

SPATIAL ECONOMETRIC STAR MODELS[†]

Raymond J.G.M. Florax^{1,2}, Valerien O. Pede¹, and Matthew T. Holt¹

¹ Department of Agricultural Economics
Purdue University
403 West State Street
West Lafayette, IN 47907-2056, USA
E-mail: vpede@purdue.edu,
rflorax@purdue.edu, mholt@purdue.edu

² Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1085 HV Amsterdam, The Netherlands

Abstract. Spatial regression models incorporating non-stationarity in the regression coefficients are popular. In this paper we propose a family of spatial Smooth Transition AutoRegressive (STAR) models inspired by analogous nonlinear approaches developed in the time series literature. Spatial STAR models constitute a parsimonious, easy-to-estimate approach to modeling nonlinear spatial parameter variation and endogenous detection of spatial regimes. A distinct advantage of the approach is the integration of modeling spatial dependence and heterogeneity in an integrated model where tradeoffs between the two can be assessed in a valid statistical framework. The STAR approach is especially useful given the observational equivalence of spatial autocorrelation and spatial heterogeneity. We derive the necessary Lagrange Multiplier tests to identify nonlinearity and spatial autocorrelation individually as well as jointly. Monte Carlo simulation experiments are used to assess the small sample performance of the LM tests, including the probability of identifying the correct underlying model. Uncomplicated maximum likelihood estimation and inference procedures are demonstrated in an example using data on economic growth for United States counties.

Keywords: spatial autoregression, smooth transition, spatial econometrics, STAR, GWR
JEL Codes: C21, C51, R11, R12

[†] The authors would like to thank Dayton M. Lambert for valuable comments.

1. Introduction

In regression models utilizing spatial cross-sectional data authors typically pay attention to spatial heterogeneity and spatial dependence. Because modeling spatial dependence requires tailored techniques it has gathered substantial scrutiny. Spatial heterogeneity or non-stationarity occurs if the modeled relationships vary systematically over space. This variation can be captured in the coefficients (spatial regimes or trends), the error variance (heteroskedasticity), the functional form, or some combination of these (Anselin, 1988).

Geographically Weighted Regression (GWR) has popularized the notion of spatial variation in the regression parameters (Fotheringham et al., 2002). In GWR each parameter is determined locally by exploiting the variation in distance-weighted subsets of neighbors. This approach has several, potentially serious, drawbacks (Cho et al., 2009). For instance, repeated usage of overlapping neighborhood sets and the conditional nature of the regression coefficients induced by prior kernel selection and bandwidth calibration challenge statistical inference based on standard least squares principles (LeSage, 2004). In addition, neighborhood sets need to be large for the estimator to be efficient. This creates a delicate tradeoff between efficiency and the desired (lack of) smoothness of the parameter surface.

In comparison, the use of spatial regimes is statistically unproblematic and more parsimonious, but requires a potentially *ad hoc* division of the sample into a disjoint, possibly predetermined number of groups. This issue has surfaced prominently in the economic growth literature in an attempt to validate the existence of convergence clubs. Solutions vary from exogenous fixes using cutoff points or exploratory (spatial) data analysis, to endogenous approaches based on regression trees, predictive density, or stationary tests (Abreu et al., 2005). In a spatial setting Dall'erba et al. (2008) propose a series of tests to endogenize regime choice,

and Basile and Gress (2005) and Basile (2008) suggest a semi-parametric approach that endogenizes regime selection and simultaneously accounts for spatial processes.

In this paper we propose a spatial version of the Smooth Transition AutoRegressive (STAR) modeling technique as an obvious pendant to the regime-switching approaches developed in time series econometrics (Teräsvirta, 1994). The spatial STAR model provides a lucid framework for statistical inference on the extent of spatial parameter variation and endogenously determines a (potentially) smooth transition between regimes.

2. Spatial STAR model

A family of spatial STAR models can be developed from the basic spatial autoregressive error specification:

$$y = X\beta + (I - \rho W)^{-1}\mu, \quad (1)$$

where y is an $(n \times 1)$ spatial data series, X an $(n \times k)$ matrix of explanatory variables, μ a vector of innovations, and W an $(n \times n)$ spatial weights matrix. Equation (1) can be extended to include the spatially lagged dependent variable Wy , resulting in the so-called spatial ARAR model. It may also contain one or more linearly independent spatial cross-regressive terms taken from WX . The parameters in β are stationary across space, except if the specification contains the spatially lagged dependent variable, which forces the parameters into a smooth autoregressive pattern due to the presence of the spatial multiplier in the reduced form (Anselin, 2006).

