Province of KwaZulu-Natal

Provincial Treasury

IMES Unit

# Estimating the Spatial Regression Function for KwaZulu-Natal[1]

**Clive Coetzee**

General Manager: IMES Unit

Economist

clive.coetzee@kzntreasury.gov.za

033 897 4538

*Working Paper 2E: April 2015*

[1] The views expressed in this working paper are the views of the author and might not necessarily reflect the views of the Provincial Treasury. All rights reserved – 2009-2015

## Introduction

The general purpose of linear regression analysis is to find a (linear) relationship between a dependent variable and a set of explanatory variables:

$$\gamma = X\beta + \varepsilon$$

The method of ordinary least squares (OLS) estimation is referred to as a Best Linear Unbiased Estimator (BLUE). The OLS estimates $\beta$ by minimizing the sum of squared prediction errors, hence, least squares. In order to obtain the BLUE property and make statistical inferences about the population regression coefficients from the estimated b, certain assumptions about the random error of the regression equation need to be made. These, according to Anselin (2005) include:

a) the random errors have a mean of zero (there is no systematic misspecification or bias in the population regression equation);

b) the random errors have a constant variance (homoskedasticity) and are uncorrelated;

c) the random errors have a normal distribution.

These assumptions may not always be satisfied in practice. When a value observed in one location depends on the values observed at neighboring locations, there is a spatial dependence. And spatial data may show spatial dependence in the variables and error terms.

Why should spatial dependence occur? There are two reasons commonly given (see LeSage, 2009). First, data collection of observations associated with spatial units may reflect measurement error. This happens when the boundaries for which information is collected do not accurately reflect the nature of the underlying process generating the sample data.

A second reason for spatial dependence is that the spatial dimension of a social or economic characteristic may be an important aspect of the phenomenon. For example, based on the premise that location and distance are important forces at work, regional science theory relies on notions of spatial interaction and diffusion effects, hierarchies of place and spatial spillovers.

Spatial regression models therefore include relationships between variables and their neighboring values, for example it allows us to examine the impact that one observation has on other proximate observations. In this article the focus will be on regional unemployment and the possible impact of neighboring variables.

## Provincial Unemployment Characteristics

The table below displays some of the major characteristics of the provincial labour market from 2008 until 2014. The economically active population increased from about 6 million to 6.6 million over the period where as the number of people employed stayed almost constant at 2.52 million. The number of people unemployed declined over the period by about 69 000, however the number of people not economically active and discouraged both increases significantly over the period by 724 000 and 437 000, respectfully. It thus seems that the decrease in the number of people unemployed in the province is not because of job creation, but rather because of people withdrawing from the job market.

**Table 1: Provincial Labour Market Trends, 2008 to 2014 (total per quarter)**

| . | Population of working age (15–64 years) | Labour Force | Employed | Unemployed | Not Economically Active | Discourage Work Seekers |
|---|---|---|---|---|---|---|
| Q1 2008 | 5 993 000 | 3 257 000 | 2 525 000 | 732 000 | 2 736 000 | 179 000 |
| Q2 2008 | 6 016 000 | 3 280 000 | 2 560 000 | 720 000 | 2 736 000 | 165 000 |
| Q3 2008 | 6 041 000 | 3 246 000 | 2 541 000 | 705 000 | 2 795 000 | 201 000 |
| Q4 2008 | 6 065 000 | 3 250 000 | 2 584 000 | 666 000 | 2 815 000 | 215 000 |
| Q1 2009 | 6 090 000 | 3 194 000 | 2 490 000 | 704 000 | 2 896 000 | 260 000 |
| Q2 2009 | 6 114 000 | 2 995 000 | 2 430 000 | 565 000 | 3 119 000 | 436 000 |
| Q3 2009 | 6 139 000 | 2 979 000 | 2 433 000 | 546 000 | 3 160 000 | 463 000 |
| Q4 2009 | 6 164 000 | 2 944 000 | 2 386 000 | 558 000 | 3 220 000 | 435 000 |
| Q1 2010 | 6 188 000 | 2 947 000 | 2 385 000 | 562 000 | 3 241 000 | 471 000 |
| Q2 2010 | 6 213 000 | 2 937 000 | 2 332 000 | 605 000 | 3 276 000 | 478 000 |
| Q3 2010 | 6 238 000 | 2 854 000 | 2 296 000 | 558 000 | 3 384 000 | 527 000 |
| Q4 2010 | 6 262 000 | 2 913 000 | 2 348 000 | 565 000 | 3 349 000 | 499 000 |
| Q1 2011 | 6 286 000 | 2 915 000 | 2 337 000 | 578 000 | 3 371 000 | 548 000 |
| Q2 2011 | 6 311 000 | 3 000 000 | 2 399 000 | 601 000 | 3 311 000 | 555 000 |
| Q3 2011 | 6 336 000 | 2 973 000 | 2 417 000 | 556 000 | 3 363 000 | 508 000 |
| Q4 2011 | 6 360 000 | 3 047 000 | 2 473 000 | 574 000 | 3 313 000 | 504 000 |
| Q1 2012 | 6 384 000 | 3 027 000 | 2 423 000 | 604 000 | 3 357 000 | 548 000 |
| Q2 2012 | 6 408 000 | 3 008 000 | 2 429 000 | 579 000 | 3 400 000 | 556 000 |
| Q3 2012 | 6 432 000 | 3 073 000 | 2 441 000 | 632 000 | 3 359 000 | 534 000 |
| Q4 2012 | 6 456 000 | 3 073 000 | 2 399 000 | 674 000 | 3 383 000 | 553 000 |
| Q1 2013 | 6 479 000 | 3 049 000 | 2 424 000 | 625 000 | 3 430 000 | 544 000 |
| Q2 2013 | 6 502 000 | 3 136 000 | 2 440 000 | 696 000 | 3 366 000 | 573 000 |
| Q3 2013 | 6 527 000 | 3 235 000 | 2 569 000 | 666 000 | 3 292 000 | 541 000 |
| Q4 2013 | 6 549 000 | 3 154 000 | 2 527 000 | 627 000 | 3 395 000 | 573 000 |
| Q1 2014 | 6 572 000 | 3 186 000 | 2 527 000 | 659 000 | 3 386 000 | 620 000 |
| Q2 2014 | 6 596 000 | 3 249 000 | 2 480 000 | 769 000 | 3 347 000 | 615 000 |
| Q3 2014 | 6 619 000 | 3 187 000 | 2 419 000 | 768 000 | 3 432 000 | 638 000 |
| Q4 2014 | 6 643 000 | 3 183 000 | 2 520 000 | 663 000 | 3 460 000 | 616 000 |

(Source, Statistics SA)

The table below shows that the provincial unemployment rate (narrow definition) decreased slightly from about 22 percent to about 20 percent over the period. However, the employment rate (number of people employed divided by the economically active population) also decreased over the period from 42 percent to 37 percent. The labour force participation rate (labour force divided by the economically active population) also unfortunately deceased whilst the expanded unemployment rate increased significantly from 28 percent to over 40 percent over the period, respectively.

**Table 2: Provincial Labour Market Rates, 2008 to 2014 (% per quarter)**

| | Unemployment rate | Employed / population ratio (Absorption) | Labour force participation rate | Unemployment Rate (Expanded Def) |
|---|---|---|---|---|
| Q1 2008 | 22.47 | 42.13 | 54.35 | 27.97 |
| Q2 2008 | 21.95 | 42.55 | 54.52 | 26.98 |
| Q3 2008 | 21.72 | 42.06 | 53.73 | 27.91 |
| Q4 2008 | 20.49 | 42.61 | 53.59 | 27.11 |
| Q1 2009 | 22.04 | 40.89 | 52.45 | 30.18 |
| Q2 2009 | 18.86 | 39.74 | 48.99 | 33.42 |
| Q3 2009 | 18.33 | 39.63 | 48.53 | 33.87 |
| Q4 2009 | 18.95 | 38.71 | 47.76 | 33.73 |
| Q1 2010 | 19.07 | 38.54 | 47.62 | 35.05 |
| Q2 2010 | 20.60 | 37.53 | 47.27 | 36.87 |
| Q3 2010 | 19.55 | 36.81 | 45.75 | 38.02 |
| Q4 2010 | 19.40 | 37.50 | 46.52 | 36.53 |
| Q1 2011 | 19.83 | 37.18 | 46.37 | 38.63 |
| Q2 2011 | 20.03 | 38.01 | 47.54 | 38.53 |
| Q3 2011 | 18.70 | 38.15 | 46.92 | 35.79 |
| Q4 2011 | 18.84 | 38.88 | 47.91 | 35.38 |
| Q1 2012 | 19.95 | 37.95 | 47.42 | 38.06 |
| Q2 2012 | 19.25 | 37.91 | 46.94 | 37.73 |
| Q3 2012 | 20.57 | 37.95 | 47.78 | 37.94 |
| Q4 2012 | 21.93 | 37.16 | 47.60 | 39.93 |
| Q1 2013 | 20.50 | 37.41 | 47.06 | 38.34 |
| Q2 2013 | 22.19 | 37.53 | 48.23 | 40.47 |
| Q3 2013 | 20.59 | 39.36 | 49.56 | 37.31 |
| Q4 2013 | 19.88 | 38.59 | 48.16 | 38.05 |
| Q1 2014 | 20.68 | 38.45 | 48.48 | 40.14 |
| Q2 2014 | 23.67 | 37.60 | 49.26 | 42.60 |
| Q3 2014 | 24.10 | 36.55 | 48.15 | 44.12 |
| Q4 2014 | 20.83 | 37.93 | 47.92 | 40.18 |

(Source, Statistics SA)

The statistics suggest that on average about 40 percent of the people that were unemployed in the province were new entrants. Job losers on average accounted for about 32 percent. Job leavers only account on average for about 8 percent, whilst re-entrants and other on average account for the remainder (4 percent and 26 percent. respectively).

