

Spatial Filtering with EViews and MATLAB

Robert Ferstl

Vienna University of Economics and Business Administration

Abstract: This article summarizes the ideas behind a few programs we developed for spatial data analysis in EViews and MATLAB. They allow the user to check for spatial autocorrelation using Moran's \mathcal{I} and provide a spatial filtering procedure based on the \mathcal{G}_i statistic by Getis and Ord (1992). We have also implemented graphical tools like Moran Scatterplots for the detection of outliers or local spatial clusters.

Zusammenfassung: Dieser Artikel beschreibt einige Programme, die wir zur Analyse räumlicher Daten in EViews und MATLAB entwickelt habe. Sie erlauben dem Anwender eine Überprüfung auf räumliche Autokorrelation mittels Moran's \mathcal{I} und ermöglichen ein räumliches Filterungsverfahren basierend auf der \mathcal{G}_i Statistik von Getis and Ord (1992). Außerdem haben wir grafische Werkzeuge (z.B. Morans Streudiagramme) zur Analyse auf Ausreisser und zum Auffinden lokaler räumlicher Zusammenschlüsse implementiert.

Keywords: Spatial Autocorrelation, Moran Scatterplots, Spatial Filtering.

1 Introduction

In recent years spatial econometric methods gained in popularity. Nevertheless, the widely used econometric software package EViews¹ does not contain functions for spatial data analysis. Therefore, we have developed a few scripts which can be started by a short entry in the command line. For users of the spatial econometrics toolbox by James P. LeSage² the MATLAB³ version of the functions is an appropriate extension.

The programs calculate the global and local Moran's \mathcal{I} statistic for spatial autocorrelation and its moments using the normal approximation and the more accurate saddlepoint approximation by Tiefelsdorf (2002). It is possible to check the data with graphical tools like Moran Scatterplots for outliers or local spatial clusters. Finally, we have implemented a filtering procedure based on the \mathcal{G}_i statistic, which is another measure for local spatial dependence. So the user has an uncomplicated possibility to handle the presence of spatial autocorrelation in the data. For further details about the programs and the theoretical background see Ferstl (2004).

¹<http://www.eviews.com/>

²<http://www.spatial-econometrics.com/>

³<http://www.mathworks.com/>

2 Spatial Structures

2.1 Global Spatial Structures

The analysis of spatial data begins with a feasible measure to represent the inherent spatial dependence. The elements of a symmetric $n \times n$ spatial distance matrix D are defined by

$$d_{ij} = \begin{cases} \sqrt{(\theta_i - \theta_j)^2 + (\omega_i - \omega_j)^2}, & i \neq j, \\ 0, & i = j, \end{cases}$$

where (θ, ω) are the cartesian coordinates. In contrast to such a matrix, which measures dissimilarities, it is possible to use a spatial weight matrix to describe the similarities between the i -th and j -th spatial object.

There are several methods to transform a distance matrix to a global spatial weight matrix G . We implemented the widely used negative exponential function to model the distance decay function (see e.g. Badinger et al., 2004). Therefore, the geographical distances are transformed to spatial weights by

$$g_{ij} = \begin{cases} \exp(-\delta d_{ij}) > 0 & \text{if objects } i \text{ and } j \text{ are spatially connected,} \\ 0 & \text{otherwise,} \end{cases}$$

with $0 < \delta < \infty$.

2.2 Coding Schemes for Spatial Weight Matrices

The programs provide three common coding schemes for standardization of the spatial weight matrix G . Tiefelsdorf (2000) defines the linkage degree of the i -th spatial object by $d_i = \sum_{j=1}^n g_{ij}$. Therefore, the overall connectivity in G is $D \equiv \sum_{i=1}^n \sum_{j=1}^n g_{ij}$.