A spatial error STAR model can be easily constructed from (1) by adding a set of coefficients δ for a second regime, interacted with a transition function:

$$y = X\beta + X\delta \circ G(s; \gamma, c) + (I - \rho W)^{-1} \mu, \quad (2)$$

where $G(s; \gamma, c)$ is a potentially smooth, real-valued transition function bounded between zero and one, γ and c are respectively the slope and location parameter, and \circ is the Hadamard product (element-by-element multiplication). In time series, (2) is often parameterized by using a logistic or exponential function, with a lagged endogenous variable, an exogenous variable, or a time trend identified as the transition variable, s . Analogous definitions of the transition variable can be used in a spatial setting, although the use of the spatially lagged dependent variable results in a highly nonlinear specification. Employing a spatially lagged exogenous variable or a polynomial in the coordinates of the observations is, however, straightforward.

In this paper we use the logistic function with a spatially lagged exogenous variable Wx as the transition variable:

$$G(Wx; \gamma, c) = \left[1 + \exp(-\gamma(Wx - c)/\sigma_{Wx}) \right]^{-1}, \quad (3)$$

where the exponential part is scaled by the standard deviation of Wx to facilitate estimation.

The spatial error STAR model is an attractive alternative to GWR. The specification is parsimonious, and exhibits data-determined, non-stationary coefficients across space. In the extreme case, if γ and/or δ are not statistically different from zero, the coefficients β are “global” in the GWR terminology. Alternatively, for large γ the transition function approaches 0 or 1 depending on whether $(Wx - c) \lesseqgtr 0$, effectively creating one or two clearly separated spatial

regimes. In general, the transition function G will have a smooth logistic shape causing the coefficients to be predominantly “local”.

For estimation we use standard maximum likelihood principles for spatial process models (Anselin, 2006). If the errors are *iid* normal with mean 0 and variance σ^2 the log-likelihood function for the spatial error STAR model is:

$$L(\theta, \rho, \sigma^2) = -\frac{n}{2} \ln \sigma^2 + \ln |I - \rho W| - \frac{1}{2\sigma^2} \mu' \mu, \quad (4)$$

where $\mu = (I - \rho W)(y - f(X; \theta))$, f is the function provided in (2) and (3) with parameters $\theta = (\beta', \delta', \gamma, c)'$, and $\sigma^2 = \mu' \mu / n$. Since there is no analytical expression for θ , optimization cannot be based on a “one-shot” maximization of the concentrated likelihood. An iterative feasible generalized least-squares (GLS) approach is, however, appropriate as long as a consistent estimate of ρ is attained through nonlinear optimization of (4) given μ . The estimation procedure is based on the following steps:

1. estimate $y = X\beta + X\delta \circ G(Wx; \gamma, c) + \mu$ using an appropriate nonlinear least squares estimator and obtain the estimated residuals $\hat{\mu}$;
2. substitute $\hat{\mu}$ in (4) and optimize to obtain a consistent estimate for ρ ;
3. use $\hat{\rho}$ to employ the Cochrane-Orcutt transformations $\check{y} = (I - \hat{\rho}W)y$ and $\check{X} = (I - \hat{\rho}W)X$; and
4. repeat the above steps with transformed variables \check{y} and \check{X} until convergence.

The information matrix is block-diagonal, implying invariance between θ and the covariance parameters ρ and σ^2 . Under appropriate regularity conditions asymptotic inference for the spatial autoregressive parameter can be based on the lower right elements of:

$$\text{AsyVar}(\theta, \sigma^2, \rho) = \begin{bmatrix} \frac{1}{\sigma^2} f(W_A X; \theta)' f(W_A X; \theta) & 0 & 0 \\ 0 & \frac{n}{2\sigma^4} & \frac{\text{tr}(W_B)}{\sigma^2} \\ 0 & \frac{\text{tr}(W_B)}{\sigma^2} & \text{tr}(W_B)^2 + \text{tr}(W_B' W_B) \end{bmatrix}^{-1}, \quad (5)$$

where $W_A = I - \rho W$, and $W_B = W(I - \rho W)^{-1}$.

The specification of the spatial error STAR model cannot be used directly to derive a nonlinearity test utilizing conventional maximum likelihood theory because of the presence of unidentified nuisance parameters under the null hypothesis. In other words, the asymptotic distribution for γ is non-standard. Luukkonen et al. (1988) suggest using a first order Taylor series approximation of G around $\gamma=0$, which in the spatial error STAR case reads as:

$$\begin{aligned} y &= X\beta + Wx\varphi_1 + (Wx \circ x)\varphi_2 + (I - \rho W)^{-1} \mu \\ &= X\beta + Z\varphi + (I - \rho W)^{-1} \mu, \end{aligned} \quad (6)$$

where $Z = [Wx, Wx \circ x]$, and $\varphi = (\varphi_1, \varphi_2)'$ with the individual φ 's being functions of the original parameters in (2) and (3). Lagrange Multiplier (LM) tests for spatially autoregressive errors and

nonlinearity can be easily derived following the general principles outlined in Anselin et al. (1996):

$$LM_{\rho=0} = \frac{1}{T} \left(\frac{e'We}{\sigma^2} \right)^2, \quad (7)$$

$$LM_{\varphi=0} = \frac{e'Pe}{\sigma^2}, \quad (8)$$

$$LM_{\rho=\varphi=0} = \frac{1}{T} \left(\frac{e'We}{\sigma^2} \right)^2 + \frac{e'Pe}{\sigma^2} = LM_{\rho=0} + LM_{\varphi=0}, \quad (9)$$

where e are the residuals of the model under the null hypothesis estimated using an adequate GLS estimator, $\sigma^2 = e'e/n$, $T = \text{tr}((W' + W)W)$, and P is the projection matrix $\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ with $\tilde{X} = [X, Z]$. The tests are asymptotically distributed with 1, 2 and 3 degrees of freedom, respectively. Details about the derivations of the tests are given in Appendix 1.

