The Regional Growth Pattern in Sweden - A Search for Hot Spots

Johan Lundberg*

Centre for Regional Science (CERUM), Umeå University, Sweden

July 9, 2002

Abstract

This paper gives an exploratory description of the regional growth pattern in Sweden during the period 1981-1999. The main issue is to test the hypothesis that municipalities with higher average income growth are more clustered that could be caused by pure chance. The paper is purely descriptive where we make use of statistical tests for spatial correlation as well as maps to identify what we reefer to as 'regional hot spots'. Our results are however to some extent sensitive for the specification of the weighting matrix.

*Corresponding address: Johan Lundberg, Centre for Regional Science (CERUM), Umeå University, SE-901 87 Umeå, Sweden. E-mail: johan.lundberg@econ.umu.se.

1 Introduction

This paper concern spatial clusters of average income growth at the local level of government in Sweden during the period 1981-1999. The main purpose is to identify what we refer to as hot spots. That is, to test the hypothesis that municipalities with similar average income growth rates are more clustered than would be expected from merely a coincidence. The analysis of average income growth is motivated from a local public finance perspective, as the local income taxes constitute the major source of funds for the Swedish municipalities. Hence, changes in average income levels is one component[1] that affect the local tax base and, consequently, the local public sector's ability to provide services imposed on them by the central government. The analysis is based on two commonly used test statistics for spatial clustering, the Moran's I and the new G_i^*-statistic proposed by Ord and Getis (1995). These two statistics complement each other and are, according to Ord and Getis, preferable used in combination. The Moran's I is a global test for spatial correlation and tells us if high or low values are more clustered than would be expected by pure chance (positive spatial correlation). If the Moran's I reveals a negative spatial correlation, the data is organized as a checkerboard pattern. The new G_i^*-statistic, which is a local test for spatial correlation, complement the Moran's I in at least two ways: 1) it reveals if there is a cluster of high or low values, not only that there is a cluster of high or low values, and 2) it reveals where these clusters are located.

Before we proceed, let us discuss some stylized facts regarding the regional growth pattern in Sweden. During the last decades, there has been a tendency that individuals move from the sparsely populated areas in the northern and western parts of the country to more densely populated areas as the major cities or university towns (see for instance Aronsson, Lundberg and Wikström (2001) and Aronsson and Lundberg (2002)). At the same time, the local public sector, which is mainly financed through a local personal income tax, has expanded quite dramatically both in terms of expenditures per capita and in local income tax rates.[2] This expansion

[1] The other component relates to population growth and the age distribution.

[2] The local income tax rate has increased from an average of 16.68-percent in 1981 to 20.55-percent in 1999. During the same period, the average income has increased from Y to YY SEK

has mainly been driven by decisions made at the national level of government as they has to an increasing extent delegated and also imposed new obligations on the local public authorities. The combination of an receding population base and an expanding local public sector may put a lot of economic stress on the local public governments. However, this effect may be neutralized by a higher average income growth.

Much of the empirical literature on regional growth has been explanatory in the sense that they have tried to find what factors are important determinants of regional growth. Many of these studies has taken the so called convergence hypothesis as point of departure, e.g. that initially 'poor' regions tend to grow faster and eventually catch up with 'richer' region. This hypothesis is predicted by neoclassical growth theory as presented by Solow (1956) and Swan (1956). For instance, Barro and Sala-i-Martin (1992, 1995) find support for the convergence hypothesis for the U.S. states. In an application using data on Swedish counties, Persson (1997) reports similar results. However, some authors have taken the lack of convergence as evidence against the neoclassical growth model; see for instance Romer (1986) and Lucas (1988). Still other studies has focused attention on a broader set of possible determinants of regional growth, such as human capital, labor market characteristics, local public expenditures and investments, intergovernmental grants, demographic characteristics and measures of political stability and leadership. One example is Aronsson, Lundberg and Wikström (2001) who, besides convergence, also find labor market characteristics to be important factors of regional growth in Sweden. Using data on Swedish municipalities during the period 1981-1990, Lundberg (2001) finds, in addition to support for the convergence hypothesis, local public expenditures and income tax rates to be important determinants of average income growth and net migration at the local level of government.

This paper complement previous studies of average income growth using Swedish data in that we do not try to explain the causes of regional growth. Instead, this paper focus attention on which regions that has experienced a larger average income growth and to what extent these regions are more spatially clustered than could be

and the local public expenditures has increased from X to XX SEK per capita measured in 2001 money value.

expected from pure chance. In addition, we also focus attention on the importance of the specification of neighbors, an issue that we often find neglected, or at least not sufficiently discussed in previous papers which makes use of spatial econometric methods or tests for spatial correlation. The main purpose is to analyze if the results differ depending on the weights matrix used. Hence, we allow for a large set of possible definitions of neighbors based on travelling time by car between municipal centers and on the criterion that the municipalities share a common border.