The job losses in the province occurred in almost all of the economic sectors, for example the number of people employed in the agriculture sector decreased by about 50 000 or by almost 40 percent. Manufacturing lost about 84 000 jobs, etc.  The sectors that did see an increase in jobs are construction with about 35 000 jobs and community and social services with 107 000 jobs.

The number of people employed in the provincial formal sector stayed almost constant at about 1.7 million over the period, whilst the number of people in the informal sector decreased from about 490 000 to about 470 000.

**Table 3: Employment per Economic Sector, 2008 to 2014, (total per quarter)**

| | Agriculture | Mining | Manufacturing | Utilities | Construction | Trade | Transport | Finance | Community and social services | Private households |
|---|---|---|---|---|---|---|---|---|---|---|
| Q1 2008 | 130 000 | 7 000 | 436 000 | 13 000 | 205 000 | 578 000 | 166 000 | 282 000 | 483 000 | 224 000 |
| Q2 2008 | 157 000 | 9 000 | 423 000 | 16 000 | 214 000 | 570 000 | 165 000 | 289 000 | 493 000 | 225 000 |
| Q3 2008 | 148 000 | 8 000 | 382 000 | 14 000 | 222 000 | 605 000 | 167 000 | 262 000 | 473 000 | 260 000 |
| Q4 2008 | 163 000 | 9 000 | 390 000 | 11 000 | 240 000 | 579 000 | 182 000 | 266 000 | 468 000 | 277 000 |
| Q1 2009 | 136 000 | 9 000 | 384 000 | 12 000 | 219 000 | 502 000 | 185 000 | 271 000 | 489 000 | 283 000 |
| Q2 2009 | 103 000 | 12 000 | 411 000 | 8 000 | 232 000 | 511 000 | 174 000 | 273 000 | 460 000 | 245 000 |
| Q3 2009 | 113 000 | 7 000 | 390 000 | 4 000 | 237 000 | 505 000 | 172 000 | 298 000 | 469 000 | 238 000 |
| Q4 2009 | 110 000 | 7 000 | 384 000 | 9 000 | 229 000 | 506 000 | 154 000 | 290 000 | 471 000 | 227 000 |
| Q1 2010 | 115 000 | 8 000 | 403 000 | 5 000 | 231 000 | 501 000 | 152 000 | 264 000 | 480 000 | 224 000 |
| Q2 2010 | 110 000 | 8 000 | 358 000 | 8 000 | 225 000 | 487 000 | 155 000 | 298 000 | 462 000 | 221 000 |
| Q3 2010 | 118 000 | 6 000 | 358 000 | 9 000 | 218 000 | 485 000 | 185 000 | 263 000 | 449 000 | 204 000 |
| Q4 2010 | 114 000 | 13 000 | 358 000 | 16 000 | 218 000 | 502 000 | 177 000 | 255 000 | 493 000 | 203 000 |
| Q1 2011 | 102 000 | 14 000 | 374 000 | 17 000 | 225 000 | 513 000 | 167 000 | 246 000 | 492 000 | 187 000 |
| Q2 2011 | 94 000 | 5 000 | 383 000 | 18 000 | 228 000 | 534 000 | 171 000 | 259 000 | 485 000 | 222 000 |
| Q3 2011 | 95 000 | 15 000 | 385 000 | 8 000 | 229 000 | 579 000 | 149 000 | 272 000 | 468 000 | 218 000 |
| Q4 2011 | 95 000 | 12 000 | 417 000 | 6 000 | 233 000 | 584 000 | 163 000 | 262 000 | 499 000 | 201 000 |
| Q1 2012 | 92 000 | 17 000 | 369 000 | 8 000 | 203 000 | 539 000 | 175 000 | 260 000 | 529 000 | 232 000 |
| Q2 2012 | 92 000 | 19 000 | 372 000 | 12 000 | 204 000 | 518 000 | 188 000 | 270 000 | 532 000 | 222 000 |
| Q3 2012 | 90 000 | 18 000 | 376 000 | 9 000 | 213 000 | 497 000 | 197 000 | 284 000 | 539 000 | 218 000 |
| Q4 2012 | 98 000 | 30 000 | 343 000 | 13 000 | 213 000 | 470 000 | 179 000 | 289 000 | 549 000 | 216 000 |
| Q1 2013 | 95 000 | 25 000 | 348 000 | 10 000 | 208 000 | 476 000 | 188 000 | 301 000 | 545 000 | 229 000 |
| Q2 2013 | 92 000 | 25 000 | 353 000 | 9 000 | 232 000 | 482 000 | 179 000 | 285 000 | 564 000 | 219 000 |
| Q3 2013 | 108 000 | 21 000 | 352 000 | 14 000 | 227 000 | 554 000 | 200 000 | 280 000 | 590 000 | 223 000 |
| Q4 2013 | 96 000 | 6 000 | 345 000 | 9 000 | 221 000 | 567 000 | 198 000 | 282 000 | 572 000 | 231 000 |
| Q1 2014 | 96 000 | 5 000 | 361 000 | 22 000 | 247 000 | 570 000 | 178 000 | 274 000 | 561 000 | 213 000 |
| Q2 2014 | 85 000 | 6 000 | 329 000 | 18 000 | 241 000 | 550 000 | 184 000 | 243 000 | 594 000 | 230 000 |
| Q3 2014 | 79 000 | 8 000 | 343 000 | 20 000 | 241 000 | 511 000 | 173 000 | 237 000 | 599 000 | 210 000 |
| Q4 2014 | 102 000 | 4 000 | 362 000 | 17 000 | 280 000 | 530 000 | 169 000 | 254 000 | 574 000 | 228 000 |

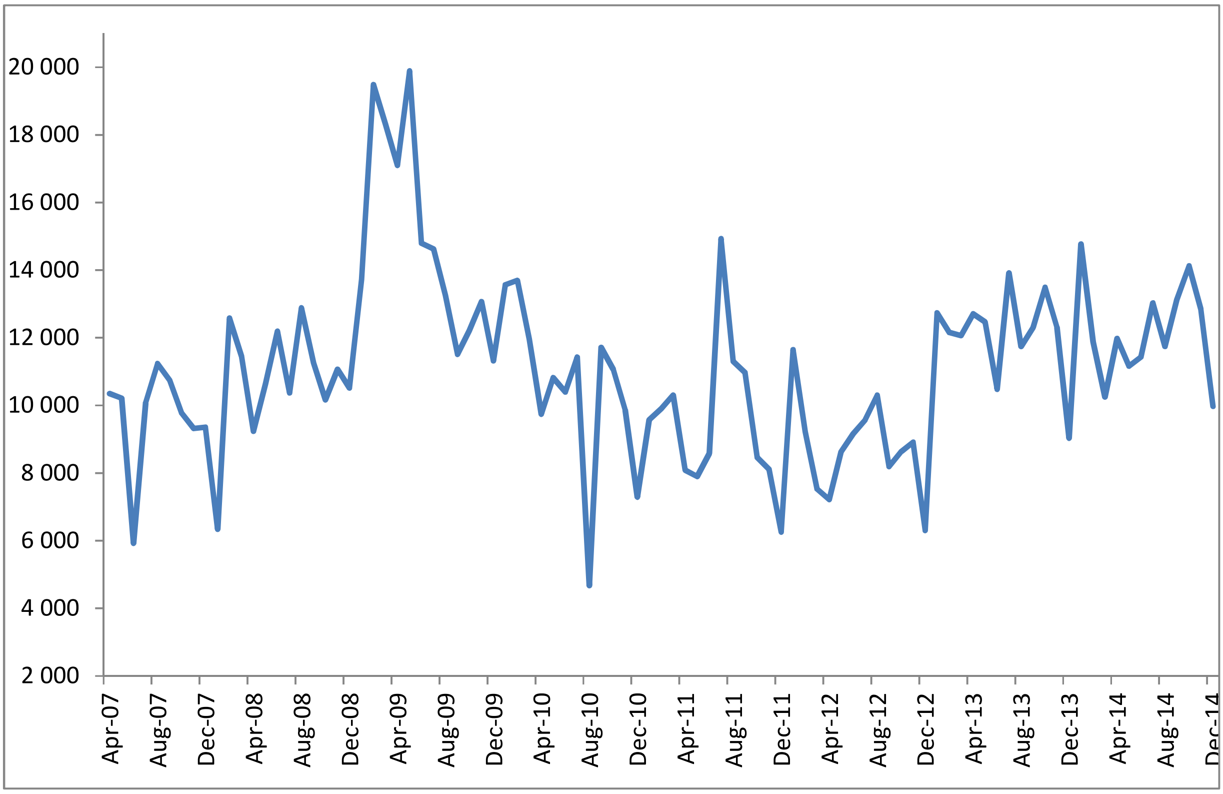
(Source, Statistics SA)

The graph below displays the number of new unemployment insurance applications in the province from April 2007 until December 2014. On average there are about 11 000 new applications per month. Unemployment insurance benefits are available if an employee loses his/her job, because of dismissal, contract termination by the employer or their insolvency, i.e., it is therefore a measure or proxy for the number of people becoming unemployed per month.

However it's not a perfect measure or proxy since it also includes temporary unemployment, for example, if a contributor's fixed-term contract has come to an end, or retirement. Bhorat et al. (2013) indicates that more than 40 percent of all claims in the period originated from claimants whose contracts had expired, and they therefore found themselves out of work. In turn, claims from dismissals (24.2 percent) and retrenchments (23.1 percent) each accounted for just under one-fifth of total claims between 2005 and 2011.

The data shows that the number of new applications increased significantly during the 2007 and 2008 financial crises and decreased significantly during the so-called economic recovery period. The number of new applications then decreased fairly slowly during 2011 and 2012, but has been increasing from 2013 and 2014.