3. Application to economic growth

We demonstrate these estimation and inference procedures for an unconditional neoclassical growth model where the annual growth of per capita income is regressed on the level of *per capita* income, both in logarithmic form and with nominal values deflated by a regional consumer price index series for four regions in the US (2003 = 100). The dataset comprises observations for 3,074 counties in the lower 48 states of the U.S., for 1963 and 2003. Income data are available from the Bureau of Economic Analysis (BEA), and price data from the Bureau of Labor Statistics (BLS). We focus on parameter variation and spatial autocorrelation, and use a second-order contiguity matrix of the queen type (lower orders included). Heteroskedasticity of

the innovations is ignored, but can be easily incorporated using a robust covariance estimator. Computer code for the example is given in Appendix 2.

Table 1
Different specifications and estimators for the unconditional growth model, U.S. counties, 1969–2003^a

	OLS	Spatial STAR	Spatial Error	Spatial Error STAR
β_0 , constant_1	4.622** (0.104)	4.984** (0.132)	3.937** (0.141)	4.365** (0.166)
δ_0 , constant_2		-11.768** (2.666)		-8.297** (1.925)
β_1 , ln(GDP ₁₉₆₉)_1	-0.428** (0.011)	-0.465** (0.014)	-0.356** (0.015)	-0.402** (0.017)
δ_1 , ln(GDP ₁₉₆₉)_2		1.189** (0.268)		0.849** (0.197)
γ , slope parameter		2.020** (0.421)		3.271** (1.204)
c , location parameter		9.909** (0.054)		9.933** (0.041)
ρ , AR parameter			0.762** (0.023)	0.747** (0.004)
convergence rate ^b	1.64 (fixed)	1.66 (0.69)	1.29 (fixed)	1.22 (0.69)
LM _{$\rho=0$}	2139.78**	1993.133**		
LM _{$\varphi=0$}	90.29**		35.57**	
LM _{$\rho=\varphi=0$}	2230.07**			

^a Standard errors in parentheses. The probability of falsely rejecting the null hypothesis is flagged by ** and * for $p \leq 0.01$, and $0.01 < p \leq 0.05$, respectively.

^b Mean annual convergence rate in percent, with the standard deviation in parentheses where appropriate. The convergence rate equals $100 \times (\ln(\beta_1+1))/-T$, where T is the length of the time period in years.

Table 1 shows that the annual convergence rate is 1.64% for the OLS estimator, but the LM tests provide strong evidence for misspecification. Individual tests as well as the joint test for spatially correlated errors and nonlinearity are significant. Omitting cross-regressive terms evokes omitted variable bias whereas ignoring spatially correlated errors merely causes the OLS estimator to be inefficient. The preferable strategy is therefore to first estimate the spatial STAR model. The second column of results shows clear evidence in favor of this specification. All coefficients are significantly different from zero, resulting in a spatially varying annual rate of convergence, which on average equals 1.66%. The LM test for this model, however, points to

spatially correlated errors. Subsequent estimation of a spatial error STAR model shows that ignoring spatial error correlation clearly leads to inflated t -values. The magnitude of the coefficients is similar—as expected. An alternative specification strategy, based on first correcting for spatially correlated errors and subsequent testing for erroneously omitted variables, leads to the same specification.

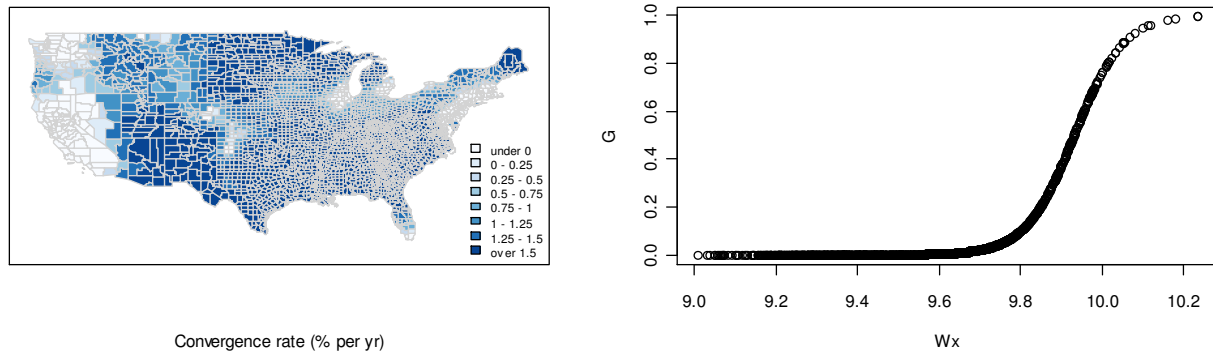


Figure 1 Convergence rates in $\% \text{ yr}^{-1}$ for U.S. counties (map) and estimated transition function (diagram) based on estimates of the spatial error STAR model

