Modeling Non-Linear Spatial Dynamics: A Family of Spatial STAR Models and an Application to U.S. Economic Growth

Valerien O. Pede
Department of Agricultural Economics, Purdue University
West Lafayette, IN 47907, USA, vpede@purdue.edu

Raymond J.G.M. Florax
Department of Agricultural Economics, Purdue University
West Lafayette, IN 47907, USA, rflorax@purdue.edu
Department of Spatial Economics, VU University
Amsterdam, The Netherlands

Matthew T. Holt
Department of Agricultural Economics, Purdue University
West Lafayette, IN 47907, USA, mholt@purdue.edu

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008.

preliminary version
not for quotation – comments invited

Abstract

This paper investigates non-linearity in spatial processes models and allows for a gradual regime-switching structure in the form of a smooth transition autoregressive process. Until now, applications of the smooth transition autoregressive (STAR) model have been largely confined to the time series context. The paper focuses on extending the non-linear smooth transition perspective to spatial processes models, in which spatial correlation is taken into account through the use of a so-called weights matrix identifying the topology of the spatial system. We start by deriving a non-linearity test for a simple spatial model, in which spatial correlation is only included in the transition function. Next, we propose a non-linearity test for a model that includes a spatially lagged dependent variable or spatially autocorrelated innovations as well. Monte Carlo simulations of the various test statistics are performed to examine their power and size. The proposed modeling framework is then used to identify convergence clubs in the context of U.S. county-level economic growth over the period 1963–2003.

Keywords: spatial econometrics, non-linearity, autoregressive smooth transition
JEL Classification: C12, C21, C51, O18, R11

1. Introduction

Over the last decade, nonlinear times series modeling has gained considerable attention in the applied economics literature. Nonlinear models can capture certain features of business cycles, which linear models cannot. The presence of asymmetry and nonlinearities in business cycles can be incorporated in smooth transition autoregressive (STAR) models. The STAR model is a member of the family of nonlinear models that exhibit regime-dependent or regime-switching behavior. As implied by its name, the STAR framework allows the model parameters to take on different values across regimes following a potentially smooth transition process. The STAR model has been used to model nonlinearity and asymmetric response in applications pertaining to industrial production (Terasvirta and Anderson, 1992), the hog-corn cycle (Holt and Craig, 2006), exchange rates (Baharumshah and Khim-Sen Liew, 2006), interest rates (van Dijk and Franses, 2000), and unemployment rates (Skalin and Terasvirta, 2002).

Applications of the smooth transition approach have been largely confined to time series models; even although spatial econometrics has evolved as one of the fastest growing sub-fields of econometrics. Spatial econometrics allows modeling spatial interactions (spatial autocorrelation) and spatial structure (spatial heterogeneity) in regressions using cross-sectional or panel data. Considerable progress has recently been made in developing consistent and efficient estimators for spatial regressions (Kelejian and Prucha, 2007; Lee, 2002, 2007). Spatial dependence represents a functional relationship between what happens at one point in space and what happens elsewhere. Spatial heterogeneity occurs when there is no uniformity of the observed effects across space. Two types of spatial process are common in the spatial econometrics literature: the spatial lag model which considers spatial spillovers in the dependent variable in the form of the spatially lagged dependent and the spatial error model in which the disturbances are generated by a spatial autoregressive process. Recently, models that combine both processes in terms of spatial autoregressive models with autoregressive disturbances (ARAR models) have received considerable attention (Kelejian and Prucha, 2004, 2007; Lee, 2002, 2007). Spatial processes have often been assumed to be linear, typically in terms of a linear autoregressive or moving average process. It is much more likely, however, that spatial dynamics exhibit non-linear features in a way that is similar to time series models.

In spite of the popularity of STAR models in depicting nonlinear models, there have been no applications of this general approach to spatial processes. That is, spatial models that incorporate nonlinearities in the form of regime switching have to date not been explored. This research proposes a statistical methodology to formally test for non-linearity in spatial processes models and allowing for a gradual regime switching structure in the form of a smooth transition autoregressive process. We start by deriving the non-linearity test for a simple non-spatial model. Next, we propose a non-linearity test for a model that includes a spatially lagged dependent variable or spatial autocorrelated innovations. Monte Carlo simulations of the various test statistics are performed to examine their small sample power and size. The proposed modeling framework is then used in the analysis of convergence clubs with respect to U.S. county-level economic growth over the period 1963-2003.

The rest of the paper is organized as follow. The next section provides a background review on the STAR model in the time series contest. Section 3 describes the spatial smooth transition autoregressive model. Estimation and Monte Carlo simulations are presented in section 4. The empirical application on the endogenous determination of "convergence clubs" for U.S. counties is presented section 5. Finally, section 6 concludes the paper.

2. Background on STAR Model

A basic STAR model with time series data can be represented as follow:

$$y_t = \varphi_1' x_t (1 - G(s_t, \gamma, c)) + \varphi_2' x_t G(s_t, \gamma, c) + \varepsilon_t , \tag{1}$$

or alternatively,

$$y_t = \phi_1' x_t + \phi_2' x_t G(s_t, \gamma, c) + \varepsilon_t \tag{2}$$

where $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, ..., y_{t-p})$ and $\varphi_i = (\varphi_{i0}, \varphi_{i,1}, ..., \varphi_{ip})'$, $i = 1, 2$, $\phi_1 = \varphi_1$, $\phi_2 = \varphi_2 - \varphi_1$, ε_t is the error term distributed independently and identically with mean zero and variance σ^2 . $G(s_t, \gamma, c)$ is the transition function bounded between zero and one, allowing for a smooth transition between regimes.

