otherwise, we do not reject the null hypothesis and conclude that there is no spatial heterogeneity. The forward searched estimates of δ are expected to be free from the effect of spatial heterogeneity. Thus, the robustness of the resulting bootstrap confidence interval will make a reasonable benchmark in deciding which hypothesis to favor that best reflects the data.

6. Simulation Study

The performance of the proposed procedures was assessed by using the procedures to simulated data with no structural change or spatial heterogeneity and to data with embedded structural change or spatial heterogeneity. The specificity of the tests is determined by applying the proposed procedures to data with no structural change and data with no spatial heterogeneity. The test is said to be specific if the null hypotheses of no temporary structural change or spatial heterogeneity are not rejected when there are none. In other words, if there is really no structural change or spatial heterogeneity (as simulated), we want the percentages of $\hat{\rho}'s\ and\ \hat{\delta}'s$ covered in the corresponding 100(1-α)% BCI's to be greater than 100(1-α)%.

On the other hand, the sensitivity of the procedures is established if the null hypotheses of no structural change and no spatial heterogeneity are rejected when there is really temporary structural change or spatial heterogeneity in the simulated data. In terms of the percentages of $\hat{\rho}'s\, and\, \hat{\delta}'s$ covered in their respective 100(1-α)% BCI's, the test is sensitive if less than 100(1-α)% of the *(Nxm)* $\hat{\rho}'s$ or $T\ \hat{\delta}'s$ are covered.

The response variable ***Y*** was computed using the following model:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \delta W_{it} + \varepsilon_{it}$$
$$\varepsilon_{it} = \rho\varepsilon_{it-1} + a_{it} \text{ where } a_{it} \sim N(0,4)$$

where $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ were each sampled from the Normal population with means 100 and 50, respectively, and variances which are both equal to 100. To introduce spatial dependence in the data, the total number of spatial units was divided into four clusters. This was done by obtaining samples for the neighborhood variable ***W*** by sampling from the Poisson distribution with means $\lambda_k, \lambda_k \in \{2,4,6,10\}$. Each λ_k corresponds to one neighborhood or cluster. Meanwhile, the error term was simulated from the AR(1) process, $\varepsilon_{it} = \rho * \varepsilon_{it-1} + a_{it}$, with ρ equal to 0.5, and a_{it} was sampled from the Normal distribution with zero mean and variance equal to 4. The values of the coefficients used are as follows: $\beta_0 = 40.00; \beta_1 = 0.70; \beta_2 = 0.45;\ and, \delta = 0.25.$ To induce structural change and spatial heterogeneity in the space-time data, $\rho = 0.75\, and\, \delta = 1.25$ were considered as the temporary values.

The choice of the population where in the covariates and neighborhood variable are selected is arbitrary and will not affect the conduct of the procedure. In addition, we considered simulating datasets for different values of the coefficient of determination, R^2. The values for R^2 considered were in the vicinity of 20%, 50%, and 95%, to represent low, moderately high, and high R^2. From the dataset with high R^2, the datasets with moderately high and low R^2 were constructed by simply multiplying a constant to the error terms.

6.1. Simulating Structural Change

Consider a space-time dataset given by $\underline{Y} = \{Y_{it} : i = 1, 2, ..., N, t = 1, 2, ..., T\}$. In matrix notation, $\underline{Y} = \begin{pmatrix} y_{11} & y_{21} & \cdots & y_{N1} \\ y_{12} & y_{22} & \cdots & y_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{NT} \end{pmatrix}$ and note that $\underline{Y}$ is composed of *N (Tx1)* vectors, $\underset{\sim}{y_i} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}$, $i = 1, 2, ..., N$. Three cases of structural change were considered depending on the part of the series where it occurred: start, middle, or end of the time series. Furthermore, it is assumed that if a temporary structural change occurs, all of the *N* spatial units are affected and that the occurrence of structural change is the same across space. That is, if a structural change was observed at the start of the series for spatial unit *i*, that for spatial unit *j* $(i \neq j;\ i, j = 1, ..., N)$, should also be observed at the start of the series. Note that the preceding assumption can be observed in real life situations. For instance, suppose we are interested in modeling crop yield for each province over time, a temporary structural change in this case may have been caused by a natural calamity that has impact on the yield of all provinces, for example, an El Niño episode.