The **globally standardized C-coding scheme** is defined as

$$C \equiv \frac{n}{D} G.$$

In the **row-sum standardized W-coding scheme** the spatial weight matrix is transformed by

$$W \equiv [\text{diag}(d)]^{-1} G,$$

where d is a vector with the row-sums of G . In the **variance stabilizing S-coding scheme** Tiefelsdorf (2000) defines the vector q by

$$q \equiv \left(\sqrt{\sum_{j=1}^n g_{1j}^2}, \sqrt{\sum_{j=1}^n g_{2j}^2}, \dots, \sqrt{\sum_{j=1}^n g_{nj}^2} \right)^\top.$$

As in the row-sum standardization the spatial weight matrix is transformed by $S^* \equiv [\text{diag}(q)]^{-1} G$. Finally, it is scaled by

$$S \equiv \frac{n}{Q} S^*,$$

where $Q \equiv \sum_{i=1}^n \sum_{j=1}^n s_{ij}^*$.

Following the notation of Tiefelsdorf (2000), we use V as a placeholder for a spatial weight matrix which has been transformed by one of the three mentioned coding schemes.

2.3 Local Spatial Structures

The local spatial weight matrix for the i -th spatial object consists of a copy of the i -th row and column of the global spatial weight matrix and zeros elsewhere. This definition leads to the following star-shaped form for the standardized local spatial weight matrices

$$\mathbf{V}_i \equiv s_i \cdot \begin{pmatrix} 0 & \cdots & 0 & g_{1i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & g_{i-1,i} & 0 & \cdots & 0 \\ g_{i1} & \cdots & g_{i,i-1} & 0 & g_{i,i+1} & \cdots & g_{in} \\ 0 & \cdots & 0 & g_{i+1,i} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & g_{ni} & 0 & \cdots & 0 \end{pmatrix}.$$

The scaling coefficient s_i depends on the applied coding scheme (see Tiefelsdorf, 2000, p. 35)

$$s_i = \begin{cases} n^2/2D, & \text{for the } C\text{-coding scheme,} \\ n/2d_i, & \text{for the } W\text{-coding scheme,} \\ n^2/2Qq_i, & \text{for the } S\text{-coding scheme.} \end{cases}$$

The sum over all local spatial weight matrices equals the global spatial weight matrix, i.e.

$$\sum_{i=1}^n \mathbf{V}_i = \frac{1}{2} (\mathbf{V} + \mathbf{V}^\top). \quad (1)$$

3 Spatial Autocorrelation

3.1 Global Moran's \mathcal{I}

Having the spatial structure summarized in a spatial weight matrix, it is possible to define a measure for the spatial autocorrelation. The most prominent one is a statistic developed by Moran (1948), which is called Moran's \mathcal{I} in the literature. According to Tiefelsdorf (2002) it is defined as follows.

Let \mathbf{y} be a system of n spatially distributed observations that are related to a set of k exogenous variables via a linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The $n \times k$ design matrix includes the usual constant vector. The disturbances are distributed as $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Omega})$. The $n \times n$ matrix $\boldsymbol{\Omega}$ reflects the covariance structure and σ^2 the variance of the disturbances. For independent disturbances the covariance structure is $\sigma^2 \mathbf{I}$. The OLS regression residuals $\hat{\boldsymbol{\epsilon}} = \mathbf{M}\mathbf{y}$ are distributed as $\hat{\boldsymbol{\epsilon}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{M}\boldsymbol{\Omega}\mathbf{M})$ using the projection matrix $\mathbf{M} \equiv \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. For an underlying autoregressive spatial process $\boldsymbol{\Omega} \equiv (\mathbf{I} - \rho \mathbf{V})^{-1} (\mathbf{I} - \rho \mathbf{V}^\top)^{-1}$, where ρ is the spatial autocorrelation coefficient.

Global Moran's \mathcal{I} is defined as a scale invariant ratio of quadratic forms in the normally distributed vector of regression residuals $\hat{\boldsymbol{\epsilon}}$

$$\mathcal{I} = \frac{\hat{\boldsymbol{\epsilon}}^\top \frac{1}{2} (\mathbf{V} + \mathbf{V}^\top) \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}. \quad (2)$$

The multiplication with $(\mathbf{V} + \mathbf{V}^\top)/2$ in the numerator ensures symmetry of the standardized spatial weight matrix.