Within the transition function, s_t represents the transition variable and it could be a lagged endogenous variable $s_t = y_{t-d}$, where d is referred to as the delay parameter; a function of lagged endogenous variables $s_t = f(\tilde{x}_t, \alpha)$ where f is a function and α a parameter vector; an exogenous variable $s_t = z_t$, or a linear time trend $s_t = t$.[1] The arguments γ and c are slope and location parameters, respectively. The parameter γ is also referred to as the smoothness parameter, and c as the threshold between the two regimes. Two functional forms of the logistic function are common in the time series literature of STAR models: the logistic function and the exponential function. Using the lagged endogenous variable y_{t-d} as transition variable, the logistic form is expressed as:

$$G(y_{t-d}, \gamma, c) = [1 + \exp - \{\gamma(y_{t-d} - c)/\sigma(y_{t-d})\}]^{-1}. \tag{3}$$

Combining (1) or (2) with (3) gives the logistic STAR (LSTAR) model. The exponential form is given as:

$$G(y_{t-d}, \gamma, c) = 1 - \exp - \left\{\gamma((y_{t-d} - c)/\sigma(y_{t-d}))^2\right\} . \tag{4}$$

Combining (1) or (2) with (4) gives the exponential STAR (ESTAR) model. For large values of the parameter γ, the logistic function converges to one when $y_{t-d} - c \succ 0$ and to zero when $y_{t-d} - c \prec 0$. When $\gamma \to 0$ the LSTAR converges to an autoregression model of order p (AR(p)). The ESTAR shows a slightly different pattern with respect to γ. For large values of γ, the exponential function converges to one for values of y_{t-d} below or above the threshold parameter c.

The STAR model offers the possibility to investigate the presence of nonlinearity in time series data. To this end, the STAR model is tested against the linear AR model. Luukkonen et al. (1988) suggested replacing the transition function $G(s_t, \gamma, c)$ by a suitable first order Taylor series approximation and deriving the LM test of nonlinearity. Taking the

---

[1] When the transition variable is a linear time trend ($s_t = t$), the STAR model is called a Time-Varying Autogressive Model (TVAR).

first order Taylor series approximation of the transition function $G(s_t, \gamma, c)$ and substituting into (2) yields a simplified form of the STAR model given as:

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \mu_t \tag{5}$$

where β_0 and β_1 are functions of original parameters in (2). The nonlinearity test consists in deriving an LM test to test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$.

Luukkonen et al. (1988) also noticed that the LM test involving a third order Taylor series approximation has better power than the LM test obtained with the first order approximation. With a third order Taylor series approximation, the STAR model is given as:

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + \beta_3' x_t s_t^3 + \mu_t \tag{6}$$

The test of nonlinearity consists in deriving an LM test to test the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ against the alternative $H_1 : \beta_1 \neq 0,\ \beta_2 \neq 0,\ \beta_3 \neq 0$.

3. Specification of Spatial STAR Model

Two types of spatial regressions are common in the spatial econometrics literature: the spatial error model and the spatial lag model. The specification of the spatial error model is relevant when the dependence works through the error process. The errors are then assumed to be generated by a spatial autoregressive process. The spatial error model may be written as:

$$y = X\beta + \varepsilon \text{ , where } \varepsilon = \lambda W \varepsilon + \mu , \tag{7}$$

where y is an $N \times 1$ vector of observations on the dependent variable, X is an $N \times K$ matrix of explanatory variables, β is a vector of unknown parameters, W is an $N \times N$ weight matrix which defines the spatial structure of regions, λ is a scalar parameter, μ is an $N \times 1$ vector of random error terms with mean 0 and constant variance. OLS estimation of the spatial error is unbiased and consistent but inefficient (Anselin, 1988). The spatial error model is appropriately estimated using Maximum Likelihood approach or General Moment. The spatial lag model is relevant when the variable under investigation depends on its spatial lag. In other words, the model considers that spatial spillovers is present and is captured through the spatially lagged dependent variable. The spatial lag model may be expressed as:

$$y = \rho W y + X\beta + \mu \tag{8}$$

where ρ is a scalar parameter, and all other defined as before. OLS estimation of the spatial lag model is biased and inconsistent. Appropriate estimation is obtained by using Maximum Likelihood or Instrumental Variables approach (Anselin, 1988).

The objective of this paper is to develop a methodology to test for nonlinearity in the above-defined spatial processes, allowing for a gradual regime switching structure in the form of a smooth transition autoregressive process. We first start with a basic specification of the STAR model for spatial variables by analogy to the model presented for time series:

$$y = (\alpha_0 + \alpha_1 x) + (\delta_0 + \delta_1 x)G(s, \gamma, c) + \varepsilon \tag{9}$$

Where y represents $N \times 1$ vector of observations on N regions, x is an $N \times 1$ vector representing an explanatory variable, G the transition function, s the transition variable, γ the smoothness parameter, c the location parameter and ε is the error term.

By analogy to the time series context, a possible candidate for the transition variable could be the spatially lagged dependent variable or a spatially lagged independent variable. In this paper, we will consider the latter. We therefore define the spatial lag of the independent variable x as Wx, where W represents the exogenously defined weight matrix.[2] W is a Boolean matrix, with element taking value 1 when regions are neighbors and 0 when they are not. If the weight matrix W is standardized, then Wx simply represents the average value of x at the neighbors.

The transition function could therefore be defined for both the logistic and exponential form respectively as:

---

[2] Wx is an $n \times 1$ vector of observations, and simply represents the cross-product of the weight matrix W and x

$$G(Wx,\ \gamma,\ c) = [1 + \exp - \{\gamma(Wx - c)\}]^{-1} \tag{10}$$

$$G(Wx,\ \gamma, c) = 1 - \exp - \{\gamma(Wx - c)^2\} \tag{11}$$

In this paper, we will only consider the LSTAR model, which is simply the combination of equation (9) and (10). The logistic function changes monotically and smoothly from zero to one depending on values of the transition variable Wx. For large values of the parameter γ, the logistic function converges to one when Wx is above the threshold value c, and converges to zero when Wx is below the threshold value c. Intuitively, this implies that the transition variable changes monotically and smoothly from zero to one as the characteristics of the neighbors (Wx) changes.

Under the assumption of homoskedastic errors, the LSTAR model obtained by combining equation (9) and (10) could be viewed as a simple spatial model, in which spatial correlation is only included in the transition function. We first start by deriving nonlinearity tests on this model. Considering a first order Taylor series approximation of the logistic function in equation (10) and substituting back into equation (9) yields a nonlinear LSTAR model of the form:

$$y = \beta_0 + \beta_1 x + \beta_2 Wx + \beta_3 Wx * x + \varepsilon \tag{12}$$

where β_i, $i = 0,1,2,3$ are function of the original parameters in (9) and ε are assumed to be independently and identically distributed with mean zero and a constant variance.[3] A test of linearity involves testing the null hypothesis $H_0 : \beta_2 = \beta_3 = 0$ against the alternative that H_0 is not true. We employ an LM test with χ^2 distribution and degree of freedom equal to 2. Rejection of the null hypothesis would imply that the model is nonlinear.