Structural change was introduced into the data by simply replacing the error terms for the selected part of the series with no structural change with the error terms that were sampled from an autoregressive process of the same order but with different parameter value, in this case, $\rho = 0.75$. Furthermore, since the structural change to be considered is temporary, only 5% to 15% of the given number of time points was allowed to be affected. Also, to determine the effect of sample size, we considered 40, 50, and 75 time points to represent small, medium and large sample sizes. The number of spatial units was also varied. Thus, the performance of the proposed procedure for testing temporary structural change was evaluated for each combination of *T*, the number of time points, and *N*, the number of spatial units.

For the AR-sieve bootstrap, 100 replicates were obtained for each spatial unit, *i*, so that the number of replicated $\hat{\rho}$'s equals 100 times the specified number of spatial units. Meanwhile, 1000 resamples with replacement were considered for the ordinary bootstrap used in the construction of the 100(1-α)% bootstrap confidence interval and the size of each resample is equal to 20% of the total number of available replicates.

Because similar results were obtained for the tests using datasets with low and high R^2, the proposed procedure for testing the presence of temporary structural change was no longer tried on a dataset with 50% R^2.

6.2. Simulating Spatial Heterogeneity

Consider the space-time data set given by $\underline{Y} = \{Y_{it} : i = 1, 2, ..., N, t = 1, 2, ..., T\}$, whose matrix notation is also given in Section 6.1. $\underline{Y}$ is composed of *T (Nx1)* vectors, $\underset{\sim}{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}$, $t = 1, 2, ..., T$. Spatial dependence was introduced by dividing the data collected across space into four neighborhoods given by $\underline{Y} = (\underline{Y}1, \underline{Y}2, \underline{Y}3, \underline{Y}4)$ where $\underset{\sim}{Y}1 \sim Poisson(2)$, $\underset{\sim}{Y}2 \sim Poisson(4)$, $\underset{\sim}{Y}3 \sim Poisson(6)$, $\underset{\sim}{Y}4 \sim Poisson(10)$. That is, a neighborhood is defined be a group of spatial units that share some common characteristics in space.

Four scenarios of occurrence of spatial heterogeneity will be considered depending on the number of neighborhoods affected. This is done in order to see if the procedure is sensitive to the number of affected neighborhoods. Similar to what was done in the case for structural change, the procedure for testing spatial heterogeneity is assessed for different proportions of affected spatial units. The proportions of spatial units exhibiting spatial heterogeneity considered were 5%, 10%, and 15%. The number of affected spatial units was distributed in one, two, three, or all four neighborhoods.

Spatial heterogeneity was introduced for each set of spatial units, collected over time, by changing the value of the spatial parameter in the selected spatial units from 0.25 to 1.25. The effect of sample size was investigated by considering 20, 40, and 60 as the number of spatial units available. In addition, the number of time points was also allowed to vary. In the conduct of the ordinary bootstrap, 1000 bootstrap samples of size T were considered.

6.3. Constructing the 100(1-α) % Bootstrap Confidence Interval

In the construction of the 100(1-α)% bootstrap confidence interval (BCI), we considered constructing a 95% BCI for each of the temporal and spatial parameters. The mean and the median of estimates were computed for each resample. Two sets of 95% BCIs were then constructed one, based on the mean and another, based on the median. Since the estimates obtained were skewed, the median seemed to be more appropriate to be used. In addition, in the presence of structural change, it was observed that tests based on the mean were not performing well because of the sensitivity of the mean to extreme observations. On the other hand, the tests based on the median were observed to be consistent.

7. Results and Discussion

A summary of the results of the tests for structural change and spatial heterogeneity is presented in Tables 1 to 5.

7.1. Test for Structural Change

The proposed test is specific enough, since it accepts the null hypothesis of no structural change when there is no structural change. The method is also sensitive to structural change since most of the time it rejects the null hypothesis of no temporary structural change if indeed the data is embedded with structural change.

For structural change that occurred at the start, the test appears to be sensitive whether the sample size is small or large. On the other hand, for structural change that occurred in the middle, the test appears to be sensitive provided that if the sample size is small to moderate, the proportion of observations with structural change is relatively large. Meanwhile, for structural change that occurred at the end of the series, the test appears to detect structural change provided that the proportion of observations with structural change is not that large. See Table 1 for details.