3.2 Local Moran's \mathcal{I}

Frequently the spatial dependencies are not the same for the whole dataset. Therefore, it is advisable to calculate a local test statistic. Local Moran's \mathcal{I}_i is defined by using the local spatial weight matrix introduced in Section (2.3), i.e.

$$\mathcal{I}_i = \frac{\hat{\boldsymbol{\epsilon}}^\top \mathbf{V}_i \hat{\boldsymbol{\epsilon}}}{\hat{\boldsymbol{\epsilon}}^\top \hat{\boldsymbol{\epsilon}}}.$$

Due to the additivity property in (1) the global Moran's \mathcal{I} can be expressed as sum over all local Moran's \mathcal{I}_i .

3.3 Moments of Moran's \mathcal{I}

The moments of the global and local Moran's \mathcal{I} statistic are calculated under the assumption of spatial independence. The moments of \mathcal{I} can be expressed in terms of the resequenced spectrum of eigenvalues $\{\lambda_1, \dots, \lambda_{n-k}, 0, \dots, 0\}$ of $\mathbf{K} \equiv \mathbf{M}(\mathbf{V} + \mathbf{V}^\top)\mathbf{M}/2$ (see Tiefelsdorf, 2000, chapter 9).

Expected value $E[\mathcal{I}] = \mu_1$, variance $\text{var}[\mathcal{I}] = \mu_2$, skewness $\mu_3/(\mu_2)^{3/2}$ and kurtosis $\mu_4/(\mu_2)^2$ of Moran's \mathcal{I} are given by

$$\begin{aligned} \mu_1 &\equiv \frac{\sum_{i=1}^{n-k} \lambda_i}{n-k} = \bar{\lambda} \\ \mu_2 &\equiv \frac{2 \sum_{i=1}^{n-k} (\lambda_i - \bar{\lambda})^2}{(n-k)(n-k+2)} \\ \mu_3 &\equiv \frac{8 \sum_{i=1}^{n-k} (\lambda_i - \bar{\lambda})^3}{(n-k)(n-k+2)(n-k+4)} \\ \mu_4 &\equiv \frac{48 \sum_{i=1}^{n-k} (\lambda_i - \bar{\lambda})^4 + 12 \left(\sum_{i=1}^{n-k} (\lambda_i - \bar{\lambda})^2 \right)^2}{(n-k)(n-k+2)(n-k+4)(n-k+6)}. \end{aligned}$$

3.4 Distribution of Moran's \mathcal{I}

The z -transformed observed Moran's \mathcal{I}_o is asymptotically normal distributed, i.e.

$$z(\mathcal{I}_o) = \frac{\mathcal{I}_o - E[\mathcal{I}]}{\sqrt{\text{var}[\mathcal{I}]}} \approx \mathcal{N}(0, 1).$$

Tiefelsdorf (2000) was the first who published the exact reference distribution of Moran's \mathcal{I} . Its evaluation is complicated and computationally expensive. He proposed a saddlepoint approximation which is fast and accurate. Our software implementation yields the same results as the SPSS macro on the authors homepage⁴. For the mathematical background of the saddlepoint approximation see e.g. Tiefelsdorf (2002) or Ferstl (2004).

⁴<http://geog-www.sbs.ohio-state.edu/faculty/tiefelsdorf/GeoStat.htm>

4 Moran Scatterplots

Anselin (1996) describes the Moran scatterplot as a tool for exploratory spatial data analysis (ESDA). Using the definition of global Moran's \mathcal{I} in (2), it is possible to interpret \mathcal{I} as the coefficient in a regression of the spatially lagged variable $(\mathbf{V} + \mathbf{V}^\top)\hat{\epsilon}/2$ on $\hat{\epsilon}$. In other words, the slope of the regression line equals global Moran's \mathcal{I} .