Next, using the model in equation (9) again, we also develop a spatial model where spatial autocorrelation is present not only in the transition function but also in the form of spatial lag. After taking the first order Taylor series approximation of the transition function in equation (10), this model could be expressed as:

[3] Proof of derivation of the model in equation (12) is provided in appendix 1.

$$y = \rho Wy + \beta_0 + \beta_1 x + \beta_2 Wx + \beta_3 Wx * x + \varepsilon \tag{13}$$

where ρ represents the scalar spatial autoregressive parameter, the errors ε are assumed to be independently and identically distributed with mean zero and a constant variance σ^2, all other terms are as previously defined. Considering the model in equation (13), it would therefore be possible to test for both spatial dependence and nonlinearity jointly. This will involve testing the null hypothesis $H_0 : \rho = 0 \;\; \textit{and} \;\; \beta_2 = \beta_3 = 0$. To this end, an LM test with χ^2 distribution is used. Failure to reject the null hypothesis would indicate that the model is linear and does not show presence of spatially lagged dependence. However, rejection could lead to the test of other null hypotheses:

- Testing for nonlinearity only, assuming there is presence of spatial lag $H_0 : \beta_2 = \beta_3 = 0 \quad \rho \neq 0$
- Testing for spatial dependence only, assuming the presence of nonlinearity $H_0 : \rho = 0 \quad \beta_2 \neq \beta_3 \neq 0$

From a different perspective of the above, we also develop a spatial model where spatial autocorrelation is present in the transition function but in addition there are autoregressive error processes. After taking the first order Taylor series approximation of the transition function in (10) and substituting back into equation (9), this model is expressed as:

$$y = \beta_0 + \beta_1 x + \beta_2 Wx + \beta_3 Wx * x + \varepsilon \quad \text{and} \quad \varepsilon = \lambda W\varepsilon + \mu \tag{14}$$

Where λ represents the coefficient of the autoregressive error term, μ is a vector of random errors with mean zero and variance σ^2 and all other terms are as previously defined. Considering the model in equation (14), it is also possible to test for both nonlinearity and spatial error dependence jointly. This will involve testing the null hypothesis $H_0 : \lambda = 0 \;\; \textit{and} \;\; \beta_2 = \beta_3 = 0$. The LM test with χ^2 distribution is also used for this test. The acceptance of the null hypothesis would indicate that the model is linear and there is no presence of spatial autocorrelation in the form of autoregressive errors. However, if the null hypothesis is rejected, other possible specifications need to be tested as well, through the following null hypotheses:

- Testing for nonlinearity only, assuming there is presence of autoregressive errors $H_0 : \beta_2 = \beta_3 = 0 \quad \lambda \neq 0$
- Testing for spatial dependence only, assuming the presence of nonlinearity $H_0 : \lambda = 0 \quad \beta_2 \neq \beta_3 \neq 0$

4. Estimation and Hypothesis Testing

In this section, we derive analytical solution of the LM tests for the above-described models. Maximum likelihood estimation is used to estimate the various parameters in each model.

Starting with the model in equation (12) the null hypothesis of nonlinearity $H_0 : \beta_2 = \beta_3 = 0$ is simply tested with an LM test given as:

$$LM_\beta = \frac{e^{*T} X \left[X^T X\right]^{-1} X^T e^*}{e^{*T} e^* / N} \tag{15}$$

where $e^* = y - X\beta^*$ is the residual from the restricted model, $\beta^* = (\beta_0^*, \beta_1^*, 0, 0)$ represents the parameters of the restricted model, $X = (1, x, Wx, Wx * x)$ and N is the number of observations. [4]

The model in equation (13) can be reformulated in intensive form as:

$$y = \rho W y + X\beta + \varepsilon \tag{16}$$

where $X = (1, x, Wx, Wx * x)$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ and the errors ε are assumed to be independently and identically distributed with mean zero and a variance σ^2. The LM test for testing the null hypothesis $H_0 : \rho = 0 \;\; and \;\; \beta_2 = \beta_3 = 0$ is given as:

[4] It could be shown that the LM test equals to NR^2, where R^2 is the uncentered R^2 from the regression of the residual e^* on X .

$$LM_{\rho\beta} = \frac{\left[\frac{\left[e^{*T}X\right]\left[X^TX\right]^{-1}\left[X^TWX\beta^*\right]}{\sigma^2} - \left[\frac{1}{\sigma^2}e^{*T}\left[Wy\right]\right]\right]^2}{\left[T + \frac{\left[WX\beta^*\right]^T M\left[WX\beta^*\right]}{\sigma^2}\right]} + \frac{\left[\left[e^*\right]^T X\right]\left[X^TX\right]^{-1}\left[X^Te^*\right]}{\sigma^2} \qquad (17)$$

Where $e^* = y - X\beta^*$ is the residual from the restricted model, β^* represents the parameters of the restricted model, $\sigma^2 = (e^{*T}e^*)/N$, $T = tr\left[W^2 + W^TW\right]$. The proof of the derivation of the LM test in (17) is provided in Appendix 2a.

The LM test for the presence of spatial lag assuming the presence of nonlinearity ($H_0 : \rho = 0$) is given as:

$$LM_{\rho/\beta^*} = \frac{\left[\frac{1}{\sigma^2}e^TWy\right]^2}{NJ} \qquad (18)$$

where $e = y - X\beta$, $X = (1, x, Wx, Wx * x)$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $NJ = \left[tr\left[W^2 + W^TW\right] + \frac{\left[WX\beta\right]^T M\left[WX\beta\right]}{\sigma^2}\right]$.

Alternatively, it is also possible to test for nonlinearity assuming the presence of spatial lag. The ML test is given as:

$$LM_{\beta^*/\rho} = \frac{e^{*T}X\left[X^TX\right]^{-1}X^Te^*}{e^{*T}e^*/N} \qquad (19)$$

where $e^* = y - \rho Wy - X\beta^*$ is the residual from the restricted model, and the other terms are defined as before.