The remaining of this paper is organized as follows. The two test statistics used, the Moran's I and the new G_i^*-statistic, are described in Section 3. The data set used, definition of the concept 'neighbors', e.g. the definition of the weights matrix, and potential data problems are discussed in Section 3. This section is followed by the empirical results presented in Section 4. Concluding remarks are given in Section 5.

2 Statistical tests for spatial correlation

2.1 Moran's I

The Moran's I, which build on the work by Moran (1948, 1950), is probably the most frequently used test for spatial correlation. Consider a data set on average income growth rates (y) covering n Swedish municipalities. Assume that $\mathbf{W}$ is a weighting matrix of dimension $(n \times n)$ whose elements assigns the neighbors to each municipality. The weighting matrices used here can be characterized as $\mathbf{W} = \{w_{ij}\}$ such that $w_{ij} > 0$ if i and j are neighbors, otherwise $w_{ij} = 0$. Using row-standardized weights, which is the preferable way of interpreting this test, $1 \geq w_{ij} \geq 0 \; \forall \; i, j$ and $\sum W_i = 1$. Then, the Moran's I is calculated as

$$I = \frac{n}{c} \sum_i \sum_j w_{ij} \left(y_i - \mu\right) \left(y_j - \mu\right) \frac{1}{\sum_i \left(y_i - \mu\right)^2}$$

where c is a scaling constant, y_i and y_j are observation for locations i and j with mean μ. The test statistic is compared with its theoretical mean, $E(I) = -1/(n-1)$. Hence, $E(I) \to 0$ as $n \to \infty$. The null hypothesis $H_0 : I = -1/(n-1)$ is tested against the alternative $H_a : I \neq -1/(n-1)$. If H_0 is rejected then there are two alternative interpretations depending on whether the test statistic I is significantly larger or lower than its expected value. If H_0 is rejected and $I > -1/(n-1)$, this indicates a positive spatial correlation meaning that municipalities with similar values are more spatially clustered than could be caused by chance. If H_0 is rejected and $I < -1/(n-1)$ this indicates a negative spatial correlation, municipalities with high and low values are mixed together. A perfect negative spatial correlation is characterized by a checkerboard pattern of high and low values. As the test statistic is to be compared to its theoretical mean, inferences is often based on the z-statistic

$$z = [I - E(I)] / SD(I)$$

where $SD(I)$ is the theoretical standard deviation of I. If $z > |1.98|$, I is at the 95-percent level of significance different from $-1/(n-1)$ indicating either a negative or a positive spatial correlation.

2.2 G-statistic

The other test for spatial correlation used here is the new G_i^*-statistics developed by Ord and Getis (1995) which build on the 'old' G_i^*-statistics suggested by Getis and Ord (1992). Like the Moran's I, the basic idea behind this test is to define a set of neighbors for each municipality, i.e. municipalities that fall within a specified distance from the municipality in which we are interested. The new G_i^*-statistic then indicates whether a particular municipality is surrounded by a cluster of other municipalities with equivalent growth rates. The new G_i^*-statistic differ from the 'old' version in that the new G_i^*-statistic allow for nonbinary weights. Moreover, the new G_i^*-statistic differ from the new G_i-statistic in that y_i is included in the calculation of G_i^*.

To be more specific, the new G_i^*-statistic is calculated as

$$G_i^* = \frac{\sum_j w_{ij} y_j - (W_i + w_{ii}) \overline{y}}{s \left\{ \left[\left(\left(n \sum_j w_{ij}^2 - (W_i + w_{ii})^2 \right) / (n-1) \right) \right] \right\}^{1/2}}, \text{for all } j$$

Here H_0 :$G_i^* = 0$ is tested against the alternative H_a :$G_i^* \neq 0$ where H_0 is the absence of spatial clustering. If H_0 is rejected, two possible interpretations arise. A positive and significant test statistic indicates that the municipality is surrounded by other municipalities with high growth rates, a negative and significant test statistic indicates the opposite while $G_i^* = 0$ indicate no spatial correlation. This test complements the Moran's I in two ways. The G_i^*-statistic tells us which municipalities are surrounded by other municipalities with similar growth rates, not just that there is a cluster of municipalities with similar growth rates. It also tells us whether there is a clustering of high or low growth rates.