**Graph 1: Number of New Unemployment Insurance Applications, 2007 to 2014 (total per month)**



(Source, SA Department of Labour, Own calculations)

The below table displays the average monthly number of new unemployment insurance applications per region over the period. The Durban/Pinetown region experienced the largest number of new applications followed by the Pietermaritzburg and Richards Bay regions. In general all regions experienced an increase in the number of new applications over the period. In general the number of new applications increased by about 6 percent in the more urban regions compared to 4 percent in the more rural regions over the period. On average, for every one new application in the more rural regions there are four new applications in the more urban regions.

**Table 4: Number of New Unemployment Insurance per Region, 2007 to 2014 (average per month)**

| | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|
| Ethekwini | 1 653 | 2 028 | 2 897 | 2 246 | 1 986 | 1 926 | 2 746 | 2 954 |
| Dundee | 87 | 66 | 93 | 128 | 109 | 140 | 213 | 198 |
| Estcourt | 217 | 186 | 208 | 189 | 402 | 263 | 302 | 224 |
| Kokstad | 273 | 234 | 248 | 154 | 146 | 162 | 286 | 247 |
| Ladysmith | 298 | 374 | 424 | 358 | 351 | 387 | 442 | 382 |
| Newcastle | 508 | 448 | 397 | 256 | 309 | 250 | 397 | 463 |
| Pietermaritzburg | 1 103 | 1 155 | 1 297 | 1 085 | 1 337 | 1 201 | 1 385 | 1 648 |
| Pinetown | 914 | 1 068 | 1 664 | 1 130 | 967 | 792 | 1 207 | 1 300 |
| Port Shepstone | 579 | 712 | 684 | 639 | 551 | 587 | 778 | 605 |
| Prospecton | 1 056 | 1 312 | 2 333 | 1 349 | 999 | 771 | 1 021 | 970 |
| Richards Bay | 929 | 1 057 | 1 328 | 922 | 809 | 775 | 1 077 | 1 031 |
| Richmond | 248 | 324 | 240 | 243 | 168 | 126 | 215 | 170 |
| Stanger | 634 | 583 | 1 067 | 507 | 390 | 465 | 718 | 706 |
| Ulundi | 244 | 215 | 229 | 214 | 222 | 227 | 321 | 359 |
| Verulam | 737 | 730 | 1 414 | 876 | 621 | 552 | 762 | 718 |
| Vryheid | 184 | 228 | 416 | 221 | 164 | 155 | 237 | 198 |

(Source, SA Department of Labour, Own calculations)

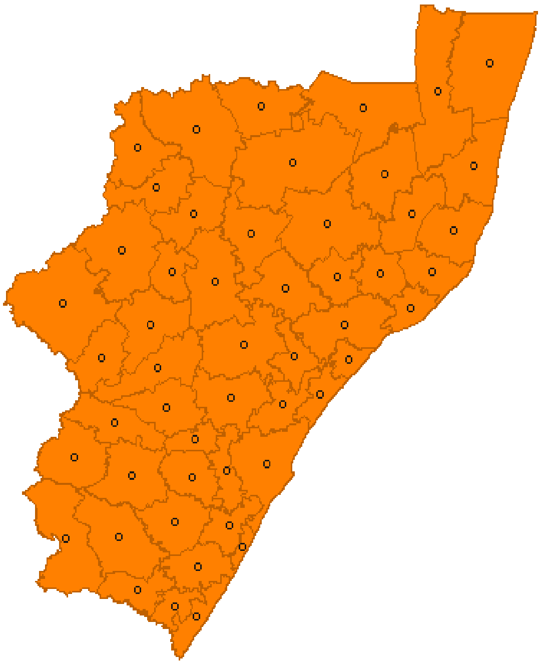
## Visualizing the Regional Unemployment Data

Geographical clusters and/or regions can be analyzed and described by a number of different spatial association statistics as well as visually (quickly and intuitively when the eye and brain look at the map). Each of the graphs below presents a colored map that allows the visualization of the spatial pattern of the unemployment rate per municipality during 1996, 2004 and 2013, respectively. The GIS (geographical information system) programme QGIS was used for the analysis. Data was sourced from various sources including Statistics SA, Global Insight and own sources. Dark areas indicate high levels or concentration of unemployment in that municipality whilst light areas indicate low levels or concentration of unemployment in that municipality.

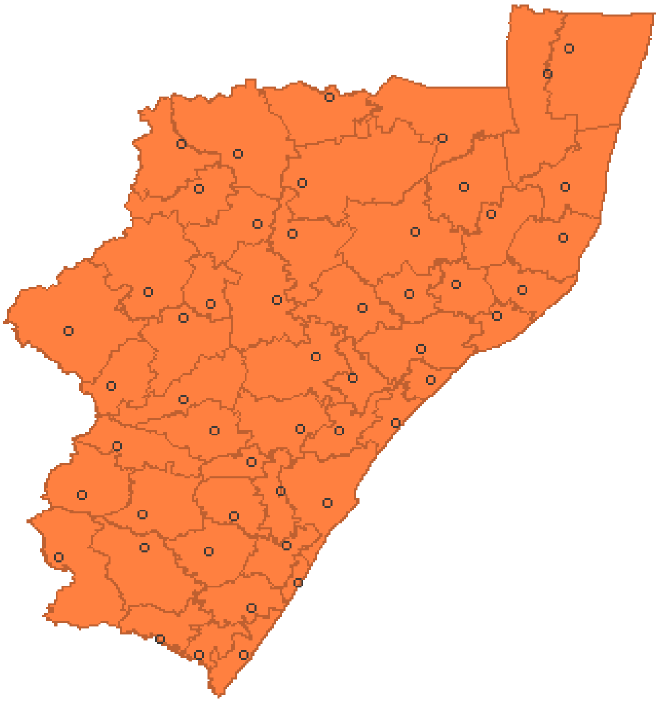
Graph 2 displays the centroids of the 51 municipalities, whereas graph 3 displays the mean centers of each municipality.

The unemployment graphs (graphs 4, 5 and 6) display the number of unemployed people as a percentage of the total labour force per municipality during 1996, 2004 and 2013 in the province. It seems evident that the percentage of people unemployed in the province has been spatially fairly dispersed. However there does seem to be some spatial association, i.e., dark shaded regions (high unemployment) seem to cluster together whilst light shaded regions (low unemployment) seem to cluster together. The graphs also suggest that the unemployment situation worsened from 1996 to 2004 and improved from 2004 to 2013. The graphs in general suggest that nearby or neighboring areas are more alike, i.e., existence of some sort of spatial unemployment relationship.