The transition function in Figure 1 shows that the degree to which local variation in convergence rates is statistically warranted is actually rather limited in this example. The estimation results reveal diverging or slowly converging economies around the major conurbations on the East and West coast, Chicago, Detroit, Denver and southern Florida, and economies converging at the same speed in most of the rest of the country. The discernable spatial pattern in the map illustrates that the STAR technique does not “over-smooth” the results, nor does it introduce unwieldy local spatial parameter variation. The estimated annual convergence rate is 1.22%, and approximately two-thirds of the sample can statistically be associated with one spatial regime with an annual convergence rate of 1.51%.

4. Conclusions

Spatial STAR models constitute a parsimonious, easy-to-estimate approach to modeling spatial parameter variation and endogenous detection of spatial regimes. A distinct advantage of the approach is the integration of modeling spatial dependence and heterogeneity in an integrated model where tradeoffs between the two can be assessed in a valid statistical framework. This approach is especially useful given the observational equivalence of spatial autocorrelation and spatial heterogeneity.

The spatial STAR model can be extended to various other situations commonly addressed in spatial econometrics. Further research is warranted and could focus on alternative nonlinear transition functions possibly including a spatially lagged dependent variable, the extension to a multivariate setting with multiple regimes, alternative estimators (nonlinear GMM), and the coverage of space-time data.

Appendix 1: Lagrange Multiplier tests for nonlinearity and autoregressive errors

Consider a spatial error STAR model with a logistic transition function defined in terms of a spatially lagged exogenous variable Wx as the transition variable:

$$y = X\alpha + X\delta \circ G(Wx; \gamma, c) + (I - \rho W)^{-1}\mu, \tag{A1}$$
$$G = \left[1 + \exp(-\gamma(Wx - c)/\sigma_{Wx}) \right]^{-1},$$

where y is an $(n \times 1)$ dependent variable, X an $(n \times k)$ matrix of non-stochastic regressors, W an $(n \times n)$ weights matrix, μ an $(n \times 1)$ vector of independent and identically distributed errors, α and δ are $(k \times 1)$ vectors with unknown parameters, and γ , c and ρ unknown scalars. The

transition variable is a spatially lagged exogenous variable Wx , with x potentially being a column vector of the design matrix X . The exponent is scaled by σ_{Wx} , the standard deviation of the transition variable.

The first order Taylor series approximation of the transition function around $\gamma=0$ is:

$$G \approx \frac{1}{2} + \frac{Wx - c}{4} \frac{\gamma}{\sigma_{Wx}} = \eta_0 + \eta_1 Wx, \quad (\text{A2})$$

where $\eta_0 = (2\sigma_{Wx} - c\gamma) / 4\sigma_{Wx}$ and $\eta_1 = \gamma / 4\sigma_{Wx}$. Substitution provides the Taylor series approximation for the spatial error STAR model:

$$y \approx X\alpha + X\delta \circ (\eta_0 + \eta_1 Wx) + (I - \rho W)^{-1} \mu, \quad (\text{A3})$$

which, after collecting terms, can be compactly rewritten as:

$$\begin{aligned} y &= X\xi + (X \circ Wx)\varphi + (I - \rho W)^{-1} \mu \\ &= Z\beta + (I - \rho W)^{-1} \mu, \end{aligned} \quad (\text{A4})$$

where Z is the $(n \times 2k)$ matrix containing both X and $X \circ Wx$, and $\beta = (\xi', \varphi)'$ is a $(2k \times 1)$ vector defined in the original parameters as $\gamma = (\eta_0 \delta' + \alpha)'$ and $\varphi = (\eta_1 \delta)'$.

The first order approximation of the spatial error STAR model in (A4) is essentially a spatially autoregressive error model where the design matrix Z contains X supplemented by a

matrix containing the Hadamard product of X and the transition variable Wx . We can therefore use the maximum likelihood principles outlined in Anselin (1988, Chapter 6; 2006) to derive Lagrange Multiplier (LM) tests for nonlinearity and spatial autoregressive errors.