[Table 1 Here]

AR models put more weight to recent observations than older ones, thus, the procedure can detect temporary structural change that occurs at the start of the series because the structure of the observations at the start is different from the more recent observations. Meanwhile, the procedure can only detect temporary structural change that occurs at the end of the series if the proportion of observations with structural change is low to moderate since if the extent of structural change is large, then, given that the AR estimation puts more weight to recent observations, it might consider the structure of the new observations as the "true" pattern of the data itself and not just a temporary structural change.

Finally, the extent of temporary structural change should be large enough for the procedure to detect it when it occurred at the middle of the series because the weights on the observations at the middle of the series can either be small or large depending on the length of the time series.

The test finds it difficult to correctly decide that there is no temporary structural change when there is really none when the number of spatial units is small and the number of time points is small to moderately large. Similarly, the abovementioned difficulty is observed as the number of spatial units is increased and the number of time points available is large.

Meanwhile, when the number of spatial units is large and there is relatively fewer number of time points, the test can only detect the temporary structural change if it occurs at the start of the series and the proportion of affected observations is large. Temporary structural change that occurs in the middle or at the end of the series is easily detected as the number of time points is increased for large number of spatial units. It was noted earlier that temporary structural change that occurs in the middle of the series can easily be detected as long as the proportion of affected time points is large; however, there is also evidence that when both the number of spatial units and time points are large, the temporary structural change that occurs in the middle is easily detected regardless of the size of the affected time points. In spite of this, it also seems that when the number of spatial units and the number of time points are both large, the temporary structural change that occurs at the start or at the end of the series becomes negligible and so, are no longer detected by the test.

The performance of the proposed procedure was also evaluated for the cases when the fit of the model is good and not good by considering high and low values of the coefficient of determination, R^2. The values of R^2 considered were in the vicinity of 20% and 95%. The proposed procedure for testing temporary structural change is unaffected by the change in the level of R^2. See Table 2 for details. However, this is to be expected because of the manner by which the dataset with low R^2 was obtained, i.e., by multiplying a constant to the error term of the data with R^2=95%. Multiplying a constant to the error term resulted in higher variance but the autoregressive parameter remained the same. And so, it can be observed from the tables that the median of the estimates, as well as the low- and high-end values of the 95% BCI for the temporal parameter is more or less the same for both levels of R^2.

[Table 2 Here]

7.2. Test for Spatial Heterogeneity

The proposed test for spatial heterogeneity is specific since it accepts the null hypothesis of no spatial heterogeneity when there is none. It is also sensitive since it rejects the null hypothesis of no spatial heterogeneity when the data is embedded with spatial heterogeneity.

When the fit of the model is good, spatial heterogeneity is actually present, it is easily detected when the number of spatial units is small and the number of time points is relatively large. The number of neighborhoods affected with spatial heterogeneity does not matter. On the contrary, for large number of spatial units, for small to moderately large number of time points, spatial heterogeneity must be present in all the neighborhoods in order for the test to detect it. Meanwhile, as the number of spatial units and number of time points become large, the proposed procedure can no longer detect spatial heterogeneity as the change in the spatial effect is diluted with more observations with no spatial heterogeneity. See Table 3 for details.

[Table 3 Here]

When the number of time points is small to moderately large and the number of spatial units is small, the proportion of units with spatial heterogeneity must be small to moderately large for the procedure to detect spatial heterogeneity. However for large number of time points, the extent of spatial heterogeneity must be large enough for spatial heterogeneity to be detected. For moderately large number of spatial units (N=40), it seems that the test is sensitive to the presence of spatial heterogeneity regardless of the number of time points available. For large number of spatial units, the number of time points should not be that large and spatial heterogeneity must be present in almost all neighborhoods for the procedure to detect it.