3 Data, definition of neighbors and potential data problems

3.1 Data

The data set used in this study originate from two sources. Information on average income growth is based on the official statistics provided by Statistics Sweden and refer to the Swedish municipalities during the period 1981-1999. During this period, the number of municipalities varied between 279 in 1981 and 288 in 1999. Those municipalities whose borders have been changed during this period are excluded from the analysis. The reason is that it is difficult to obtain comparable data on average income growth for those municipalities. This leaves us with a data set covering 269 municipalities during a period of 19 years. The growth rate of the average income level is calculated as $y_i = \ln (Y_{i,t}/Y_{i,t-T})$ where Y is the average income level for the subpopulation aged 20 or above.

The weighting matrices used here are based on the travelling time by car between municipal centers. This information has been provided by The Swedish Road Administration and is based on the road network and speed limits in 1985. Descriptive statistics are presented in Table 1. To clarify, the average income growth has on average been 142-percent during this period and the average traveling time between municipal centers are 330.74 minutes or 5 hours and 30 minutes.

Table 1. Descriptive statistics.

Variable	Mean	Standard deviation	Min	Max
y_i	1.42	0.08	1.25	2.06
w_{ij}	330.74	-	5.87	1 212.00

3.2 Definition of neighbors, the weights matrix W

One of the more crucial and delicate problems in most empirical studies where the spatial dimension in the data is an issue is the specification of the weights matrix **W**. As, in this case, **W** is an $n \times n$ matrix it is impossible to estimate its elements. This means that the elements in **W** has to be specified a priori from some criteria. The definition of the elements in **W** is of grate importance as **W** is crucial for the results. So, the question is which municipalities are to be considered as neighbors and why?

As we focus attention on geographical clusters, it seems natural to base the definition of neighbors on some geographical criteria. One obvious definition of neighbors are municipalities that share the same border. However, consider the situation where municipality i border on j and k, and l border on m where l and m do not border on i. There are no roads directly connecting i and j while there is a highway connecting i and k. Should i and j be regarded as as close neighbors as i and k? Should i and j be considered as neighbors at all? Furthermore, assume that if you are travelling by car, you have to pass through k to get from i to j. This trip takes 30 minutes. Instead, if you are to take the car between i and l it takes 20 minutes even though i and m do not share a common border. Then, are i and l to be considered as more closely related compared to i and j? And if it takes you 45 minutes to travel by car from l to m should l and m be considered as neighbors at all?

Here, the definition of neighbors is based on two criteria, either as two municipalities who share a common border or on the travelling time by car between municipal canters. We elaborate with the following weights matrices;

- $\mathbf{W1}, \mathbf{W2}, \mathbf{W5}, \mathbf{W10}$: Neighbors are defined as the nearest, the two nearest, the five nearest and the 10 nearest municipalities respectively.

- **WBin30**, **WBin45**, **WBin60**, **WBin75**: Neighbors are defined as those municipalities located within the range of 30, 45, 60 and 75 minutes travel time by car.

- **WB**: Neighbors are defined as those municipalities who share a common border.

- **WInv**: The elements in **W** are defined as $w_{ij} = 1/d_{ij}$ where d_{ij} is the travelling time by car between municipalities i and j.

- **WInv30**, **WInv45**, **WInv60**, **WInv75**: The elements in **W** are defined as $w_{ij} = 1/d_{ij}$ where d_{ij} is the travelling time by car between municipalities i and j with cut off values of 30, 45 60 and 75 minutes respectively.

As a large set of different weights matrices are used, we reduce the risk for misinterpretations due to the fact that the weights matrix is incorrectly specified.

- Descriptive statistics of the elements in the different weights matrices are presented in Table 2.

Table 2. Descriptive statistics of the **W** matrices.

Weights matrix	Mean	Min	Max	Weights matrix	Mean	Min	Max
W1	1.00	0.00	1.00	**WB**	0.21	0.10	1.00
W2	0.50	0.00	0.50	**WInv**	0.00	0.00	0.02
W5	0.20	0.00	0.20				
W10	0.10	0.00	0.10				
WBin30	0.27	0.06	1.00	**WInv30**	0.05	0.03	0.17
WBin45	0.16	0.05	1.00	**WInv45**	0.04	0.02	0.17
WBin60	0.10	0.04	1.00	**WInv60**	0.03	0.02	0.17
WBin75	0.07	0.03	1.00	**WInv75**	0.02	0.01	0.17

3.3 Potential data problems

As described in the data section above, the municipalities whose borders have been changed during this period are excluded from the data. This is, of course, unfortunate as it will automatically induce spatial 'holes' in the data set. However, on average, the municipalities excluded from the analysis are quite small both in geographic and population terms.