Moran scatterplots can be used to find local patterns of spatial association. They also provide an easy way to detect outliers. Additionally, our programs plot a smoothed curve through the datapoints (instead of the regression line), and line diagrams of the studentized residuals and the Cook's distances. It is also possible to generate a so called Moran scatterplot matrix, which helps to visualize spatial interactions between two or more variables (see Anselin et al., 2002)

5 Spatial filtering

After having successfully detected spatial autocorrelation in the data, the question is how to handle it. The simplest alternative is to spatially filter the data, and estimate the model with the usual OLS procedure.

5.1 The \mathcal{G}_i Statistic

A spatial filter is based on a local statistic of spatial dependence. Contrary to the version developed by Getis and Ord (1992), we use a non-binary weight matrix like in Section (2.1). This extension was introduced by Getis and Ord (1995) and was recently used in Badinger et al. (2004)

$$\mathcal{G}_i(\delta) = \frac{\sum_{j=1}^n v_{ij}(\delta)x_j}{\sum_{j=1}^n x_j}, \quad j \neq i.$$

The v_{ij} terms are the elements of a standardized weight matrix \mathbf{V} . The x_j represent observations of a random variable X_j . The expected value of the \mathcal{G}_i statistic is given by

$$E[\mathcal{G}_i] = \frac{V_i}{(n-1)}, \quad (3)$$

with $V_i = \sum_{j \neq i} v_{ij}(\delta)$. The variance is

$$\text{var}[\mathcal{G}_i] = \frac{V_i(n-1-V_i)}{(n-1)^2(n-2)} \frac{Y_{i2}}{Y_{i1}^2},$$

where $\sum_{j=1}^n x_j/(n-1) = Y_{i1}$ and $\sum_{j=1}^n x_j^2/(n-1) - Y_{i1}^2 = Y_{i2}$.

The distribution of the z -standardized statistic $z(\mathcal{G}_i)$ is like Moran's \mathcal{I} asymptotically $\mathcal{N}(0, 1)$. The \mathcal{G}_i^* statistic differs from the \mathcal{G}_i statistic by including the value of x_i if $i = j$.

5.2 The Filtering Process

The expected value in (3) represents the realized value in the i -th region when no autocorrelation occurs. Dividing it by the \mathcal{G}_i statistic results in a ratio that represents the spatially uncorrelated part of the data, i.e. the filtered variable X^*

$$x_i^* = \frac{x_i (V_i / (n - 1))}{\mathcal{G}_i(\delta)}.$$

Purely the spatial effects are stored in a new variable $L = X - X^*$. The goal is to minimize the remaining spatial autocorrelation in the filtered variable by varying the distance decay parameter δ . That leads to the optimization of the objective function

$$\delta_{opt} = \min_{\delta} |z_I(\mathbf{X}^*)|, \quad 0 < \delta < \infty. \quad (4)$$

6 Programs

Table 1 shows a list of available program files for EViews and MATLAB. Detailed descriptions of the syntax are included as comments. Demos of the functions can be found in the example files mentioned in Table 2.

Table 1: Programs

EViews subroutines	MATLAB functions	Description
moransad.prg	moransad.m	Global and local Moran's I (including saddlepoint approximation)
moranplot.prg	moranplot.m	Moran scatterplot (with outlier detection)
opt_delta.prg	moranplotmatrix.m	Moran scatterplot matrix
distance2weight.prg	opt_delta.m	optimum delta for weights matrix
getis.prg	distance2weight.m	convert distance matrix to weight matrix
optim.prg	getis.m	spatial filtering with Getis statistic
	objfct.m	one-dimensional optimization

Table 2: Examples

EViews subroutines	MATLAB functions	Description
ex_a.prg	ex_a.m	Example A: calculate Moran's I
ex_b.prg	ex_b.m	Example B: Moran scatterplot for one variable
ex_c.prg	ex_c.m	Example C: Moran scatterplot for one variable (with outliers)
ex_d.prg	ex_d.m	Example D: Moran scatterplot matrix for two variables
ex_e.prg	ex_e.m	Example E: calculate optimum δ for spatial filtering
ex_f.prg	ex_f.m	Example F: spatial filtering with a given δ