When the fit of the model is not that good (R^2 = 50% and 20%), the proposed procedure for testing spatial heterogeneity is still found to be specific enough (see Tables 4 and 5). However, for 50% R^2, it was observed that the procedure already finds it difficult to detect spatial heterogeneity when the number of spatial units is small. Based on the results of the simulation study, when the number of spatial units is small, spatial heterogeneity is detected only when it is present in all of the neighborhoods and the number of time points is large enough. The proposed procedure is sensitive to spatial heterogeneity for moderately large number of spatial units (N=40), provided that the extent of spatial heterogeneity is large enough. Meanwhile, for large number of spatial units, spatial heterogeneity is detected when it is scattered in the neighborhoods and the extent of spatial heterogeneity is large enough.

When R^2 =20%, the proposed procedure is sensitive to spatial heterogeneity when the extent of spatial heterogeneity is small to moderately large for small N and moderately large to large T and for small T when N is relatively large. In addition, when the fit of the model is not good, it was observed that there is a possibility that the spatial effect is not significantly different from zero and the 95% BCI's obtained are very wide.

[Table 4 Here]

[Table 5 Here]

8. Conclusions

The paper proposed a nonparametric procedure for testing the presence of temporary structural change and spatial heterogeneity in the spatio-temporal model proposed by Landagan and Barrios (2007). The proposed procedure utilized the AR-sieve bootstrap to detect temporary structural change and the forward search algorithm to detect spatial heterogeneity. Bootstrap was used to construct the 95% confidence intervals for the temporal and spatial parameters, which, in turn, were used to test whether a temporary structural change or spatial heterogeneity is present or not.

The proposed procedure for testing temporary structural change and spatial heterogeneity is specific in a sense that under some conditions, the proposed procedure can correctly identify non-occurrence of structural change or spatial heterogeneity when they are actually non-existent. It is also sensitive since it can aptly identify cases where structural change or spatial heterogeneity indeed happened.

The nonparametric procedure for testing temporary structural change is recommended for use when we suspect that there has been a change in the temporal effect with recent observations and the available space-time data is composed of a relatively short time series and relatively small number of spatial units. On the other hand, the use of the nonparametric procedure for testing spatial heterogeneity is recommended when dealing with relatively small to moderate number of spatial units and assuming that the fit of model considered is good enough.

References:

Atkinson, A. C. and M. Riani (2002), 'Forward search added-variable *t*-tests and the effect of masked outliers on model selection', *Biometrika*, 89, 939-946.

Atkinson, A. C. and M. Riani (2007), 'Building regression models with the forward search', *Journal of Computing and Information Technology – CIT*, 15, 287-294.

Bühlmann, P. (1997), 'Sieve bootstrap for time series', *Bernoulli*, 3**,** 123-148.

Bühlmann, P. (2002), 'Bootstraps for time series', *Statistical Science,* 17**,** 52-72.

Dumanjug, C., E. Barrios, and J. Lansangan (2009), 'Bootstrap procedures in a spatio-temporal model', forthcoming in *Journal of Statistical Computation and Simulation*.

Hering, A.S. , C. L. Bell, and M. G. Genton (2006), 'Modeling spatio-temporal wildfire ignition point patterns', *Environ. Ecol. Stat.* DOI: 10.1007/s10651-007-0080-6.

Ioannidis, E. A. (2005), 'Residual-based block bootstrap unit root testing in the presence of trend breaks', *Econometrics Journal*, 8, 323 – 351.

Konis, K. and F. Laurini (2007), 'Fitting a forward search in linear regression', *Bulletin of the 56th Meeting of the International Statistical Institute*, Portugal.

Landagan, O.Z. and E. Barrios (2007), 'An estimation procedure for a spatio-temporal model', *Statistics and Probability Letters,* 77, 401-406.

Mavridis, D. and I. Moustaki (2007), 'Implementing the forward search algorithm in latent variable models for identifying outliers and extreme response patterns', *Bulletin of the 56th Meeting of the International Statistical Institute*, Portugal.

Paparoditis, E. and D. N. Politis (2005), 'Bootstrap hypothesis testing in regression models', *Statistics and Probability Letters*, 74, 356-365.

Paparoditis, E. and D. N. Politis (2003), 'Residual-based block bootstrap for unit root testing', *Econometrica,* 71**,** 813-855.

Park, J. Y. (2003), 'Bootstrap unit root tests', *Econometrica*, 17, 1845-1895.

Politis, D. N. (2003), 'The impact of bootstrap methods on time series analysis', *Statistical Science*, 18, 219 - 230.