Reformulating the model in equation (14) in intensive form yields:

$$y = X\beta + \varepsilon \quad \text{and} \quad \varepsilon = \lambda W\varepsilon + \mu \qquad (20)$$

where $X = (1, x, wx, wx * x)$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ and the errors μ are assumed to be independently and identically distributed with mean zero and variance σ^2. The LM test for jointly testing for nonlinearity and the presence of autoregressive errors is given as:

$$LM_{\lambda/\beta^*} = \frac{\left[\frac{1}{\sigma^2} e^{*T} W e^*\right]^2}{T} + \frac{e^{*T} X \left[X^T X\right]^{-1} X^T e^*}{e^{*T} e^* / N} \tag{21}$$

where $e^* = y - X\beta^*$ and the other terms defined as before. The proof of the derivation of the LM test in (21) is provided in Appendix 2b.

The LM test for the presence of autoregressive errors assuming nonlinearity ($H_0 : \lambda = 0$) is given as:

$$LM_{\lambda/\beta^*} = \frac{\left[\frac{1}{\sigma^2} e^T W e\right]^2}{T} \tag{22}$$

where $e = y - X\beta$ and the other terms are defined as before.

Alternatively, to test for nonlinearity, assuming the presence of autoregressive errors, the LM is given as:

$$LM_{\beta^*/\lambda} = \frac{e^{*T} X \left[X^T X\right]^{-1} X^T e^*}{e^{*T} e^* / N} \tag{23}$$

where $e^* = (I - \lambda W)(y - X\beta^*)$, N is the number of observation and the other terms are defined as before.

5. Monte Carlo Simulation

In this section we investigated the performance of the above-described LM tests using Monte Carlo simulations. We first start with a data generating process where variables are artificially created to fit models described in equations (12), (13) and (14) respectively. We generated a 625 x 625 weight matrix corresponding to a regular 25 x 25 grid structure, using a queen criterion.[5] Next we generated the independent variable x from a random uniform distribution. Subsequently, we also created the lagged dependent Wx and the interaction term $Wx * x$. The matrix of independent variables $X = (1, x, wx, wx * x)$ is maintained fixed in the replications. For the simulation, we consider 1000 replications. The dependent variable is generated for each model by fixing the parameters $\alpha_0, \alpha_1, \delta_0, \delta_1, c$ to unity and following the structure of equations. We started the simulation by looping over the following values for the parameter $gamma$: 0, 0.1, 0.3, 0.5, 1, 10 and 100. For the spatial lag and spatial error model, we also looped over following values for parameters ρ and λ: 0, 0.1, 0.2, 0.3, 0.5, 0.7 and 0.9. For each replication, the LM tests are computed and compared to their asymptotic critical value at $\alpha = 0.05$. The proportion of time the null hypothesis is rejected is reported. Table 1a, b, c shows the percentage of rejection of the null hypothesis corresponding to each combination of ρ and λ in the case of the spatial lag model. The size of each corresponds to the probability of rejection when $gamma = rho = 0$. All three tests show relatively good size (about 10%). These tests also show high power, especially for large values of gamma. Similar pattern is observed with in Table 2a, b, c for the spatial error model. The size of each corresponds to the probability of rejection in case $gamma = lambda = 0$. The size of all tests is about 10% and they all show strong power against the null.

6. Empirical Application: Economic Growth Analysis of U.S. Counties, 1969-2003

6.1. Background Review on Economic Growth and Convergence Clubs

Countries of the world are characterized by larges disparities in terms of per capita income and growth rates. While some countries are extremely rich, others are extremely poor. Also, while some countries are growing fast, others are experiencing slow growth. Similar patterns can be observed at lower spatial scale levels, for instance for counties and states in the U.S. Economic growth studies try to explain disparities between countries or regions in terms of real per capital income or growth rates. Despite differences between regions, the neoclassical growth theory predicts that in the long run economic forces will contribute to regions becoming similar in terms of per capita income. This key proposition of neoclassical growth theory is known as the convergence property. In particular, regions with similar

[5] The queen criterion consider all the regions (cells) having a side in common to the north, south, east and west as neighbors, as well as those having a vertex in common.

characteristics will converge to the same steady state.[6] These groups of regions are known as convergence clubs. Identifying these groups has been a challenging task in the economic growth literature. Some studies have proposed exogenous approaches, oftentimes based on spatial statistics or threshold values in initial per capita income or human capital variables to identify groups (Florax and Nijkamp 2005, Le Gallo and Dall'erba 2006, Pede et al. 2007). These methods are rather ad hoc and assume an abrupt transition between groups, which is not always realistic. Other studies propose endogenous procedures based on the regression tree method (Johnson and Durlauf 1995) or a predictive density approach (Canova 2004). Unlike the previous methods, the latter allow for a relatively gradual transition between groups and the number of groups is endogenously determined rather than determined a priori and exogenously. Recently O'hUallachain (2007) proposed an endogenous procedure based on principal components analysis and cluster analysis to identify convergence clubs as well as transition clubs. Their procedure assumes that growth transitions are uniform, but in fact, it is also likely that the spatial dynamic of the growth process is gradual, and follows a smooth transition. In other words, nonlinear spatial dynamic may be the driving force in growth transitions.

This empirical application proposes a procedure to endogenously determine the convergence clubs for the economic growth analysis of U.S. counties, allowing for a gradual regime switching structure in the form of a smooth transition autoregressive process. The methodology described in section 3 is applied to per capita income data for 3074 U.S. counties provided by the Bureau of Economic Analysis (BEA) over the period 1969-2003 to determine convergence clubs.

6.2. Estimation and Testing

We first start with estimation of parameters in the spatial logistic STAR model (SLSTAR) in equation (9). Following the neoclassical growth theory we consider an unconditional growth model as in Rey and Montouri (1999). The real per capita income growth over the period 1969-2003 is expressed as a function of the real per capita income in the initial year 1969, and we consider a smooth transition of the growth which is governed by the transition function G. The model reads as:

[6] The steady state is the state where capital and output are no longer growing over time.

$$\ln\left[\frac{y_{2003}}{y_{1969}}\right] = \alpha_0 + \alpha_1 \ln(y_{1969}) + (\delta_0 + \delta_1 \ln(y_{1969}))G[W\ln(y_{1969}), \gamma, c] + \varepsilon \qquad (24)$$

where y_{2003} and y_{1969} represents the real per capita income in 1969 and 2003, respectively. The errors ε are assumed to be independently and identically distributed with mean zero and a constant variance. The parameters in this model are estimated using the optimization routine in R. Results of the estimation of parameters in equation (24) are summarized in Table 3.