7 Practical Example

The following examples are taken from Ferstl (2004). The analyzed series represents the gross value added per capita in million ECU in 1975 based on price and exchange rate levels from 1990. The $i = 1, \dots, 194$ spatial objects are NUTS 2 regions of the EU-15 countries. Table 3 shows the result of a test for global spatial autocorrelation. To perform the calculations in EViews these commands must be entered in the command line.

```
'Open dataset :
open Data\eu.wfl

'Calculate global and local Moran's I:
run moransad ser01 const u w gl j
```

Table 3: Global Moran's \mathcal{I} with normal and saddlepoint approximation

Results using normal approximation	
0.6185	Global Moran's \mathcal{I}
-0.0052	Expected Value
0.0006	Variance
0.4355	Skewness
3.3800	Kurtosis
1.0000	Probability ($\mathcal{I} \leq \mathcal{I}_o$)
25.7052	$z(\mathcal{I}_o)$
Results using saddlepoint approximation	
1.2877	Saddlepoint
13.0177	r-parameter for distribution
68.1026	u-parameter for distribution
1.0000	Probability ($\mathcal{I} \leq \mathcal{I}_o$)

The analyzed variable shows a highly positive and significant spatial autocorrelation. Taking a look at a Moran scatterplot yields the same conclusion. Figure 1 also includes measures to identify influential observations. The following code constructs a Moran scatterplot in EViews.

```
'Moran scatterplot :
run moranplot ser01 w
```

The Moran scatterplot in Figure 1 shows a cluster of small values in the third quadrant. Alternatively, this local spatial dependence can be visualized by a choroplethe map of the \mathcal{G}_i^* statistics, see Figure 2. There occur negative significant values in the region around Portugal.

The spatial filtering program finds a δ^{opt} at 138.1611. Figure 3 shows a plot of the objective function. Using the optimum δ parameter the spatial autocorrelation is successfully removed from the variable. The following commands are necessary to do this in EViews.