Politis, D. and Romano, J, (1994), 'The Stationary Bootstrap', *Journal of the American Statistical Association*, 89, 1303-1313.

Singh, K. and R. H. Berk (1994), ‘A concept of type-2 *p*-value’, *Statistica Sinica*, 4 **,** 493-504.

Stroud, J.R., P. Muller, and B. Sanso (2001), ‘Dynamic models for spatiotemporal data’, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 63**,** 673-689.

Waller, L. A., et al. (2007), ‘Spatio-temporal patterns in country-level incidence and reporting of Lyme disease in the northeastern United States, 1990-2000’, *Environ. Ecol. Stat.*, 14, 83-100.

Westerlund, J. and D. L. Edgerton (2007), ‘A panel bootstrap cointegration test’, *Economics Letters*, 97, 185-190.

Table 1.
Summary of the Results of the Tests for Temporary Structural Change: $R^2 = 95\%$
(Percentage of $\hat{\rho}'s$ covered in the 95% Bootstrap Confidence Interval for ρ)

POSITION OF STRUCTURAL CHANGE	N = 20			N = 40			N = 60		
	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75
NO STRUCTURAL CHANGE	92.2	95.0	95.5	95.0	96.9	94.2	95.9	95.4	93.1
(a) Proportion of time points with structural change = 5%									
START	94.3	93.5	94.6	94.2	93.1	93.7	95.2	96.4	95.0
MIDDLE	95.0	97.4	94.2	95.9	95.7	95.3	95.3	95.7	95.0
END	93.9	94.6	93.7	95.3	94.8	93.8	95.4	94.4	95.5
(b) Proportion of time points with structural change = 10%									
START	92.8	96.4	94.7	95.0	95.1	93.4	95.7	95.5	95.8
MIDDLE	95.6	94.5	95.7	94.6	95.2	96.3	95.2	95.9	94.8
END	93.1	92.6	94.3	94.8	94.1	95.2	96.4	94.7	95.0
(c) Proportion of time points with structural change = 15%									
START	94.3	94.2	94.8	92.6	95.0	94.8	94.9	95.3	95.8
MIDDLE	93.7	94.9	95.4	96.2	94.5	95.2	95.6	94.7	93.5
END	96.5	95.9	96.0	95.6	95.5	94.1	95.0	95.7	95.5

NOTES:
1. Ho: There is no temporary structural change vs. Ha: A temporary structural change occurred. The null hypothesis is rejected when the percentage of $\hat{\rho}'s$ covered in the 95% bootstrap confidence interval is less than 95%.
2. When there is no structural change, the test is specific if the null hypothesis is not rejected. On the other hand, if a structural change occurred at the start, middle or end of the series, the test is said to be sensitive if the null hypothesis is rejected.

Table 2. Summary of the Results of the Tests for Temporary Structural Change: $R^2 = 20\%$
(Percentage of $\hat{\rho}'s$ covered in the 95% Bootstrap Confidence Interval for ρ)

POSITION OF STRUCTURAL CHANGE	N = 20			N = 40			N = 60		
	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75
NO STRUCTURAL CHANGE	92.2	94.5	94.4	95.0	96.9	94.2	95.9	95.4	93.1
(a) Proportion of time points with structural change = 5%									
START	94.3	93.5	95.0	93.8	93.1	93.7	95.2	96.4	95.0
MIDDLE	95.0	97.1	93.8	95.9	95.7	95.3	95.3	95.7	95.0
END	94.0	94.7	94.0	95.7	94.8	93.8	95.4	94.4	95.5
(b) Proportion of time points with structural change = 10%									
START	92.8	96.1	95.1	95.0	95.1	93.4	95.7	95.5	95.8
MIDDLE	95.6	93.9	96.3	94.6	95.2	96.3	95.2	95.9	94.8
END	93.0	92.8	94.3	94.0	94.1	95.1	96.4	94.7	95.0
(c) Proportion of time points with structural change = 15%									
START	94.3	94.2	93.8	92.6	95.0	94.8	94.9	95.3	95.8
MIDDLE	93.7	96.0	95.4	96.2	94.5	95.2	95.6	94.7	93.5
END	97.3	96.0	95.3	96.5	95.5	94.1	95.0	95.7	95.5

NOTES:

1. Ho: There is no temporary structural change vs. Ha: A temporary structural change occurred. The null hypothesis is rejected when the percentage of $\hat{\rho}'s$ covered in the 95% bootstrap confidence interval is less than 95%.