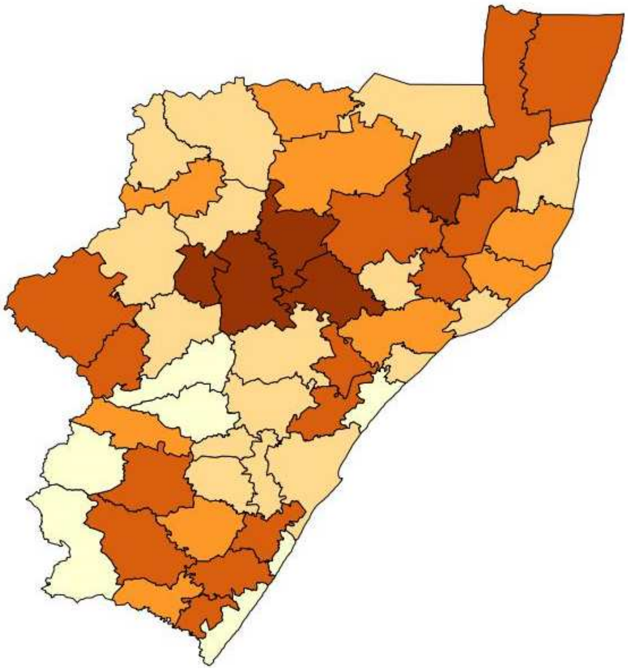
**Graph 2: Municipal Centroids**



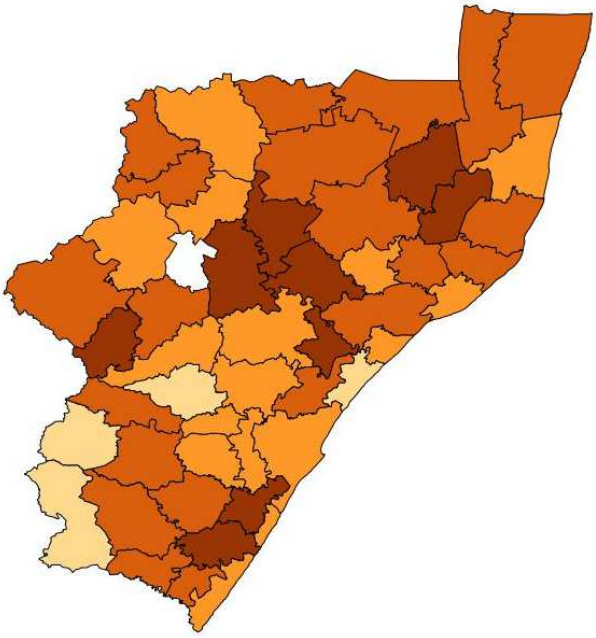
**Graph 3: Municipal Mean Centers**



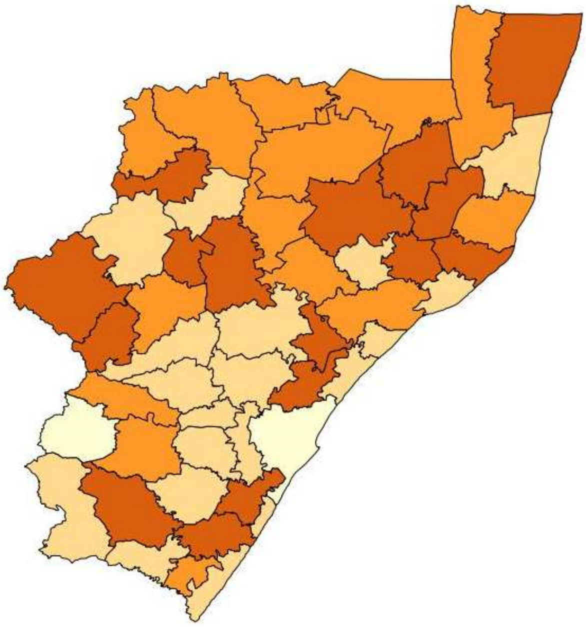
Graph 4: Unemployment 1996



Graph 5:  Unemployment 2004



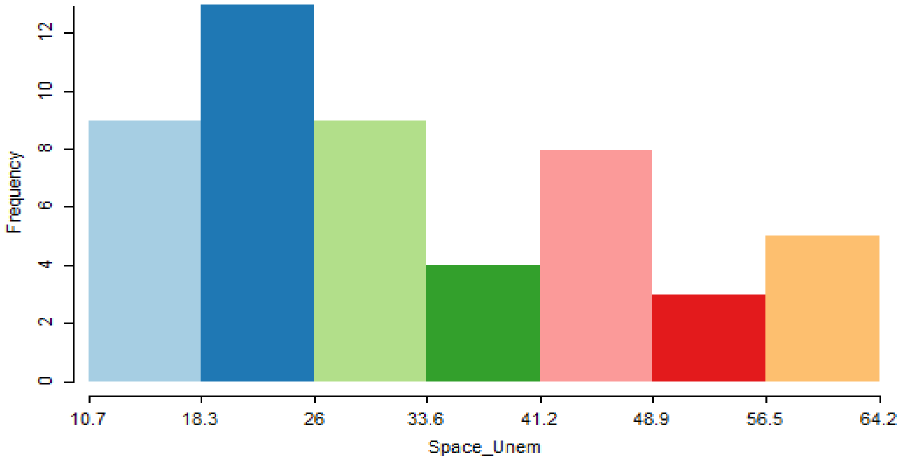
Graph 6:  Unemployment 2013



(Source, Statistics SA, own calculations)

The three graphs below (graph 7, 8 and 9) display the frequency distribution of the municipalities with regard to the unemployment rate during 1996, 2004 and 2013. The general worsening of the unemployment situation from 1996 to 2004 and the general improvement from 2004 to 2013 is again evident.

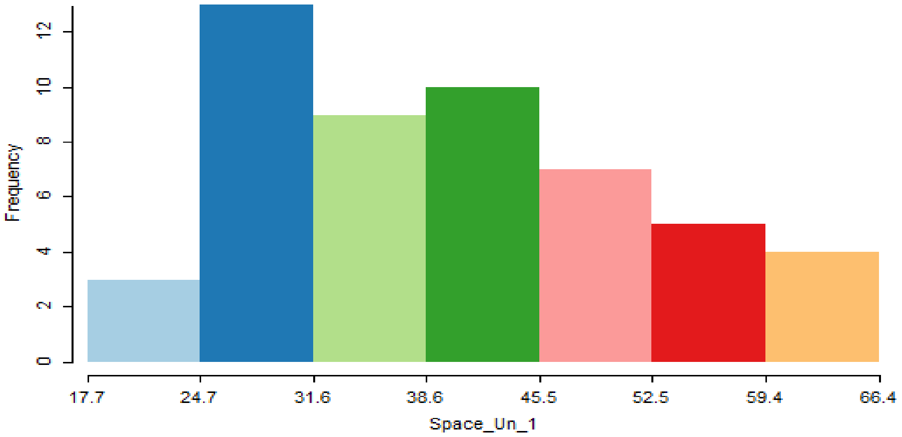
### Graph 7: Frequency Distribution 1996



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| from | 10.7 | 18.3 | 26 | 33.6 | 41.2 | 48.9 | 56.5 |
| to | 18.3 | 26 | 33.6 | 41.2 | 48.9 | 56.5 | 64.2 |
| #obs | 9 | 13 | 9 | 4 | 8 | 3 | 5 |
| % of total | 17.6 | 25.5 | 17.6 | 7.84 | 15.7 | 5.88 | 9.8 |
| sd from mean | -0.971 | -0.45 | 0 | 0.0713 | 0.592 | 1.11 | 1.63 |

min: 10.67, max: 64.16, median: 30.42, mean: 32.5492, s.d.: 14.6639, #obs: 51

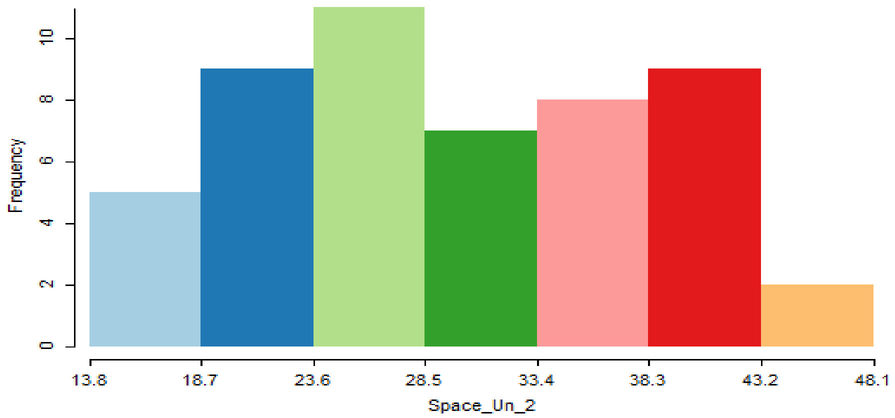
### Graph 8: Frequency Distribution 2004



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| from | 17.7 | 24.7 | 31.6 | 38.6 | 45.5 | 52.5 | 59.4 |
| to | 24.7 | 31.6 | 38.6 | 45.5 | 52.5 | 59.4 | 66.4 |
| #obs | 3 | 13 | 9 | 10 | 7 | 5 | 4 |
| % of total | 5.88 | 25.5 | 17.6 | 19.6 | 13.7 | 9.8 | 7.84 |
| sd from mean | -1.24 | -0.666 | -0.0874 | 0 | 0.491 | 1.07 | 1.65 |

min: 17.71, max: 66.4, median: 39.96, mean: 39.6276, s.d.: 12.0256, #obs: 51

**Graph 9: Frequency Distribution 2013**



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| from | 13.8 | 18.7 | 23.6 | 28.5 | 33.4 | 38.3 | 43.2 |
| to | 18.7 | 23.6 | 28.5 | 33.4 | 38.3 | 43.2 | 48.1 |
| #obs | 5 | 9 | 11 | 7 | 8 | 9 | 2 |
| % of total | 9.8 | 17.6 | 21.6 | 13.7 | 15.7 | 17.6 | 3.92 |
| sd from mean | -1.26 | -0.69 | -0.116 | 0 | 0.459 | 1.03 | 1.61 |

min: 13.8, max: 48.09, median: 28.96, mean: 29.4822, s.d.: 8.52501, #obs: 51

## Determination of the spatial weights matrices

Before constructing a spatial weights matrix, we must make a spatial contiguity matrix by using weight function (Smith, 2009). A *spatial weight matrix* summarizes potential spatial relations between *n* spatial units. Here each *spatial weight*, $w_{ij}$, typically reflects the "spatial influence" of unit *j* on unit *i*. For *n* elements in a geographical system, a spatial contiguity matrix, *C*, can be expressed in the form:

$$C= \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

where *cij* is a measurement used to compare and judge the degree of nearness or the contiguous relationships between region *i* and region *j*. Thus a spatial weights matrix can be defined as:

$$W= \frac{C}{C_o} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

where

$$C_0 = \sum_{i=0}^{n} \sum_{j=0}^{n} Cij \quad , \quad \sum_{i=0}^{n} \sum_{j=0}^{n} wij = 1$$

Spatial contiguity weights - The simplest of these weights simply indicate whether spatial units share a boundary or not. If the set of boundary points of unit i is denoted by bnd(i) then the so-called queen contiguity weights are defined by:

$$W_{ij} = \left\{ \frac{1, \quad bnd(i) \frown bnd(j) \neq \emptyset}{0, otbnd(i) \frown bnd(j) = \emptyset} \right\}$$

k-Nearest Neighbor weights - Let centroid distances from each spatial unit *i* to all units $j \neq i$ be ranked as follows: $d_{ij\,(1)} \leq d_{ij(2)} \leq --- \leq d_{ij(n\ 1)}$. Then for each *k* =1,..,*n* - 1, the set $N_k(i)$ = { j(1), j(2),..,j(k)} contains the *k* closest units to *i* (where for simplicity we ignore ties). For each given *k* , the *k-nearest neighbor* weight matrix, *W* , then has spatial weights of the form:

$$W_{ij} = \left\{ \frac{1, \quad j \in Nk(i)}{0, otherwise} \right\}$$

Radial distance weights - If distance itself is an important criterion of spatial influence, and if *d* denotes a *threshold distance* (or *bandwidth*) beyond which there is no direct spatial influence between spatial units, then the corresponding *radial distance* weight matrix, *W* , has spatial weights of the form:

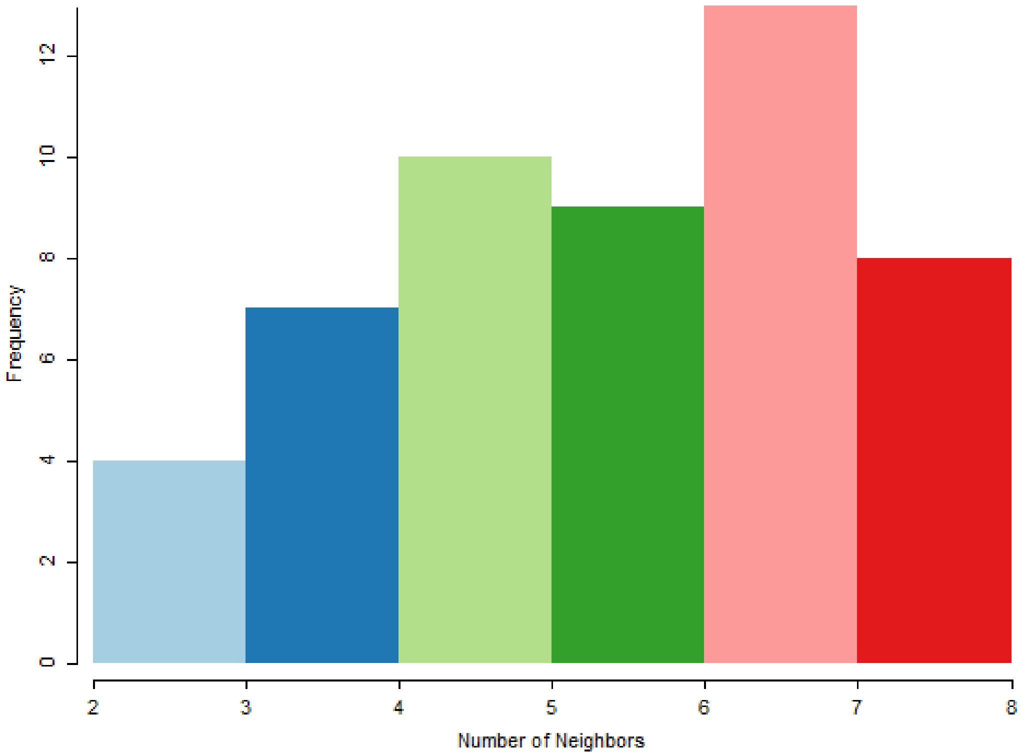
$$W_{ij} = \left\{ \frac{1, 0 \leq dij \leq d}{0, \quad dij > d} \right\}$$

Actual distance values - If distance itself is an important criterion of spatial influence, and if *d* denotes the actual *distance* (1/d = inverse of the distance) then the corresponding *actual distance* weight matrix, *W* , has spatial weights of the form:

W$_{ij}$ = $\{1, 1/dij > 0\}$

The analysis will predominantly make use of the Rook-Based Contiguity weight matrix. The distribution of neighbors is displayed in the graph below. It's evident that the majority of municipalities (12) have about 6 neighbors.

**Graph 10: Rook-Based Contiguity Weight Matrix Distribution of Neighbors**



## Applying Moran's I statistic

In statistics, Moran's I is a measure of spatial autocorrelation developed by Patrick Alfred Pierce Moran (Ward and Gleditsch, 2007). Moran's I takes the form of a classic correlation coefficient in that the mean of a variable is subtracted from each sample value in the numerator. This results in coefficients ranging from (–1) to (+1), where values between (0) and (+1) indicate a positive association between variables, values between (0) and (-1) indicate a negative association, and (0) indicates there is no correlation between variables. The expected value of Moran's I under the null hypothesis of no spatial autocorrelation is E(I) = {-1}/ {N-1}.

Moran’s I:

$$I(d) = \frac{\frac{1}{W}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{hi}(y_h - \bar{y})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{for} \quad h \neq i$$

where:

$I(d)$ = Moran’s I correlation coefficient as a function of distance

$w_{hi}$ = a matrix of weighted values, where elements are a function of distance

1 = $y_h$ and $y_i$ are within a given distance class, for $y_h \neq y_i$

0 = all other cases

$y_h, y_i$ = values of variables at locations h and I

$W$ = sum of the values of the matrix $w_{hi}$

$n$ = sample size

(Wikipedia, http://en.wikipedia.org/wiki/Moran's_I)

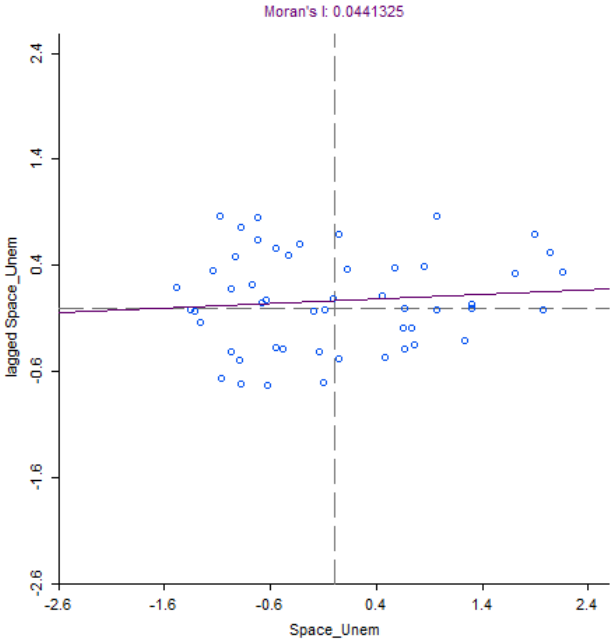
The results for unemployment during 1996, 2004 and 2014 using the contiguity weight matrix are displayed below. The results must be seen in context of the graphical illustrations presented, i.e., graphs 4 to 6. As stated, the graphs suggest some sort of spatial relation or spatial association with regard to the number of unemployed people, in that regions with high levels of unemployment tend to cluster together and vice versa.

However the Moran’s I correlation coefficient for the three periods are very low, i.e.,

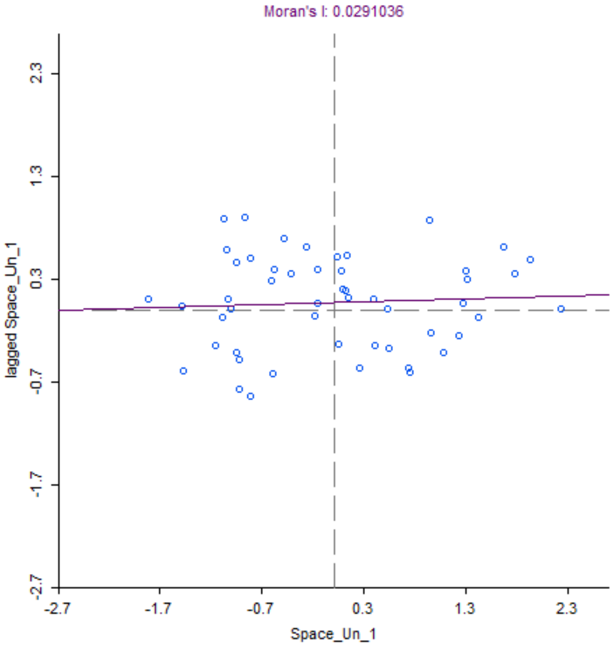
- 0.04 for 1996,
- 0.03 for 2004 and
- 0.02 for 2013

None of the Moran's I correlation coefficients are statistically significant (p=0.22, p=0.3 and p=0.28, respectively) suggesting that the number of people unemployed exhibit random patterns, i.e., no spatial autocorrelation.

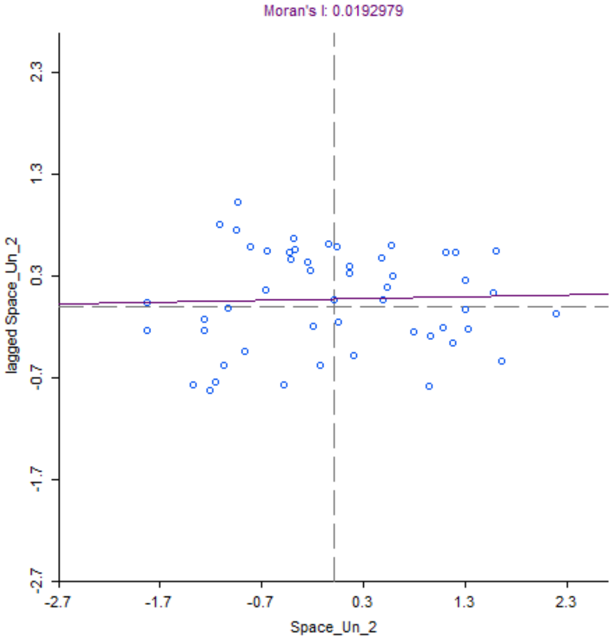
**Graphs 11: Moran's I correlation coefficient Unemployment 1996**



**Graphs 12: Moran's I correlation coefficient Unemployment 2004**



### Graphs 13: Moran's I correlation coefficient Unemployment 2013



However the underlying assumption of stationarity may be violated by the intrinsic variance instability of rates. This follows when the number of unemployed people per region varies considerable across observations. The variance instability may lead to spurious inference of Moran's I. To correct for this, GeoDa, implements the Empirical Bayes (EB) standardization suggested by Assuncao and Reis (1999).

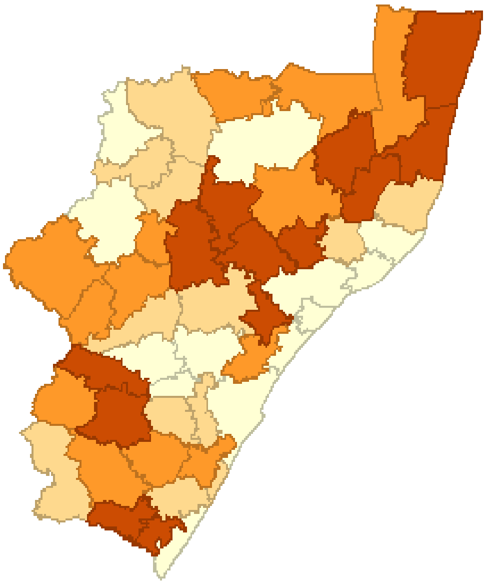
The results of the modified Moran's I correlation coefficients for the number of unemployed people per region during 1996, 2004 and 2013, using the gross domestic product (GDP Rand value, 2010 constant prices) per region for the corresponding period as the base value, are displayed in the table below.