The results presented in Table 3 indicate that there are two regimes delimited by a threshold value or location parameter (c = 9.96). When the average per capita income at the neighbors in the initial year is less than the threshold value, then the region belongs to the first regime. When it is higher than the threshold value, then the region belongs to the second regime. The transition from one regime to another is determined by the smoothness parameter of magnitude 10.37. The parameters δ_0, δ_1, and γ are insignificant and they are also sensitive to the choice of initial parameters. However, the parameters α_0, α_1, and c are significant and appear to be robust to the choice of initial parameters. The negative coefficient on the initial level of per capita income denotes that there is β-convergence occurring in the growth process. Moreover, the significance of the location parameter may suggest the presence of potential nonlinearities in the growth process.

The transition function $G(Wx, \gamma, c)$ is represented in Figure 1. As expected, for low values of the average per capita income at the neighbors, the transition function is almost zero and the growth process is simply linear. But for large values of the average per capita income at the neighbors the transition function converges to one. Between the groups represented by high and low average per capita incomes, the growth transition is nonlinear. The low value of the estimated smoothness parameter explains why the growth transition process is relatively slow.

The above analysis suggests that there are two convergence clubs delimited by a threshold value, and also there is presence of nonlinearity which has yet to be tested. Following the procedure described above, we proceed to testing for nonlinearity in the above described spatial processes. Results pertaining to the nonlinearity tests in the spatial error and spatial lag models are presented in Table 4. Starting with the spatial error model, the null hypothesis $H_0 : \lambda = 0 \text{ and } \beta_2 = \beta_3 = 0$ was rejected at 5%. The individual tests on the null hypothesis $H_0 : \beta_2 = \beta_3 = 0$ and $H_0 : \lambda = 0$ were also rejected at 5%. This indicates that there is presence of autoregressive errors and nonlinearity in the growth process. Similar results were obtained when we consider the spatial lag model. The coefficient of the spatial lag is significant and there is also presence of nonlinearity.

These results suggest that the growth process might appropriately be explained by a spatial lag process with autoregressive disturbances incorporating nonlinearity in the form of smooth transition autoregressive. These results are still preliminary and further consideration and extension are expected in the future.

7. Conclusion

This paper investigates nonlinearity in two spatial process models: the spatial error and the spatial lag. A gradual regime-switching structure is allowed in the spatial processes in the form of smooth transition autoregressive. A procedure has been proposed to test for the presence of autoregressive parameters and/or of nonlinearity. Monte Carlo results indicate that the proposed tests have high power in general, in particular when the smoothness parameter is large. With regard to size, all the tests behave relatively well, with a size of about 10%. The empirical application on the economic growth of U.S. counties indicates two convergence clubs, and evidence of the presence of nonlinearity in the spatial growth transition process has been confirmed in the proposed tests. In addition to the strong evidence of nonlinearity, both spatial error and spatial lag were found to be significant, suggesting that there is spatial dependence in the growth process.

Results presented in this paper are still preliminary. Further investigation in the testing procedure is needed to substantiate the results. Also, there are several possibilities to extend this research. First, the transition variable in the transition function could be a lagged dependent variable instead of the lagged independent. Second, we could consider a third order Taylor series approximation of the transition function as suggested by Luukkonen et al. (1988). Third, we could also consider a spatial STAR model which combine spatial lag and spatial error in the form of ARAR model. Finally, we could also consider different specification of the weight matrix.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

Anselin, L. (2003). "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics" *International Regional Science Review* 26:153-66.

Baharumshah A. Z. and V. K. Liew (2006). "Forecasting Performance of Exponential Smooth Transition Autoregressive Exchange Rates Models" *Open Economies Reviews* 17:235-251.

Conley, G. T. and F. Molinari (2007). "Spatial Correlation Robust Inference with Errors in Location or Distance" *Journal of Econometrics* 140:76-96.

Conova, F. (2004) "Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach" *International Economic Review* 45:49-77.

Florax, R.J.G.M. and Nijkamp, P. (2005). "Misspecification in linear spatial regression models", in K. Kempf-Leonard (ed.), Encyclopedia of Social Measurement, Academic Press, San Diego.

Holt, T.M. and L. A.Craig (2006). "Nonlinear Dynamic and Structural change in the U.S. Hog-corn Cycle: A Time-Varying STAR Approach" *American Journal of Agricultural Economics* 88:215-233.

Johnson, A. and Durlauf S.N. (1995). "Multiple Regimes and Cross-Country Growth Behaviour" *Journal of Applied Econometrics* 10:365-384.

Kelejian, H. H. Prucha, I.R and Y. Yuzefovich (2004). "Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results" in J. LeSage and R.K. Pace, Advances in Econometrics, New York: Elsevier, pp. 163-198.

Kelejian H.H.,. Prucha I.R and A. Drukker, 2007. "A Spatial Cliff-Ord-Type Model with Heterokedastcity Innovations: Small and Large Samples Results".

Keynes, J.M. (1936). "The General Theory of Employment, Interest, and Money". London MacMillan.

Lee, L. (2002). "Consistency and Efficiency of Least Squarres Estimation for Mixed Regressive, Spatial Autoregressive Models" *Econometric Theory* 18:252-277.

Lee, L (2007). "The Method of Elimination and Substitution in the GMM Estimation of Mixed Regressive, Spatial Autoregressive Models" *Journal of Econometrics* 140:155-189.

Le Gallo, J and Dall'erba S. (2006). "Evaluating the Temporal and Spatial Heterogeneity of the European Convergence Process, 1980-199" *Journal of Regional Science* 46:269-288.

Luukkonen, R., P. Saikkonen, and T. Terasvirta. (1988). "Testing Linearity against Smooth Transition Autoregressive Models." *Biometrika* 75:491–9.

Mitchell, W.C., (1927). Business Cycles. The Problem and its Setting. NBER, New York.

Paelinck and Klaassen (1979). Spatial Econometrics. Farnborough: Saxon House.

O'hUallachain, B. (2007). "Regional Growth Transition Clubs in the United States" *Papers in Regional Science* 87:33-53

Pede, O. V., R.G.M. Florax and H. de Groot (2008 forthcomming). "The Role of Knowledge Externalities in the Spatial Distribution of Economic Growth: A Spatial Econometric Analysis for US Counties, 1969–2003" *Annals of Economics and Statistics*.