4 Results

4.1 Moran's I

If the variable that is to be tested follows a normal distribution, the I-statistic is compared with its theoretical mean, $-1/(n-1)$. However, if this is not the case, the reference distribution for I should be generated empirically. This is done by randomly reshuffling the observed values over all locations. A Wald test statistic of 3 238 with 2 degrees of freedom reveals non-normality in the variable y.[3] Consequently, the reference distribution of the Moran's I is generated using the permutation approach.

The Moran's I for different weighting matrices are presented in Table 2. Independent of the weighting matrix used, the results suggest a positive and at the 95-percent level significant spatial correlation indicating that high or low values are spatially clustered. What differs is the level of significance which, with one exception, tend to increase with the number of neighbors assigned to each municipality.

[3]The Wald statistic is χ^2-distributed and calculated as $W = n\left[b_1^2/6 + (b_2-3)^2/24\right]$ where $b_1 = (1/n)\sum_i (y_i-\mu)^3 / \left(\sigma^2\right)^{3/2}$ (skewness), $b_2 = (1/n)\sum_i (y_i-\mu)^4 / \left(\sigma^2\right)^2$ (kurtosis) and σ is the standard deviation of y.

Table 2. Moran's I

Weights matrix	z-value	Mean	Weights matrix	z-value	Mean
W1	3.20	-0.018	**WB**	6.31	-0.009
W2	4.87	-0.011	**WInv**	7.84	-0.004
W5	4.54	-0.006			
W10	6.67	-0.005			
WBin30	2.26	-0.006	**WInv30**	2.60	-0.006
WBin45	2.40	-0.009	**WInv45**	2.62	-0.010
WBin60	2.84	-0.006	**WInv60**	3.27	-0.007
WBin75	3.94	-0.007	**WInv75**	4.24	-0.008

4.2 The new G_i^*-statistic

In order to make the results from the new G_i^*-statistic easier to overview and interpret, they are presented in map-form. Areas marked dark red indicate significant clusters of municipalities with low average income growth rates while areas marked dark blue indicate significant clusters of high average income growth rates. The colors light red and light blue indicate non-significant clusters of low and high average income growth rates respectively and those municipalities excluded from the analysis are marked in white.

Let us go through the maps and discuss how the clustering pattern change as we elaborate with different weights matrices. Figure 1 shows the results using **W1**. Here, there are only three areas displayed in red, Malmö in the south, Nynäshamn near Stockholm and Hällefors in the middle part of the country. The areas displayed in blue are Danderyd, Täby Lidingö and Värmdö near Stockholm and Gisslaved and Gnosjö south of the lake Vättern.

5 Concluding remarks

The main purpose in this paper has been to test the hypothesis that municipalities with similar average income growth rates are more spatially clustered than could be expected from pure chance and to what extent these results are sensitive to the definition of the spatial weights matrix. In order to accomplish this task, we make use of two commonly used test statistics for spatial correlation, the Moran's I and the new G_i^*-statistic.

The results from the Moran's I suggest a positive spatial cluster meaning that municipalities with high or low average income growth rates are mors spatially clustered than could be expected from pure chance. The level of significance of the Moran's I tend to increase as the number of neighbors assigned to each municipality increases. One possible interpretation of this result is that...

The results from the G_i^*-statistic suggest a spatial cluster of high average growth rates around the area of Jönköping and Stockholm. These two areas tend to increase as the number of municipalities assigned to each municipality increases. One interpretation of this result is...

References

[1] Anselin, L. (1988): Spatial Econometrics

[2] Aronsson, T., J. Lundberg and M. Wikström (2001):

[3] Aronsson, T. and J. Lundberg (2002):

[4] Burridge, P. (1980): On the Cliff-Ord Test for Spatial Correlation, *Journal of the Royal Statistical Society* B 42, 107-108.

[5] Cliff, A. and J.K. Ord (1972): Testing for Spatial Autocorrelation Among Regression Residuals, *Geographical Analysis* 4, 267-284.

[6] Getis, A. and J.K. Ord (1992): The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis* 24, 189-206.

[7] Moran, P. (1950): Notes on Continuos Stochastic Phenomenon, *Biometrica* 37, 17-23.

[8] Lundberg, J. (2001):

[9] Ord, J.K. and A. Getis (1995): Local Spatial Autocorrelation Statistics: Distributional Issues and an Application, *Geographical Analysis* 27:4, 286-305.

[10] Pinkse, J. (1999): Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models

Appendix