### Table 5: Modified Moran's I Correlation Coefficients

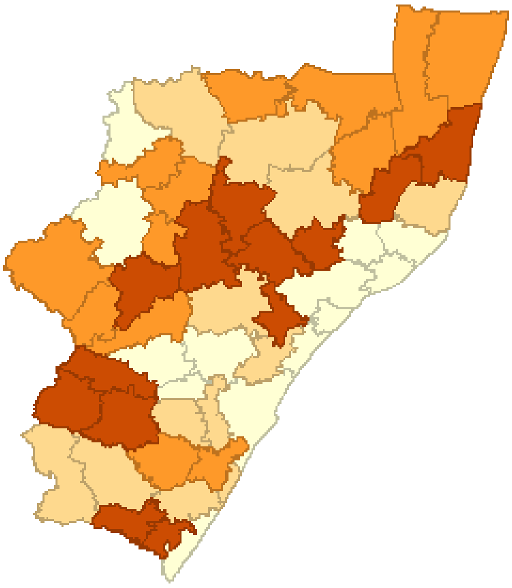
| | Moran's I | P value | |
|---|---|---|---|
| Unemployment 1996 | 0.19 | 0.02 | Statistical Significant at 5% |
| Unemployment 2004 | 0.18 | 0.02 | Statistical Significant at 5% |
| Unemployment 2013 | 0.16 | 0.03 | Statistical Significant at 5% |

The results now suggest that nearby or neighboring areas are indeed more alike, i.e., the number of unemployed people exhibit positive spatial autocorrelation. The below modified graphs (using GDP as the base value) seem to support the view of positive spatial autocorrelation.

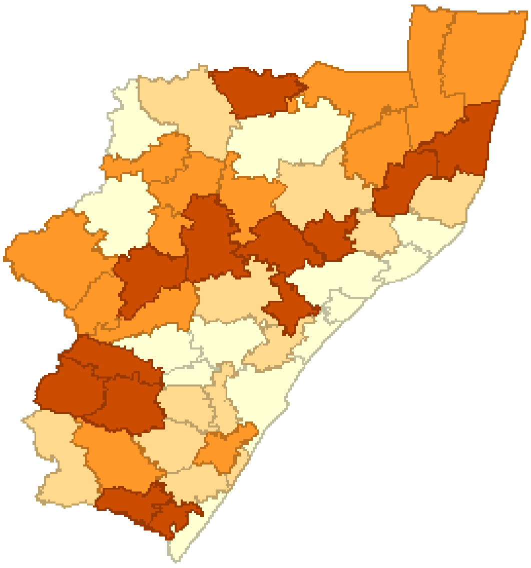
**Graph 14: Modified Unemployment 1996**



**Graph 15: Modified Unemployment 2004**



**Graph 16: Modified Unemployment 2013**



(Source, Statistics SA, own calculations)

## Spatial Regression Models

Spatial regression methods allow us to account for dependence between observations, which often arises when observations are collected from points or regions located in space. LeSage (2008) argues that it is commonly observed that sample data collected for regions or points in space are not independent, but rather spatially dependent, which means that observations from one location tend to exhibit values similar to those from nearby locations.

There are a number of theoretical motivations for the observed dependence between nearby observations. For example, Ertur and Koch (2007) use a theoretical model that posits physical and human capital externalities as well as technological interdependence between regions. They show that this leads to a reduced form growth regression that should include an average of growth rates from neighboring regions. In time series, time dependence is often justified by theoretical models that include costly adjustment or other behavioral frictions which give rise quite naturally to time lags of the dependent variable. The theoretical work of Ertur and Koch (2007) is similar in spirit, using the notion of “spatial diffusion with friction” to provide a motivation for a spatial lag, which takes the form of an average of neighboring regions (LeSage, 2008).

Spatial econometrics is a field whose analytical techniques are designed to incorporate dependence among observations (regions or points in space) that are in close geographical proximity. Extending the standard linear regression model, spatial methods identify cohorts of “nearest neighbors” and allow for dependence between these regions/observations (Anselin, 1988 and LeSage, 2005).

The most general statement of a spatial autoregressive model can be show as follows:

$$\gamma = \rho W1\gamma + X\beta + \mu$$

$$\mu = \lambda W2\mu + \varepsilon$$

$$\varepsilon \sim N(0, \sigma 2, In)$$

Where y contains an nx1 vector of cross-sectional dependent variables and X represents an nxk matrix of explanatory variables. $W_1$ and $W_2$ are known nxn spatial weight matrices, usually containing contiguity relations or functions of distance.

From the general model it is possible to derive special models by imposing restrictions. For example, setting X = 0 and $W_2 = 0$ produces a first-order spatial autoregressive model shown as follows:

$$\gamma = \rho W1 + \varepsilon$$

This model attempts to explain variation in y as a linear combination of contiguous or neighboring units with no other explanatory variables. It represents a spatial analogy to the first order autoregressive model from time series analysis, $\gamma t = \rho \gamma t - 1 + \varepsilon t$, where total reliance is on past period observations to explain variation in $y_t$.

Setting $W_2 = 0$ produces a mixed regressive-spatial autoregressive model shown below. This model is analogous to the lagged dependent variable model in time series. Here we have additional explanatory variables in the matrix X to explain variation in y over the spatial sample of observations.

$$\gamma = \rho W1\gamma + X\beta + \varepsilon$$

Letting $W_1 = 0$ results in a regression model with spatial autocorrelation in the disturbances shown as follows:

$$\gamma = X\beta + \mu$$

$$\mu = \lambda W2\mu + \varepsilon$$

A related model known as the spatial Durbin model is shown below where a “spatial lag” of the dependent variable as well as a spatial lag of the explanatory variables matrix X are added to a traditional least-squares model.

$$\gamma = \rho W_1 \gamma + X\beta_1 + W_1 X\beta_2 + \varepsilon$$

The *n* by 1 vector *y* contains our dependent variable and ρ is a scalar parameter, with *W* representing an *n* by *n* spatial weight matrix.

$$\rho W\gamma$$

An *n* by *n* binary indicator matrix *P* displayed below is derived, where the rows of the matrix correspond to observations/regions 1 to 5. Values of 1 are used in each column to indicate « neighboring » observations associated with each row. For example, *P*(1,2) = 1 and *P*(1,3) = 1 indicates that the second and third observations/regions represent the two nearest (measured using distance from the center of each region). This reflects that regions #2 and #3 are the nearest neighbors to region #1, which meets our definition of *m* = 2 neighbors. Similarly, in row 2 we have *P*(2,1) = 1 and *P*(2,3) = 1, indicating that regions 1 and 3 are the *m* = 2 nearest neighbors to region #2. Similarly, region #5 is a neighbor to regions #3 and #4.

$$P = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

The main diagonal elements of *P* are zero to prevent an observation from being defined as a neighbor to itself. It is possible to normalize the matrix *P* to have row-sums of unity by dividing all elements of the matrix *P* by the number of neighbors, in this case *m* = 2. This leads to a matrix label *W*. This *rowstochastic* form of the spatial weight matrix will be useful for expressing the spatial regression model.

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

Consider the product of the matrix *W* and a vector of observations *y* on any particular variable for the five regions displayed below. This matrix product known as a *spatial lag* produces an *n* by 1 vector containing an average of the particular variable from the regions defined as neighbors by the matrix *P*.

$$Wy = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

$$= \begin{pmatrix} 1/2y_2 + 1/2y_3 \\ 1/2y_1 + 1/2y_3 \\ 1/2y_2 + 1/2y_4 \\ 1/2y_3 + 1/2y_5 \\ 1/2y_3 + 1/2y_4 \end{pmatrix}$$

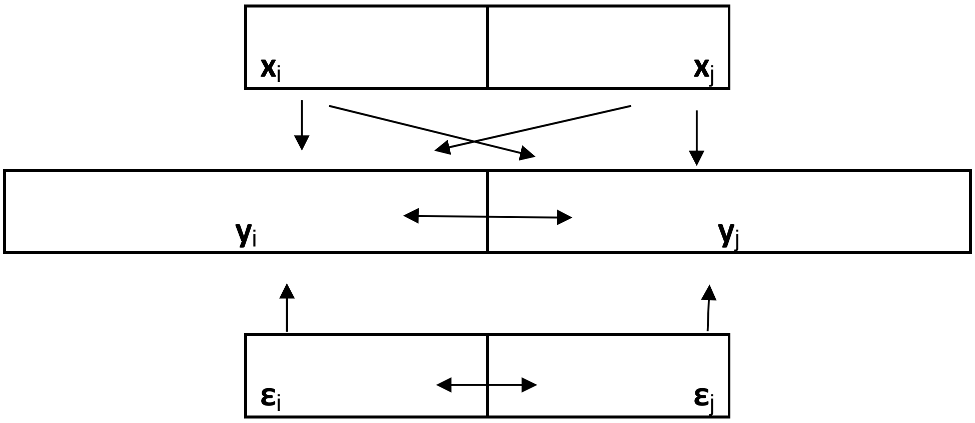
(LeSage, 2008)

It therefore seems that regions/space might be related with their neighbors in three different ways:

- The value of y in a region might impact (or be related to) the value of y in a neighboring region
- The value of X's in a region might impact (or be related to) the value of y in a neighboring region
- The residuals ε might impact (or be related to) the residuals in a neighboring region (spatial heterskedasticity)

The above can be presented as follows:

**Illustration 1: Spatial Regression Models**



## Spatial Lag Regression

Ward and Gleditsch (2007) state that the spatially lagged y model is appropriate when we believe that the values of y in one unit i are directly influenced by the values of y found in i's "neighbors." This influence is above and beyond other covariates specific to I, if we believe that y is not influenced directly by the value of y as such among neighbors, but rather that there is some spatially clustered feature that influences the value of y for i and its neighbors but is omitted from the specification. For the spatially lagged y model to be appropriate, the dependent variable y must be considered as a continuous variable.