Rey, S.J. and B.D. Montouri 1999. "U.S. Regional Income Convergence: A Spatial Econometric Perspective" *Regional Studies* **33**:143–156.

Skalin, J and T. Terasvirta (2002). "Modeling Asymmetries and Moving Equilibria in Unemployment Rates" *Macroeconomic Dynamics* 6:202-241.

Terasvirta T. and Anderson (1992). "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models" *Journal of Applied Econometrics* 7:119-136.

Van Dijk, D. and P.H. Franses (2000). "Nonlinear Error-Correction Models for Interest Rates in the Netherlands" in W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim, and A. Würtz (ed.) Nonlinear Econometric Modelling-Proceedings of the 6th (EC)2 meeting, Cambridge: Cambridge University Press, pp. 203-227.

Table 1a: Testing for spatial lag assuming nonlinearity

	gamma						
rho	0	0.1	0.3	0.5	1	10	100
0	0.09	0.08	0.07	0.10	0.09	0.09	0.07
0.1	0.33	0.31	0.30	0.31	0.34	0.34	0.35
0.2	0.82	0.84	0.81	0.82	0.82	0.81	0.83
0.3	0.99	0.98	0.99	0.99	0.99	0.99	0.99
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 1b: Testing for nonlinearity assuming the presence of spatial lag

	gamma						
rho	0	0.1	0.3	0.5	1	10	100
0	0.07	0.15	0.54	0.85	0.98	1.00	1.00
0.1	0.08	0.14	0.55	0.86	0.99	1.00	1.00
0.2	0.07	0.13	0.56	0.84	0.99	1.00	1.00
0.3	0.08	0.15	0.53	0.85	0.98	1.00	1.00
0.5	0.07	0.12	0.55	0.84	0.97	1.00	1.00
0.7	0.06	0.14	0.54	0.86	0.99	1.00	1.00
0.9	0.08	0.15	0.55	0.88	0.98	1.00	1.00

Table 1c: Testing for both nonlinearity and spatial lag

	gamma						
rho	0.00	0.10	0.30	0.50	1.00	10.00	100.00
0	0.12	0.19	0.54	0.84	0.98	1.00	1.00
0.1	0.40	0.49	0.78	0.94	0.99	1.00	1.00
0.2	0.89	0.91	0.97	0.99	1.00	1.00	1.00
0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2a: Testing for spatial error assuming nonlinearity

	gamma						
lambda	0	0.1	0.3	0.5	1	10	100
0	0.09	0.08	0.09	0.07	0.07	0.09	0.08
0.1	0.32	0.32	0.31	0.33	0.33	0.31	0.31
0.2	0.79	0.81	0.83	0.82	0.82	0.83	0.82
0.3	0.98	0.99	0.98	0.98	0.99	0.99	0.99
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2b: Testing for nonlinearity assuming the presence of spatial error

	gamma						
lambda	0	0.1	0.3	0.5	1	10	100
0	0.10	0.19	0.60	0.88	0.99	1.00	1.00
0.1	0.10	0.20	0.61	0.90	0.99	1.00	1.00
0.2	0.13	0.22	0.61	0.90	1.00	1.00	1.00
0.3	0.13	0.21	0.60	0.85	0.99	1.00	1.00
0.5	0.20	0.27	0.62	0.86	0.98	0.99	1.00
0.7	0.30	0.34	0.60	0.80	0.95	0.98	0.98
0.9	0.45	0.48	0.55	0.66	0.80	0.86	0.89

Table 2c: Testing for both nonlinearity and spatial error

	gamma						
lambda	0	0.1	0.3	0.5	1	10	100
0	0.11	0.19	0.57	0.86	0.99	1.00	1.00
0.1	0.31	0.38	0.69	0.90	1.00	1.00	1.00
0.2	0.74	0.81	0.93	0.98	1.00	1.00	1.00
0.3	0.98	0.98	0.99	1.00	1.00	1.00	1.00
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3: Estimated parameters of the spatial STAR model

	coefficients	standard errors	t-value
α_0	5.12	0.76	6.74
α_1	-0.48	0.08	-6.00
δ_0	-7.94	8.59	-0.92
δ_1	0.81	0.86	0.94
γ	10.37	29.23	0.35
c	9.96	0.07	142.29

Table 4: Nonlinearity tests in growth process of U.S. counties

Models	Null Hypothesis	LM test value	χ^2 critical value 5%
Spatial lag	H_0: rho = 0 and $\beta_2=\beta_3=0$	1729.98	7.82
	H_0: rho = 0	950.80	3.84
	H_0: $\beta_2=\beta_3=0$	102.10	5.99
Spatial error	H_0: lambda = 0 and $\beta_2=\beta_3=0$	1059.17	7.82
	H_0: lambda = 0	942.18	3.84
	H_0: $\beta_2=\beta_3=0$	57.73	5.99

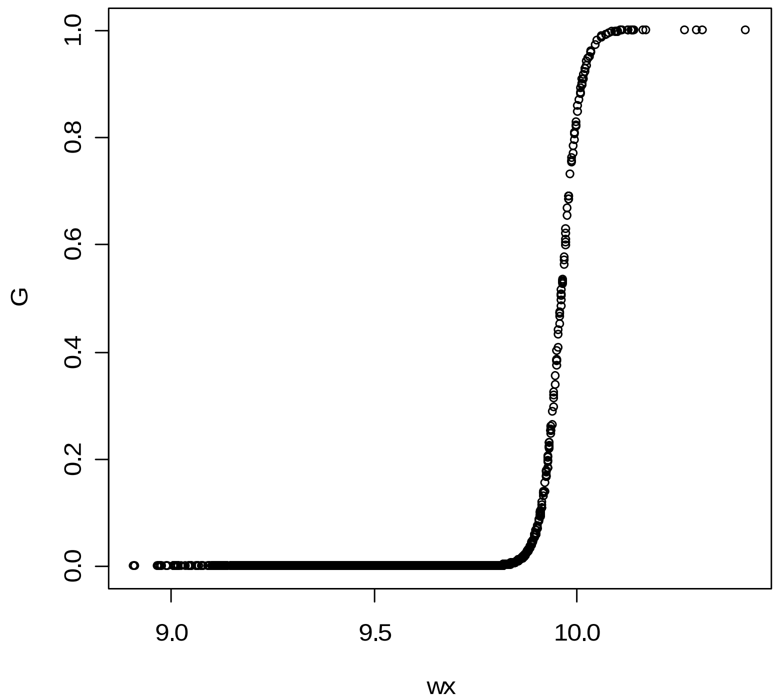

Figure 1: The Logistic Transition Function G against Wx

Appendix 1.