2. When there is no structural change, the test is specific if the null hypothesis is not rejected. On the other hand, if a structural change occurred at the start, middle or end of the series, the test is said to be sensitive if the null hypothesis is rejected.

Table 3. Summary of the Results of the Tests for Spatial Heterogeneity: R^2 = 95%
(Percentage of $\hat{\delta}'s$ covered in the 95% Bootstrap Confidence Interval for δ)

No. of Neighborhoods with Spatial Heterogeneity	N = 20			N = 40			N = 60		
	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75
NO SPATIAL HETEROGENEITY	95.0	94.0	96.0	97.5	96.0	97.3	92.5	96.0	100.0
(a) Proportion of spatial units with spatial heterogeneity = 5%									
1 NEIGHBORHOOD	92.5	88.0	97.3	92.5	98.0	98.7	100.0	100.0	98.7
2 NEIGHBORHOODS	95.0	100.0	96.0	92.5	94.0	94.7	100.0	100.0	100.0
3 NEIGHBORHOODS	92.5	96.0	97.3	95.0	94.0	93.3	97.5	92.0	98.7
4 NEIGHBORHOODS	90.0	88.0	94.7	95.0	92.0	93.3	92.5	94.0	96.0
(b) Proportion of spatial units with spatial heterogeneity = 10%									
1 NEIGHBORHOOD	95.0	94.0	97.3	95.0	98.0	96.0	100.0	94.0	97.3
2 NEIGHBORHOODS	95.0	94.0	94.7	92.5	96.0	94.7	97.5	98.0	97.3
3 NEIGHBORHOODS	100.0	98.0	94.7	95.0	96.0	92.0	100.0	100.0	100.0
4 NEIGHBORHOODS	92.5	94.0	93.3	90.0	90.0	90.7	92.5	94.0	94.7
(a) Proportion of spatial units with spatial heterogeneity = 15%									
1 NEIGHBORHOOD	90.0	98.0	89.3	100.0	94.0	94.7	95.0	96.0	97.3
2 NEIGHBORHOODS	97.5	96.0	90.7	92.5	98.0	96.0	100.0	98.0	98.7
3 NEIGHBORHOODS	95.0	96.0	93.3	90.0	90.0	92.0	97.5	92.0	98.7
4 NEIGHBORHOODS	100.0	100.0	96.0	87.5	84.0	93.3	92.5	94.0	100.0

NOTES:
1. Ho: There is no spatial heterogeneity vs. Ha: There is spatial heterogeneity. The null hypothesis is rejected when the percentage of $\hat{\delta}'s$ covered in the 95% bootstrap confidence interval is less than 95%.
2. When there is no spatial heterogeneity, the test is specific if the null hypothesis is not rejected. On the other hand, if there is spatial heterogeneity in one, two, three or four neighborhoods, the test is said to be sensitive if the null hypothesis is rejected.

Table 4. Summary of the Results of the Tests for Spatial Heterogeneity: $R^2 = 50\%$
(Percentage of $\hat{\delta}'s$ covered in the 95% Bootstrap Confidence Interval for δ)