The spatial lag regression is defined as follows:

$$y_i = pW_iy_i + x_i\beta_1 + e_i$$

The spatial lag model reduced form equation is:

$$(\lfloor - pW_i)\ y_i = x_i\beta_1 + e_i$$

The independent variables are explaining the variation in the dependent variable that is not explained by the "neighbors" values. The spatial dependence parameter p (rou) is also estimated, where a positive value for the parameter associated with the spatial lag (p) indicates that countries are expected to have higher y values if, on average, their neighbors have high y values.

## Spatial Lagged X Regression

This model is appropriate when variables' (or region's) behaviour or outcome reacts to the exogenous observable characteristics of neighbours such that;

$$y_i = x_i\beta_1 + W_iX\theta + e_i$$

$\theta$ = spatially weighted independent variables of the neighbors (theta)

This model includes the spatially lagged independent variables.

## Spatial Error Regression

In this model spatial dependence enters through the errors rather than through the systematic component of the model, i.e.,

$$y_i = x_i\beta_1 + e_i$$

$$e_i = \lambda W e_i + \varepsilon_i$$

$\lambda$ = spatially weighted errors of the neighbors (lambda)

In the spatial error model, the spatial autocorrelation term captures the spatial dependence.

## The Data

The models will make use of the following data (1996, 2004 and 2013):

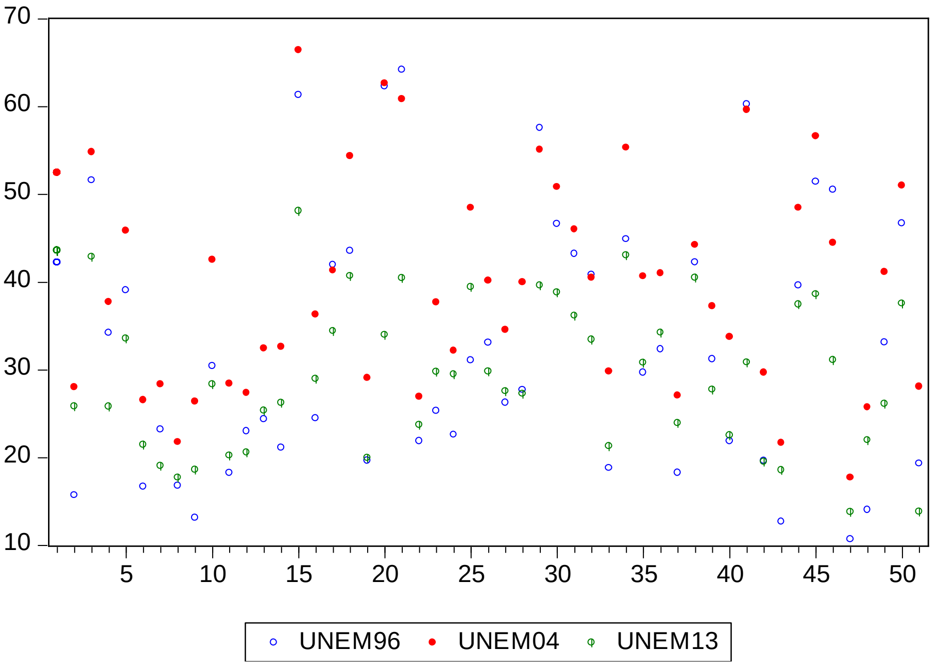
Rate of unemployment (narrow definition, percentage) per municipality as published by Global Insight and cross referenced by Statistics SA quarterly labour force survey (graph 17).

GDP (constant 2005 prices, R'000) per municipality as published by Global Insight and cross referenced by KZN Treasury provincial GDP model (graph 18).

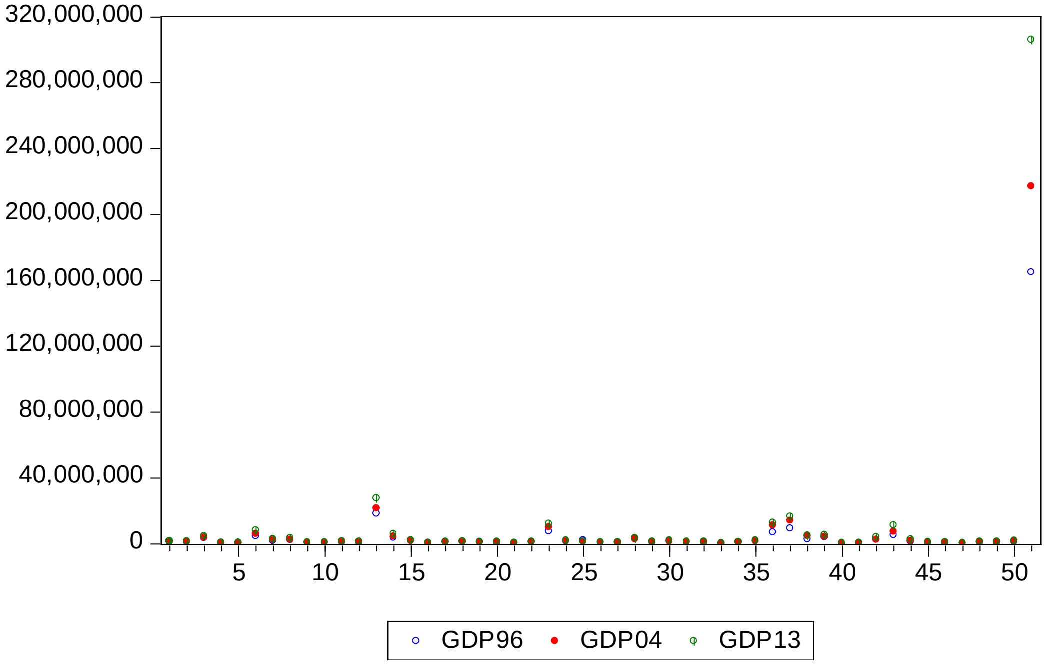
Total population (number of people) per municipality as published by Global Insight and cross referenced by Statistics SA annual population estimates (graph 19).

Literacy rate (percentage, completed grade 7 or higher) per municipality as published by Global Insight (graph 20).

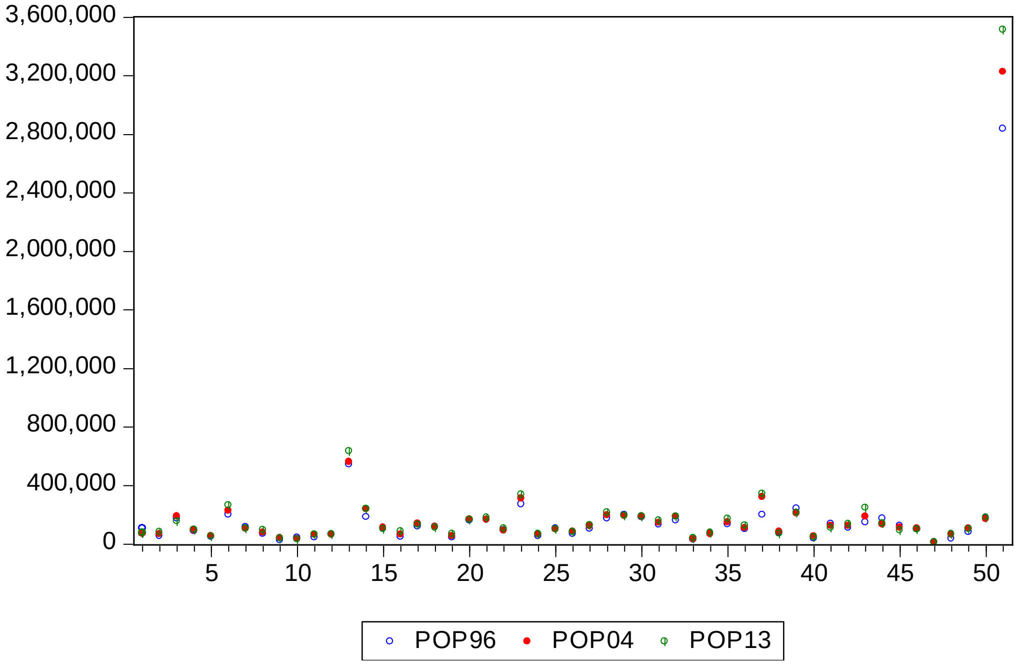
**Graph 17: Regional Unemployment Rate, 1996, 2004 and 2013 (%)**