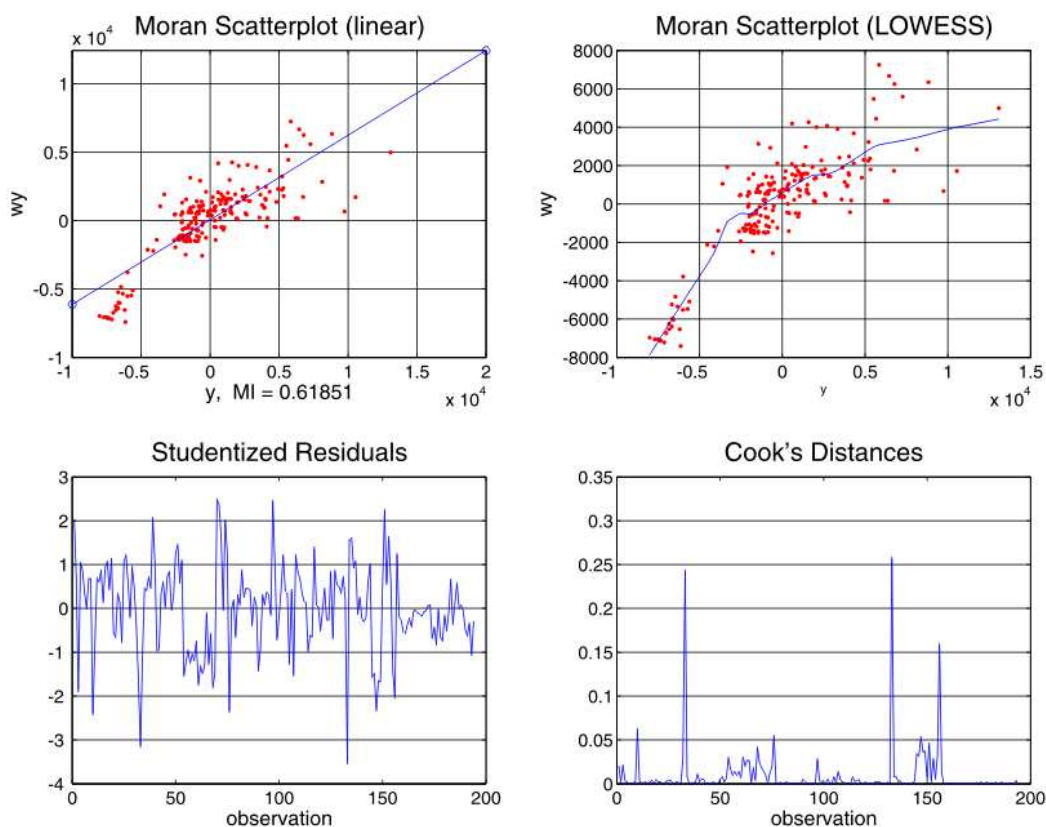


Figure 1: Moran Scatterplot with statistics for outlier detection

```
' Spatial filtering :
run optdelta ser01 dst 70 80 100 0.00001 100
```

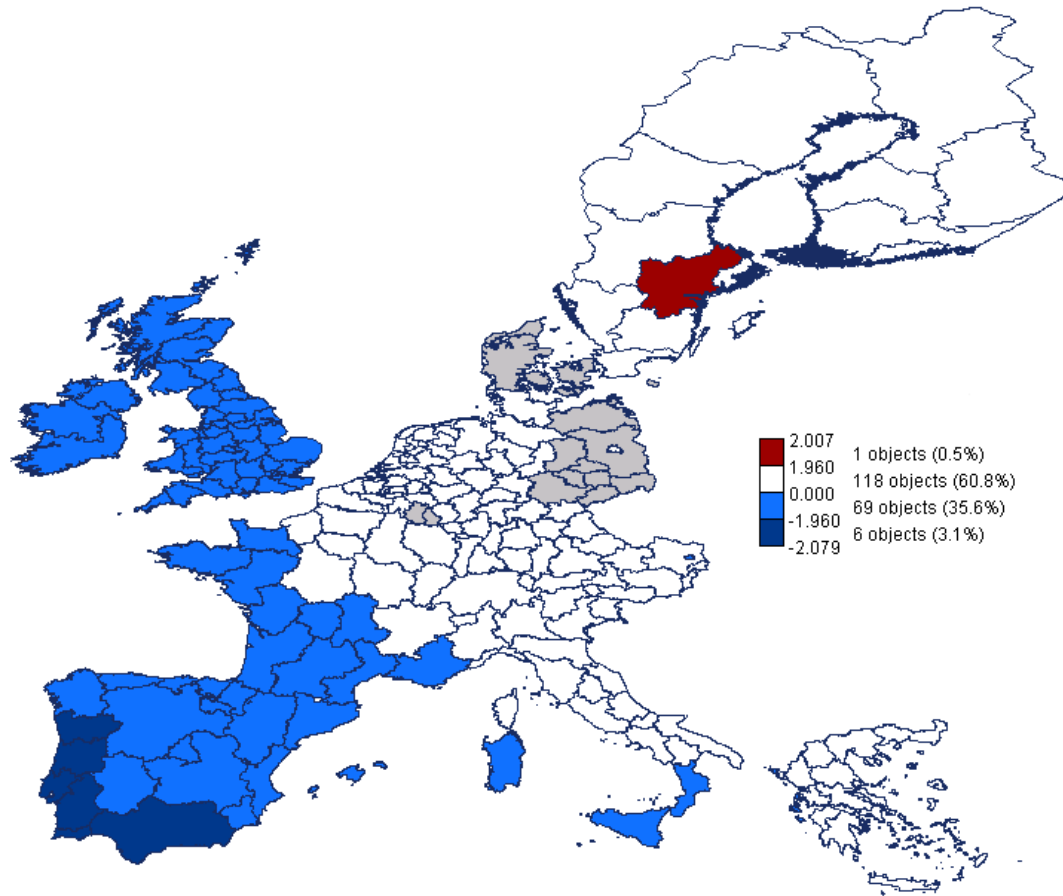
8 Conclusions

In this paper we shortly demonstrate a collection of programs we have developed for spatial data analysis in Ferstl (2004). The idea was to offer users of the widespread econometric software package EViews simple functions to analyze data for spatial dependencies. We have implemented functions to calculate Moran's \mathcal{I} and its central moments using the normal and the saddlepoint approximation. Influential observations and spatial clusters can be visualized by Moran scatterplots. One way to deal with the presence of spatial autocorrelation is to filter the data. For this purpose functions are available, too.