Assuming μ is independently and identically distributed as $N(0, \sigma^2 I)$ and Wx is independent from the errors, the joint probability density function of y is:

$$f(y) = (2\pi\sigma^2)^{-1/2} |\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \mu' \Omega^{-1} \mu\right), \quad (\text{A5})$$

where $\Omega = (I - \rho W)'(I - \rho W)$. From (A6), the log-likelihood function follows as:

$$\ell(\beta, \rho, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |W_A| - \frac{1}{2\sigma^2} \mu' W_A' W_A \mu, \quad (\text{A6})$$

where $W_A = I - \rho W$.

Defining the vector of unknown parameters as $\theta = (\beta', \rho, \sigma^2)'$, the score vector is given by:

$$d(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \beta} \\ \frac{\partial \ell(\theta)}{\partial \rho} \\ \frac{\partial \ell(\theta)}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} d_\beta \\ d_\rho \\ d_{\sigma^2} \end{bmatrix}, \quad (\text{A7})$$

and the information matrix reads as:

$$\mathfrak{I}(\theta) = -E \left[\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right] = -E \begin{bmatrix} d_{\beta}^2 & d_{\beta\rho}^2 & d_{\beta\sigma^2}^2 \\ d_{\rho\beta}^2 & d_{\rho}^2 & d_{\rho\sigma^2}^2 \\ d_{\sigma^2\beta}^2 & d_{\sigma^2\rho}^2 & d_{\sigma^2}^2 \end{bmatrix}. \quad (\text{A8})$$

The elements of the score vector containing first order derivatives are given by:

$$d_{\beta} = \frac{1}{\sigma^2} \mu' W_A' W_A Z, \quad (\text{A9.1})$$

$$d_{\rho} = -\text{tr}(W W_A^{-1}) + \frac{1}{\sigma^2} \mu' W_A' W \mu, \quad (\text{A9.2})$$

$$d_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \mu' W_A' W_A \mu. \quad (\text{A9.3})$$

The second order derivatives are:

$$d_{\beta}^2 = -\frac{1}{\sigma^2} Z' W_A' W_A Z, \quad (\text{A10.1})$$

$$d_{\rho}^2 = -\text{tr}(W_A^{-1} W W_A^{-1} W) - \frac{1}{\sigma^2} \mu' W' W \mu, \quad (\text{A10.2})$$

$$d_{\sigma^2}^2 = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \mu' W_A' W_A \mu, \quad (\text{A10.3})$$

$$d_{\beta\rho}^2 = -\frac{1}{\sigma^2} Z' W_A' W \mu - \frac{1}{\sigma^2} \mu' W_A' W Z, \quad (\text{A10.4})$$

$$d_{\beta\sigma^2}^2 = -\frac{1}{(\sigma^2)^2} \mu' W_A' W_A Z, \quad (\text{A10.5})$$

$$d_{\rho\sigma^2}^2 = -\frac{1}{(\sigma^2)^2} \mu' W_A' W \mu. \quad (\text{A10.6})$$

We can now straightforwardly derive individual and joint Lagrange Multiplier (LM) tests for nonlinearity and spatially autoregressive errors. Generally, the asymptotic LM test is defined as $LM = d(\theta) [\mathfrak{I}(\theta)]^{-1} d(\theta)'$, to be evaluated under the null hypothesis, where $d(\theta)$ is the score vector, $[\mathfrak{I}(\theta)]^{-1}$ the asymptotic variance-covariance matrix:

$$[\mathfrak{I}(\theta)]^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} Z' W_A' W_A Z & 0 & 0 \\ 0 & \text{tr}(W_B)^2 + \text{tr}(W_B' W_B) & \frac{\text{tr}(W_B)}{\sigma^2} \\ 0 & \frac{\text{tr}(W_B)}{\sigma^2} & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}, \quad (\text{A11})$$

with $W_B = W(I - \rho W)^{-1}$, and $\theta = (\beta', \rho, \sigma^2)'$ the parameter vector.

A1.1 Nonlinearity tests

The nonlinearity test is simply a standard test on omitted variables, where the omitted variables are $X \circ Wx$. Two null hypotheses can be defined, depending on whether the restricted model is the non-spatial model or the spatial error model. For the first case the parameter vector of the restricted model is $\theta_0 = (\xi', 0, 0, \sigma^2)'$ and the LM statistic for the null hypothesis $H_0 : \varphi = 0 | \rho = 0$ is:

$$\begin{aligned}
LM_{\varphi=0|\rho=0} &= \left(\frac{1}{\sigma^2} \mu'Z \right) \sigma^2 (Z'Z)^{-1} \left(\frac{1}{\sigma^2} \mu'Z \right)' \\
&= \frac{\mu'Z (Z'Z)^{-1} Z' \mu}{\sigma^2},
\end{aligned} \tag{A12}$$

where $\mu = y - X\xi$ corresponds with the OLS residuals of the restricted model, and $\sigma^2 = \mu'\mu/n$.