Considering a Logistic STAR model given as:

$$y = (\alpha_0 + \alpha_1 x) + (\delta_0 + \delta_1 x) G(wx, \gamma, c) + \varepsilon$$

The first order Taylor series approximation reads as:

$$\begin{aligned} G(wx, \gamma, c)_r &= \frac{1}{1 + \exp\left[-\gamma\left((wx)_r - c\right)\right]} \\ &\approx \frac{1}{2} + \frac{\left((wx)_r - c\right)}{4}\gamma \\ &= \frac{1}{2} - \frac{c\gamma}{4} + \frac{\gamma}{4}(wx)_r \\ &\equiv \lambda_0 + \lambda_1 (wx)_r \end{aligned}$$

Substituting back into the equation of the STAR model gives:

$$\begin{aligned} y_r &= \alpha_0 + \alpha_1 x_r + [\delta_0 + \delta_1 x_r][\lambda_0 + \lambda_1 (wx)_r] + \mu_r \\ &= \alpha_0 + \alpha_1 x_r + \lambda_0 [\delta_0 + \delta_1 x_r] + \lambda_1 (wx)_r [\delta_0 + \delta_1 x_r] + \mu_r \\ &= [\alpha_0 + \lambda_0 \delta_0] + [\alpha_1 + \lambda_0 \delta_1] x_r + \lambda_1 \delta_0 (wx)_r + \lambda_1 \delta_1 (wx)_r x_r + \mu_r \\ &= \left[\alpha_0 + \frac{1}{2}\delta_0 - \frac{c\gamma}{4}\delta_0\right] + \left[\alpha_1 + \frac{1}{2}\delta_1 - \frac{c\gamma}{4}\delta_1\right] x_r + \frac{\gamma}{4}\delta_0 (wx)_r + \frac{\gamma}{4}\delta_1 (wx)_r x_r + \mu_r \\ &\equiv \beta_0 + \beta_1 x_r + \beta_2 (wx)_r + \beta_3 (wx)_r x_r + \mu_r \end{aligned}$$

Appendix 2a.

Developing LM test for nonlinearity in the spatial Lag Model using Maximum Likelihood

The following proof shows the derivation of the LM test for testing the null hypothesis

$H_0 : \rho = 0 \textit{ and } \beta = \beta^*$

Considering the model

$y = \rho Wy + X\beta + \varepsilon$

where y is an $N \times 1$vector of observations on the dependent variable, W is an $N \times N$ weight matrix which defines the spatial structure of regions, ρ is the autoregressive coefficient, X an $N \times k$ vector of independent variables, β is $k \times 1$ vector associated with the independent variables, and ε is an $N \times 1$vector representing the error tern and is distributed with mean 0 and variance σ^2

Using the simplification suggested by Ord (1975), the log-likelihood function may be written as:

$$L = \sum_i \ln(1 - \rho w_i) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{(y - \rho Wy - X\beta)^T (y - \rho Wy - X\beta)}{2\sigma^2}$$

From the FOC, the ML estimates of β and σ^2 in a spatial lag model are obtained as:

$$\beta_{ML} = (X^T X)^{-1} X^T (I - \rho W) y$$

and

$$\sigma^2 = \frac{(y - \rho Wy - X\beta_{ML})^T (y - \rho Wy - X\beta_{ML})}{N}$$

Substituting β and σ^2 in the likelihood function yield the concentrated form given as follow:

$$L = \sum_i \ln(1 - \rho w_i) - \frac{N}{2}\ln\left[\frac{(e_0 - \rho e_L)^T (e_0 - \rho e_L)}{N}\right]$$

Where e_0 and e_L are residuals in a regression of y on X and Wy on X , respectively.

The asymptotic variance matrix follows as the inverse of the information matrix

$$AsyVar = \begin{bmatrix} tr[W_\rho]^2 + tr[W_\rho{}^T W_\rho] + \frac{[W_\rho X\beta]^T [W_\rho X\beta]}{\sigma^2} & \frac{[X^T W_\rho X\beta]^T}{\sigma^2} & \frac{tr(W_\rho)}{\sigma^2} \\ \frac{[X^T W_\rho X\beta]}{\sigma^2} & \frac{X^T X}{\sigma^2} & 0 \\ \frac{tr(W_\rho)}{\sigma^2} & 0 & \frac{N}{2\sigma^4} \end{bmatrix}^{-1}$$

Where $W_\rho = W(I - \rho W)^{-1}$

Taking the FOC

$$\frac{\partial LogL}{\partial \beta} = -\frac{1}{\sigma^2}[y - \rho Wy - X\beta]^T [-X] = \frac{1}{\sigma^2}[e]^T X$$

$$\frac{\partial LogL}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}[e]^T [e]$$

$$\frac{\partial LogL}{\partial \rho} = tr[(I - \rho W)^{-1}(-W)] - \frac{1}{\sigma^2}[e]^T [-Wy] = -tr[W(I - \rho W)^{-1}] + \frac{1}{\sigma^2}[e]^T [Wy]$$

Evaluated at the restricted values $\rho = 0$ *and* $\beta = \beta^*$ the FOC becomes

$$\frac{\partial LogL}{\partial \beta} = \frac{1}{\sigma^2}[y - X\beta^*]^T X = \frac{1}{\sigma^2}[e^*]^T X$$

$$\frac{\partial LogL}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}[y - X\beta^*]^T [y - X\beta^*]$$

$$\frac{\partial LogL}{\partial \rho} = \frac{1}{\sigma^2}[y - X\beta^*]^T [Wy] = \frac{1}{\sigma^2}[e^*]^T [Wy]$$