All programs are available on request by email from the authors. They have been tested under Windows 2000/XP in EViews 4/5 and MATLAB 6/7.

Acknowledgments

We would like to thank the Department for Economic Geography and Geoinformatics at the Vienna University of Economics and Business Administration for providing us the

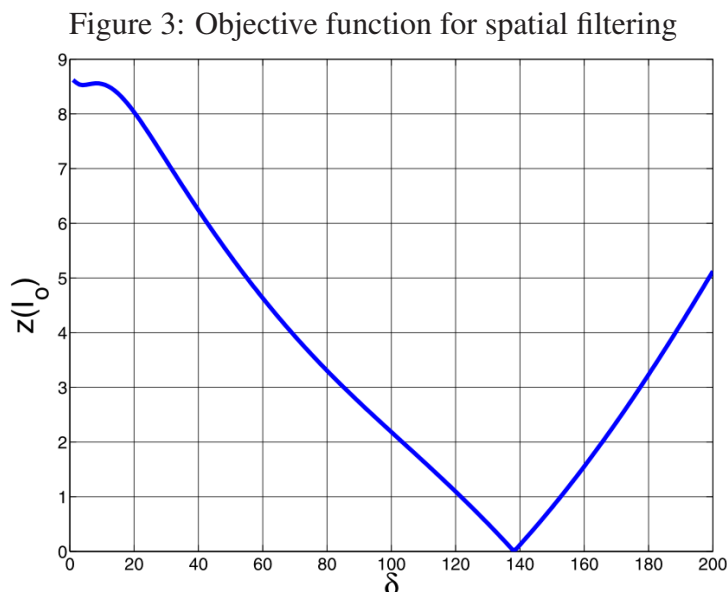
Figure 2: Classified choroplethe map of the \mathcal{G}_i^* statistics

Source: Macon AG 1999 (Geodata Europe NUTS, Department for Economic Geography and Geoinformatics, University of Economics and Business Administration, Vienna)

shape file used in the cartographic visualization.

References

- Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. M. Fischer, H. J. Scholten, and D. Unwin (Eds.), *Spatial Analytical Perspectives on GIS* (p. 111-125). London: Taylor & Francis Ltd.
- Anselin, L., Syabri, I., and Smirnov, O. (2002). *Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows*. (URL: http://agec221.agecon.uiuc.edu/users/anselin/papers/multi_lisa.pdf)
- Badinger, H., Müller, W., and Tondl, G. (2004). Regional convergence in the European Union, 1985-1999: A spatial dynamic panel analysis. *Regional Studies*, 38, 241-253.
- Ferstl, R. (2004). *Werkzeuge zur Analyse räumlicher Daten – Eine Softwareimplementierung in EViews und MATLAB*. Unpublished master's thesis, Wirtschaftsuniversität Wien.



Getis, A., and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Journal of Geographical Analysis*, 24, 189-206.

Getis, A., and Ord, J. K. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Journal of Geographical Analysis*, 27, 286-305.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B*, 10, 243-51.

Tiefelsdorf, M. (2000). *Modelling Spatial Processes – The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I*. Springer.

Tiefelsdorf, M. (2002). The saddlepoint approximation of Moran's i 's and local Moran's i_i 's reference distribution and their numerical evaluation. *Journal of Geographical Analysis*, 34, 187-206.

Author's address:

Robert Ferstl

Institute for Quantitative Management and Operations Research

Vienna University of Economics and Business Administration

Nordbergstraße 15

A-1090 Vienna, Austria

Tel. +43 1 31336 / 4560

Fax +43 1 31336 / 708

E-mail: robert.ferstl@wu-wien.ac.at

<http://www.wu-wien.ac.at/or>