Similarly, for the case where the restricted model is the spatial error model, the parameter vector under the null hypothesis is given as $\theta_0 = (\xi', 0, \rho, \sigma^2)'$ and the LM statistic for the null hypothesis $H_0 : \varphi = 0 | \rho \neq 0$ is:

$$\begin{aligned}
LM_{\varphi=0|\rho \neq 0} &= \left(\frac{1}{\sigma^2} \mu'W_A'W_A Z \right) \left(\frac{1}{\sigma^2} Z'W_A'W_A Z \right)^{-1} \left(\frac{1}{\sigma^2} \mu'W_A'W_A Z \right)' \\
&= \frac{\tilde{\mu}'\tilde{Z} (\tilde{Z}'\tilde{Z})^{-1} \tilde{Z}'\tilde{\mu}}{\tilde{\sigma}^2},
\end{aligned} \tag{A13}$$

where $\tilde{\mu} = W_A (y - X\xi)$ corresponds to the residuals of the spatial error model, $\tilde{Z} = W_A Z$ are the spatially filtered variables Z , and $\tilde{\sigma}^2 = \tilde{\mu}'\tilde{\mu}/n$. Both tests are asymptotically distributed following a χ^2 distribution with k degrees of freedom, where k equals the number of restrictions implied by $\varphi = (0, 0, \dots, 0)'$.

A1.2 Spatially autoregressive error tests

The tests for spatially autoregressive errors can be shown to be the standard result originally derived in Burridge (1980). Two null hypotheses can be distinguished, depending on whether the

restricted model is the non-spatial model or the spatial STAR model. For the first case the parameter vector of the restricted model is $\theta_0 = (\xi', 0, 0, \sigma^2)'$ and the statistic for the null hypothesis $H_0 : \rho = 0 | \varphi = 0$ is the familiar:

$$LM_{\rho=0|\varphi=0} = \frac{1}{\text{tr}((W' + W)W)} \left(\frac{\mu'W\mu}{\sigma^2} \right)^2, \quad (\text{A14})$$

where $\mu = y - X\xi$ corresponds to the OLS residuals of the restricted model, and $\sigma^2 = \mu'\mu/n$. In the second case the parameter vector of the restricted model is $\theta_0 = (\xi', \varphi', 0, \sigma^2)'$ and the test of the null hypothesis $H_0 : \rho = 0 | \varphi \neq 0$ is identical to (A14), except that in this case $\mu = y - Z\beta$. Both tests are asymptotically χ^2 distributed with one degree of freedom.

A1.3 Joint test for nonlinearity and spatially autoregressive errors

The familiar additivity property of many LM tests holds in the case of a joint test for nonlinearity and spatially autoregressive errors, because the asymptotic variance matrix in (A11) is block-diagonal under the null hypothesis $H_0 : \rho = 0, \varphi = 0$. The parameter vector of the restricted model is $\theta_0 = (\xi', 0, 0, \sigma^2)'$, and the LM statistic for the joint null hypothesis is therefore simply:

$$LM_{\rho=\varphi=0} = \begin{bmatrix} \frac{1}{\sigma^2} \mu'Z \\ \frac{1}{\sigma^2} \mu'W\mu \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} Z'Z & 0 \\ 0 & \text{tr}((W' + W)W) \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{\sigma^2} \mu'Z \\ \frac{1}{\sigma^2} \mu'W\mu \end{bmatrix} \quad (\text{A15})$$

$$= \frac{\mu'Z(Z'Z)^{-1}Z'\mu}{\sigma^2} + \frac{1}{\text{tr}((W'+W)W)} \left(\frac{\mu'W\mu}{\sigma^2} \right)^2$$

$$= LM_{\varphi=0|\rho=0} + LM_{\rho=0|\varphi=0},$$

where $\mu = y - X\xi$ corresponds to the OLS residuals, and $\sigma^2 = \mu'\mu/n$. The test asymptotically follows a χ^2 distribution with $k+1$ degrees of freedom, where k equals the number of restrictions implied by $\varphi = (0, 0, \dots, 0)'$.