The asymptotic variance matrix evaluated at the restricted value $\rho = 0$ *and* $\beta = \beta^*$ becomes:

$$AsyVar = \begin{bmatrix} tr\left[W^2 + W^T W\right] + \frac{\left[WX\beta^*\right]^T \left[WX\beta^*\right]}{\sigma^2} & \frac{\left[X^T WX\beta^*\right]^T}{\sigma^2} & \frac{tr(W)}{\sigma^2} \\ \frac{\left[X^T WX\beta^*\right]}{\sigma^2} & \frac{X^T X}{\sigma^2} & 0 \\ \frac{tr(W)}{\sigma^2} & 0 & \frac{N}{2\sigma^4} \end{bmatrix}^{-1}$$

$$LM = \begin{bmatrix} \frac{1}{\sigma^2}\left[e^*\right]^T X \\ \frac{1}{\sigma^2}\left[e^*\right]^T \left[Wy\right] \end{bmatrix}^T \begin{bmatrix} \frac{X^T X}{\sigma^2} & \frac{\left[X^T WX\beta^*\right]}{\sigma^2} \\ \frac{\left[X^T WX\beta^*\right]^T}{\sigma^2} & tr\left[W^2 + W^T W\right] + \frac{\left[WX\beta^*\right]^T \left[WX\beta^*\right]}{\sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sigma^2}\left[e^*\right]^T X \\ \frac{1}{\sigma^2}\left[e^*\right]^T \left[Wy\right] \end{bmatrix}$$

$$= \frac{\left[\frac{\left[\left[e^*\right]^T X\right]\left[X^T X\right]^{-1}\left[X^T WX\beta^*\right]}{\sigma^2} - \left[\frac{1}{\sigma^2}\left[e^*\right]^T \left[Wy\right]\right] \right]^2}{\left[T + \frac{\left[WX\beta^*\right]^T M \left[WX\beta^*\right]}{\sigma^2} \right]} + \frac{\left[\left[e^*\right]^T X\right]\left[X^T X\right]^{-1}\left[X^T e^*\right]}{\sigma^2}$$

Appendix 2b.

Developing LM test for nonlinearity in the spatial error model using Maximum Likelihood

The following proof shows the derivation of the LM test for testing the null hypothesis

$H_0 : \rho = 0 \; and \; \beta = \beta^*$

Considering the model

$y = X\beta + \varepsilon$ With $\varepsilon = \lambda W\varepsilon + \mu$

where y is an $N \times 1$ vector of observations on the dependent variable, W is an $N \times N$ weight matrix which defines the spatial structure of regions, and μ is an $N \times 1$ vector of random error terms distributed with mean 0 and variance σ^2 while ε is an $N \times 1$ vector of random error terms distributed with mean 0 and nonspherical variance-covariance matrix $\Omega = (I - \lambda W)^{-1}(I - \lambda W')^{-1}$.

The log likelihood is given as:

$$LogL = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) + \ln|I - \lambda W| - \frac{1}{2\sigma^2}[y - \lambda Wy - X\beta + \lambda WX\beta]^T [y - \lambda Wy - X\beta + \lambda WX\beta]$$

The First Order Conditions are given as:

$$\frac{\partial LogL}{\partial \beta} = -\frac{1}{\sigma^2}[y - \lambda Wy - X\beta + \lambda WX\beta]^T [-(I - \lambda W)X]$$

$$= \frac{1}{\sigma^2} e^T [(I - \lambda W)X]$$

$$\frac{\partial LogL}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} e^T e$$

$$\frac{\partial LogL}{\partial \lambda} = tr[(I - \lambda W)^{-1}(-W)] + \frac{1}{\sigma^2}[y - \lambda Wy - X\beta + \lambda WX\beta]^T [W(y - X\beta)]$$

The block diagonal asymptotic variance-covariance matrix is given as:

$$\begin{bmatrix} \frac{1}{\sigma^2}\left[(X-\lambda WX)^T(X-\lambda WX)\right] & 0 & 0 \\ 0 & \frac{N}{2\sigma^4} & \frac{tr\left[W(I-\lambda W)^{-1}\right]}{\sigma^2} \\ 0 & \frac{tr\left[W(I-\lambda W)^{-1}\right]}{\sigma^2} & tr\left[W(I-\lambda W)^{-1}\right]^2 + tr\left[\left[W(I-\lambda W)^{-1}\right]^T\left[W(I-\lambda W)^{-1}\right]\right] \end{bmatrix}^{-1}$$

At the restricted values $\beta = \beta^* \ and \ \lambda = 0$ the First Order Condition are given as:

$$\frac{\partial LogL}{\partial \beta} = \frac{1}{\sigma^2}\left[y - X\beta^*\right]^T\left[(I-\lambda W)X\right] = \frac{1}{\sigma^2} e^{*T} X$$

$$\frac{\partial LogL}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} e^{*T} e^*$$

$$\frac{\partial LogL}{\partial \lambda} = \frac{1}{\sigma^2}\left[(y - X\beta^*)\right]^T W\left[y - X\beta^*\right] = \frac{1}{\sigma^2} e^{*T} W e^*$$

And the asymptotic variance-covariance matrix as:

$$\begin{bmatrix} \frac{1}{\sigma^2} X^T X & 0 & 0 \\ 0 & \frac{N}{2\sigma^4} & 0 \\ 0 & 0 & tr\left[W^2 + W^T W\right] \end{bmatrix}^{-1}$$

The LM test for the null hypothesis $\beta = \beta^* \ and \ \lambda = 0$ is derived as:

$$LM = \begin{bmatrix} \frac{1}{\sigma^2} e^{*T} X \\ \frac{1}{\sigma^2} e^{*T} W e^{*} \end{bmatrix} \begin{bmatrix} \left[\frac{1}{\sigma^2} X^T X \right]^{-1} & 0 \\ 0 & \left[tr\left[W^2 + W^T W \right] \right]^{-1} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} e^{*T} X \\ \frac{1}{\sigma^2} e^{*T} W e^{*} \end{bmatrix}^T$$

$$= \left[\frac{1}{\sigma^2} e^{*T} X \right] \left[\frac{1}{\sigma^2} X^T X \right]^{-1} \left[\frac{1}{\sigma^2} e^{*T} X \right]^T + \left[\frac{1}{\sigma^2} e^{*T} W e^{*} \right] \left[tr\left[W^2 + W^T W \right] \right]^{-1} \left[\frac{1}{\sigma^2} e^{*T} W e^{*} \right]^T$$

$$= \frac{\left[e^{*T} X \right] \left[X^T X \right]^{-1} \left[X^T e^{*} \right]}{\sigma^2} + \frac{\left[\frac{1}{\sigma^2} e^{*T} W e^{*} \right]^2}{\left[tr\left[W^2 + W^T W \right] \right]}$$