No. of Neighborhoods with Spatial Heterogeneity	N = 20			N = 40			N = 60		
	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75
NO SPATIAL HETEROGENEITY	95.0	94.0	97.3	97.5	96.0	94.7	92.5	96.0	100.0
(a) Proportion of spatial units with spatial heterogeneity = 5%									
1 NEIGHBORHOOD	97.5	98.0	96.0	92.5	98.0	97.3	95.0	96.0	97.3
2 NEIGHBORHOODS	95.0	96.0	96.0	92.5	92.0	97.3	92.5	100.0	93.3
3 NEIGHBORHOODS	95.0	96.0	97.3	92.5	96.0	96.0	100.0	100.0	100.0
4 NEIGHBORHOODS	95.0	94.0	97.3	97.5	96.0	96.0	95.0	98.0	98.7
(b) Proportion of spatial units with spatial heterogeneity = 10%									
1 NEIGHBORHOOD	97.5	100.0	97.3	92.5	92.0	93.3	100.0	100.0	97.3
2 NEIGHBORHOODS	97.5	96.0	97.3	92.5	94.0	98.7	92.5	100.0	96.0
3 NEIGHBORHOODS	95.0	98.0	96.0	92.5	96.0	97.3	92.5	100.0	96.0
4 NEIGHBORHOODS	95.0	94.0	97.3	97.5	96.0	94.7	95.0	92.0	94.7
(a) Proportion of spatial units with spatial heterogeneity = 15%									
1 NEIGHBORHOOD	100.0	98.0	94.7	92.5	96.0	97.3	97.5	100.0	100.0
2 NEIGHBORHOODS	97.5	96.0	97.3	97.5	98.0	98.7	100.0	100.0	100.0
3 NEIGHBORHOODS	97.5	96.0	96.0	92.5	96.0	98.7	95.0	96.0	94.7
4 NEIGHBORHOODS	95.0	96.0	97.3	92.5	98.0	96.0	95.0	94.0	90.7

NOTES:
1. Ho: There is no spatial heterogeneity vs. Ha: There is spatial heterogeneity. The null hypothesis is rejected when the percentage of $\hat{\delta}'s$ covered in the 95% bootstrap confidence interval is less than 95%.
2. When there is no spatial heterogeneity, the test is specific if the null hypothesis is not rejected. On the other hand, if there is spatial heterogeneity in one, two, three or four neighborhoods, the test is said to be sensitive if the null hypothesis is rejected.

Table 5. Summary of the Results of the Tests for Spatial Heterogeneity: R^2 = 20%
(Percentage of $\hat{\delta}$'s covered in the 95% Bootstrap Confidence Interval for δ)

No. of Neighborhoods with Spatial Heterogeneity	N = 20			N = 40			N = 60		
	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75	T = 40	T = 50	T = 75
NO SPATIAL HETEROGENEITY	95.0	94.0	97.3	97.5	96.0	94.7	92.5	96.0	100.0
(a) Proportion of spatial units with spatial heterogeneity = 5%									
1 NEIGHBORHOOD	95.0	96.0	92.0	92.5	98.0	97.3	92.5	92.0	98.7
2 NEIGHBORHOODS	95.0	94.0	96.0	92.5	96.0	97.3	92.5	94.0	96.0
3 NEIGHBORHOODS	95.0	94.0	97.3	92.5	96.0	93.3	97.5	100.0	98.7
4 NEIGHBORHOODS	95.0	96.0	92.0	95.0	96.0	96.0	97.5	98.0	100.0
(b) Proportion of spatial units with spatial heterogeneity = 10%									
1 NEIGHBORHOOD	97.5	100.0	96.0	90.0	96.0	98.7	95.0	98.0	100.0
2 NEIGHBORHOODS	95.0	96.0	97.3	90.0	98.0	97.3	92.5	100.0	100.0
3 NEIGHBORHOODS	95.0	94.0	97.3	97.5	96.0	96.0	95.0	100.0	100.0
4 NEIGHBORHOODS	95.0	96.0	92.0	97.5	96.0	97.3	97.5	100.0	94.7
(a) Proportion of spatial units with spatial heterogeneity = 15%									
1 NEIGHBORHOOD	97.5	98.0	96.0	95.0	98.0	96.0	97.5	100.0	100.0
2 NEIGHBORHOODS	97.5	96.0	98.7	92.5	98.0	97.3	95.0	100.0	100.0
3 NEIGHBORHOODS	95.0	96.0	97.3	95.0	96.0	96.0	100.0	100.0	100.0
4 NEIGHBORHOODS	95.0	96.0	97.3	95.0	96.0	96.0	97.5	96.0	94.7

NOTES:
1. Ho: There is no spatial heterogeneity vs. Ha: There is spatial heterogeneity. The null hypothesis is rejected when the percentage of $\hat{\delta}$'s covered in the 95% bootstrap confidence interval is less than 95%.
2. When there is no spatial heterogeneity, the test is specific if the null hypothesis is not rejected. On the other hand, if there is spatial heterogeneity in one, two, three or four neighborhoods, the test is said to be sensitive if the null hypothesis is rejected.