Appendix 2: R script for the spatial (error) STAR model

Code for the empirical example has been written in R (version 2.8.0) using a series of external libraries. The script contains estimators for various restricted and unrestricted models, the associated LM tests, and graphical output in terms of a map and a plot of the transition function.

```
#####
# A Spatial Econometric STAR Model
# with an Application to U.S. County Economic Growth, 1969–2003
#
# (c) rjgmflorax, vopede
# Dept. of Agricultural Economics, Purdue University
# Space, Health and Population Economics (SHaPE) Program
# West Lafayette, IN, USA
# rflorax@purdue.edu, vpede@purdue.edu
#
# Date last changed: 02.14.2009
#####

# load libraries

library(spdep)
library(grid)
library(sem)
library(fields)
library(MASS)
library(fUtilities)
library(RColorBrewer)

# load data using shapefile

uscnt <- readShapePoly("C:/Program Files/R/work/STAR/rpci.shp", IDvar="ID")
```

```

# import weights matrix and calculate eigenvalues

weightgal <- read.gal("C:/Program Files/R/work/STAR/queen12.GAL", region.id=uscnt$ID)

w <- nb2mat(weightgal)
ws <- nb2listw(weightgal)

v <- eigenw(ws, quiet = TRUE) # eigenvalues
vinvmin <- 1/min(v) # minimum eigenvalue
vinvmax <- 1/max(v) # maximum eigenvalue
interval <- c(vinvmin,vinvmax) # interval to set parameter space rho

# define variables

x <- uscnt$LN_RPC_69
wx <- lag.listw(ws,x)
wxx <- wx*x
y <- uscnt$LN_PR_0369
wy <- lag.listw(ws,y)
n <- length(x)
I <- diag(n)

## spatial STAR model,  $y = Xb + Xd.G(Wx;g,c) + u$ 

funct.nls <- nls(y ~ beta0+beta1*x+(delta0+delta1*x)/(1+exp(-gamma*(wx-c)/sd(wx))),
  start = list(beta0 = 5.14218, beta1 = -0.483006, delta0 = -11.3601, delta1 = 1.1553,
  gamma = 3.1540, c = 9.97903), trace = TRUE, algorithm = "port")
summary(funct.nls)

cstar <- coef(funct.nls)
e.nonlin <- resid(funct.nls)

## null and alternative specifications

# ols, non-spatial model,  $y = Xb + u$ 
funct.lin <- lm(y ~ x)
summary(funct.lin)

clin <- coef(funct.lin)
bconv.Xb <- clin[2]
e.Xb <- resid(funct.lin)

# linear spatial error model,  $y = Xb + (I-rW)^{-1} u$ 
funct.err <- errorsarlm(y~x,data=uscnt,ws)
summary(funct.err)

cerr <- coef(funct.err)
bconv.err <- cerr[3]
e.err <- resid(funct.err)

# spatial error STAR model,  $y = Xb + Xd.G(Wx;g,c) + (I-rW)^{-1} u$ 

e <- e.nonlin
crit <- 0.000001
max.iter <- 100
obj.old <- 0

for (i in 1:max.iter) {
  flerr <- function(rho) {
    (-n/2)*log((1/n)*crossprod((I-rho*w)%*%e))+sum(log(1-rho*v))
  }
}

```

```

fmax <- optimize(flerr,interval=interval,maximum=TRUE)
rho <- fmax$maximum
irwy <- y-rho*lag.listw(ws,y)
irwx <- x-rho*lag.listw(ws,x)

funct.errstar <- nls(irwy ~ beta0*(1-rho)+beta1*irwx+(delta0*(1-rho)+
  delta1*irwx)/(1+exp(-gamma*(wx-c)/sd(wx))),
  start = cstar, trace = FALSE, algorithm = "port")

cstar.err <- coef(funct.errstar)

G <- (1+exp(-cstar.err[5]*(wx-cstar.err[6])/sd(wx)))^-1
e <- y-(cstar.err[1]+cstar.err[2]*x+(cstar.err[3]+cstar.err[4]*x)*G)

if (abs(fmax$objective - obj.old) <= crit) break
obj.old <- fmax$objective
}

# estimates and inference for beta, delta, gamma and c

summary(funct.errstar)

# estimate and inference for rho

sig2 <- crossprod(e-rho*lag.listw(ws,e))/n
varsig2 <- n/(2*sig2^2)
wirwi <- w%*%invIrW(ws,rho)
covsig2rho <- (1/sig2)*sum(diag(wirwi))
varrho <- (sum(diag(wirwi)))^2 + sum(diag(crossprod(wirwi)))

row1 <- c(varsig2,covsig2rho)
row2 <- c(covsig2rho,varrho)
mat <- rbind(row1,row2)
covmat <- solve(mat)

serho <- sqrt(covmat[2,2])
trho <- rho/serho
df <- n-length(cstar.err+1)
prho <- 2*pt(abs(trho),df=df,lower.tail=FALSE)

output.rho <- cbind(rho,serho,trho,prho)
print(output.rho)

## LM tests, based on linearization using first-order Taylor approximation

# LM rho = 0 given phi = 0

tr <- sum(diag(crossprod(w)))+sum(diag(w%*%w))
we <- lag.listw(ws,e.Xb)
ewe <- crossprod(e.Xb,we)
s2 <- crossprod(e.Xb)/n
LMerr.Xb <- (ewe/s2)^2/tr
pLMerr.Xb <- 1-pchisq(LMerr.Xb,df=1,ncp=0,log=FALSE)
LMerr.Xb
pLMerr.Xb

# LM phi = 0 given rho = 0

Z <- cbind(1,x,wx,wxx)
P <- Z%*%solve(crossprod(Z))%*%t(Z)
LMlin.Xb <- crossprod(e.Xb,P)%*%e.Xb/(crossprod(e.Xb)/n)
pLMlin.Xb <- 1-pchisq(LMlin.Xb,df=2,ncp=0,log=FALSE)

```

```

LMlin.Xb
pLMlin.Xb

# LM rho = phi = 0

LMerrlin.Xb <- LMerr.Xb+LMlin.Xb
pLMerrlin.Xb <- 1-pchisq(LMerrlin.Xb,df=3,ncp=0,log=FALSE)
LMerrlin.Xb
pLMerrlin.Xb

# LM phi = 0 given rho <> 0

Ztild <- Z-cerr[1]*lag.listw(ws,Z)
Ptild <- Ztild**solve(crossprod(Ztild))**t(Ztild)
LMlin.err <- crossprod(e.err,Ptild)**e.err/(crossprod(e.err)/n)
pLMlin.err <- 1-pchisq(LMlin.err,df=2,ncp=0,log=FALSE)
LMlin.err
pLMlin.err

# LM rho = 0 given phi <> 0

lm.nonlin <- lm(y~x+wx+wxx)
e.nonlin <- resid(lm.nonlin)
we <- lag.listw(ws,e.nonlin)
ewe <- crossprod(e.nonlin,we)
s2 <- crossprod(e.nonlin)/n
LMerr.nonlin <- (ewe/s2)^2/tr
pLMerr.nonlin <- 1-pchisq(LMerr.nonlin,df=1,ncp=0,log=FALSE)
LMerr.nonlin
pLMerr.nonlin

## Visualization

# transition function and convergence rate, -100.(ln(dydx+1)/T,
# for the spatial error STAR model

G <- (1+exp(-cstar.err[5]*(wx-cstar.err[6])/sd(wx)))^-1

aux1 <- (1+exp(-cstar.err[5]*(wx-cstar.err[6])/sd(wx)))^-2
aux2 <- cstar.err[4]+cstar.err[5]*w**((cstar.err[3]+cstar.err[4]*x)/sd(wx))
aux3 <- exp(-cstar.err[5]*(wx-cstar.err[6])/sd(wx))
bconv <- cstar.err[2] + (cstar.err[4]+aux2)*aux3*aux1
crate <- (-100/34)*log(1+bconv)

# map function

MAP <- function(plotvar,title,steps) {
  brks <- c(-4,0,0.25,0.5,0.75,1,1.25,1.50,1.60)
  colors <- brewer.pal(steps,"Blues")
  #colors <- colors[steps:1] ## reorder colors

plot(uscnt,border="lightgray",col=colors[findInterval(plotvar,brks,all.inside=TRUE)])
  box()
  legend("bottomright",legend=leglabs(brks),fill=colors,bty="n",cex=0.7,
  y.intersp=1,x.intersp=1)
  title(x=list(title))
}

# map convergence rate and plot transition function
graph <- par(mfrow = c(2, 2))
MAP(crate,"Convergence rate (% per yr)",8)
plot(wx,G,xlab="Wx")

```

References

- Abreu, Maria, de Groot, Henri L.F., Florax, Raymond J.G.M., 2005. Space and Growth: A Survey of Empirical Evidence and Methods, *Région et Développement* **21**, 13–44.
- Anselin, Luc, 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.
- Anselin, Luc, 2006. Spatial Econometrics. In: Mills, Terence C., Patterson, Kerry (eds.), *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Palgrave Macmillan, Basingstoke, 901–969.
- Anselin, Luc, Bera, Anil K., Florax, Raymond, Yoon, Mann J., 1996. Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics* **26**, 77–104.
- Basile, Roberto, 2008. Regional Economic Growth in Europe: A Semiparametric Spatial Dependence Approach. *Papers in Regional Science* **87**, 527–544.
- Basile, Roberto, Gress, Bernard, 2005. Semi-Parametric Spatial Auto-Covariance Models of Regional Growth in Europe. *Région et Développement* **21**, 93–118.
- Burridge, Peter, 1980. On the Cliff-Ord test for Spatial Correlation. *Journal of the Royal Statistical Society B* **42**, 107–108.
- Cho, Seong-Hoon, Lambert, Dayton M., Chen, Zhuo, 2009. Geographically Weighted Regression Bandwidth Selection and Spatial Autocorrelation: An Empirical Example Using Chinese Agriculture Data. *Applied Economics Letters* **16**, forthcoming.
- Dall’erba, Sandy, Percoco, Marco, Piras, Gianfranco, 2008. The European Regional Growth Process Revisited. *Spatial Economic Analysis* **3**, 7–25.

- Fotheringham, A. Stewart, Brundson, Chris, Charlton, Martin, 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons, Hoboken, NJ.
- LeSage, James P., 2004. A Family of Geographically Weighted Regression Models. In: Anselin, Luc, Florax, Raymond J.G.M., Rey, Sergio J. (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer Verlag, Berlin, 241–264.
- Luukkonen, Ritva, Saikkonen, Pentti, Teräsvirta, Timo, 1988. Testing Linearity against Smooth Transition Autoregressive Models. *Biometrika* **75**, 491–499.
- Teräsvirta, Timo, 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association* **89**, 208–218.