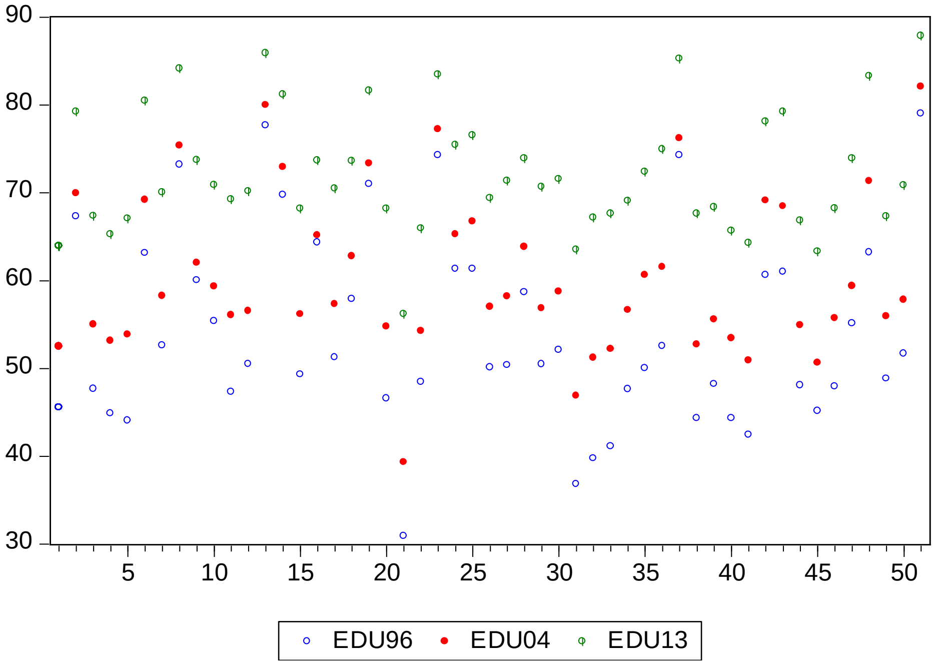
**Graph 18: Regional GDP, 1996, 2004 and 2013 (R'000)**



**Graph 19: Regional Total Population, 1996, 2004 and 2013 (number)**

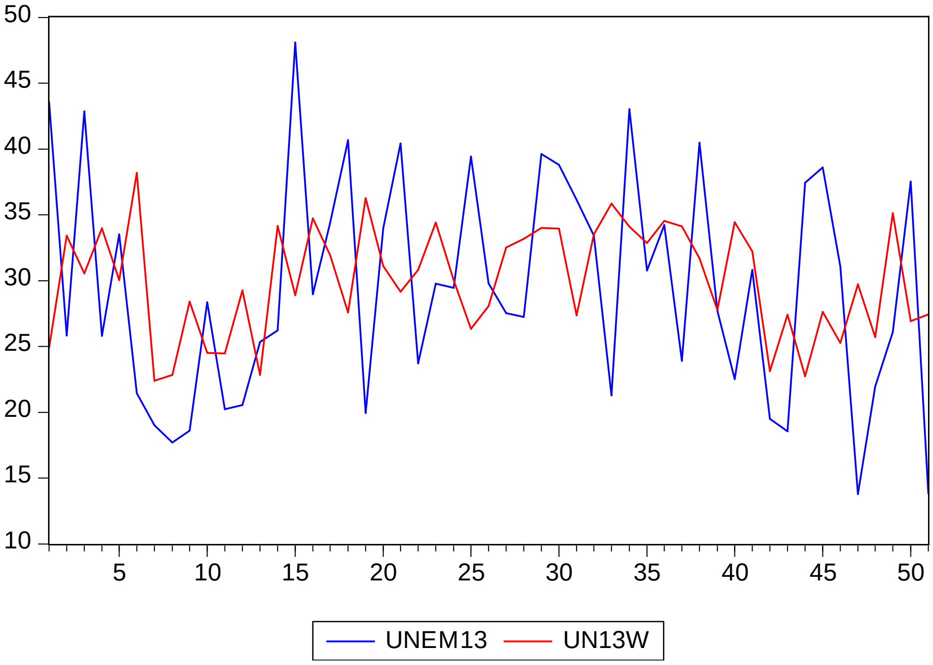


**Graph 20: Regional Literacy Rate, 1996, 2004 and 2013 (%)**



Using GeoDa it is possible to calculate the spatial lag for each of the variables (ρW). By means of an example, graph 21 displays the unemployment rate (unem) per region for 2013 as well as the spatial lag unemployment rate (unw) per region for 2013.

**Graph 21: Spatial Lag Unemployment Rate vs Unemployment rate per region, 2013 (%)**



The graph above and the summary descriptive statistics below suggest that the spatial lag operator is a much more smoothed time series than the actual time series. This holds true for all of the other variables as well.

**Table 6: Summary Descriptive Statistics, 2013**

| | UNEM13 | UN13W |
|---|---|---|
| Mean | 29.48 | 30.05 |
| Median | 28.96 | 30.04 |
| Maximum | 48.09 | 38.20 |
| Minimum | 13.80 | 22.40 |
| Std. Dev. | 8.53 | 4.16 |
| Skewness | 0.15 | -0.21 |
| Kurtosis | 2.11 | 2.00 |
| | | |
| Jarque-Bera | 1.89 | 2.52 |
| Probability | 0.39 | 0.28 |
| | | |
| Sum | 1503.59 | 1532.68 |
| Sum Sq. Dev. | 3633.79 | 866.22 |
| | | |
| Observations | 51 | 51 |

The p-values also suggest that the data of both time series is normal distributed.

## Specifying the Regression Model

The analysis will start with estimation using the classic model (default option in GeoDa), i.e., regression without spatial weights. The resulting equation is as follows:

$$\gamma t = \beta X t + \varepsilon t$$

where,

y = unemployment rate per region

X = vector of explanatory variables (GDP, population and literacy per region)

t = 1996, 2004 or 2013

ε = random error term

The results of the estimation for the three years are displayed in the table below.

**Table 7: Results of the Estimation using the Classic Model**

| Model | Classic | Classic | Classic |
|---|---|---|---|
| t | 1996 | 2004 | 2013 |
| | | | |
| Constant | 64.68967*** | 81.32516*** | 73.1142*** |
| GDP | -0.000001981223*** | -0.000001068702*** | -0.0000003998261*** |
| Population | 0.0001207952*** | 0.00007605764*** | 0.00003428265*** |
| Literacy | -0.789472*** | -0.8109855*** | -0.6500303*** |
| Adjusted R-squared | 0.54 | 0.40 | 0.29 |
| F-statistic | 20.6838*** | 11.9708*** | 7.65634*** |
| Log likelihood | -187.35 | -184.23 | -171.01 |
| Akaike info criterion | 382.706 | 376.45 | 350.015 |
| Schwarz criterion | 390.433 | 384.177 | 357.742 |
| Jarque-Bera | 0.7839*** | 1.1679*** | 1.9355*** |
| Breusch-Pagan test | 3.27 | 2.86 | 2.99 |
| Koenker-Bassett test | 4.11 | 3.39 | 4.04 |
| White test | 6.67 | 7.03 | 14.88 |

**** = significant at the 1 percent level

The Jarque-Bera test on normality of the errors is distributed as a $\chi 2$ statistic with 2 degrees of freedom. In all three cases, there is strong suggestion of normality of the errors (p>005). Both the Breusch-Pagan and Koenker-Bassett tests are implemented as tests on random coefficients, which assumes a specific functional form for the heteroskedasticity. The Koenker-Bassett test is essentially the same as the Breusch-Pagan test, except that the residuals are studentized, i.e., they are made robust to non-normality. Both test statistics indicate no problems with heteroskedasticity in each of the trend surface specifications (p>0.05).

The White test is a so-called specification-robust test for heteroskedasticity, in that it does not assume a specific functional form for the heteroskedasticity. Instead, it approximates a large range of possibilities by all square powers and cross-products of the explanatory variables in the model. The White test in this case supports the above two test statistics supports the evidence of no-heteroskedasticity.

In general the classic model seems to perform very well given that the coefficients of the X-variables are statistically significant individually and jointly, etc. However nothing has been said about a possible special relationship if any at all.

Estimating the classical model but including a spatial weight matrix (rook continuity method) yields the same results as above, but also includes the diagnostics tests against possible spatial autocorrelation as displayed in the table below.

**Table 8: Diagnostics for Spatial Autocorrelation (p values)**

| Model | Classic | Classic | Classic |
|---|---|---|---|
| t | 1996 | 2004 | 2013 |
| | | | |
| Moran's I (error) | 0.17 | 0.39 | 0.22 |
| Lagrange Multiplier (lag) | 0.74 | 0.72 | 0.92 |
| Robust LM (lag) | 0.06 | 0.09 | 0.06 |
| Lagrange Multiplier (error) | 0.29 | 0.57 | 0.35 |
| Robust LM (error) | 0.04*** | 0.08 | 0.04*** |
| Lagrange Multiplier (SARMA) | 0.10 | 0.20 | 0.11 |

**** = significant at the 1 percent level

Five Lagrange Multiplier test statistics are reported in the diagnostic output. The first two (LM-Lag and Robust LM-Lag) pertain to the spatial lag model as the alternative. The next two (LM-Error and Robust LM-Error) refer to the spatial error model as the alternative. The last test, LM-SARMA, relates to the higher order alternative of a model with both spatial lag and spatial error terms. This test is only included for the sake of completeness, since it is not that useful in practice. More specifically, in addition to detecting the higher order alternative for which it is designed, the test also has high power against the one-directional alternatives. In other words, it will tend to be significant when either the error or the lag model are the proper alternatives, but not necessarily the higher order alternative.

All one-directional test statistics are distributed as $\chi 2$ with one degree of freedom (the LM-SARMA test statistics has two degrees of freedom). To guide the specification search, the test statistics should be considered in a given sequence. The important issue to remember is to only consider the Robust versions of the statistics when the standard versions (LM-Lag or LM-Error) are significant. When they are not, the properties of the robust versions may no longer hold. The majority of the above test statistics appear not to be statistically significant suggestion no spatial autocorrelation.

## Estimating the Spatial Lag Model

Anselin (2005) states that while it is tempting to focus on traditional measures, such as the R2, this is not appropriate in a spatial regression model. The value listed in the spatial lag output is not a real R2, but a so-called pseudo-R2, which is not directly comparable with the measure given for OLS results. The proper measures of fit are the Log-Likelihood, AIC and SC.

**Table 9: Results of the Estimation using the Spatial Lag Model**

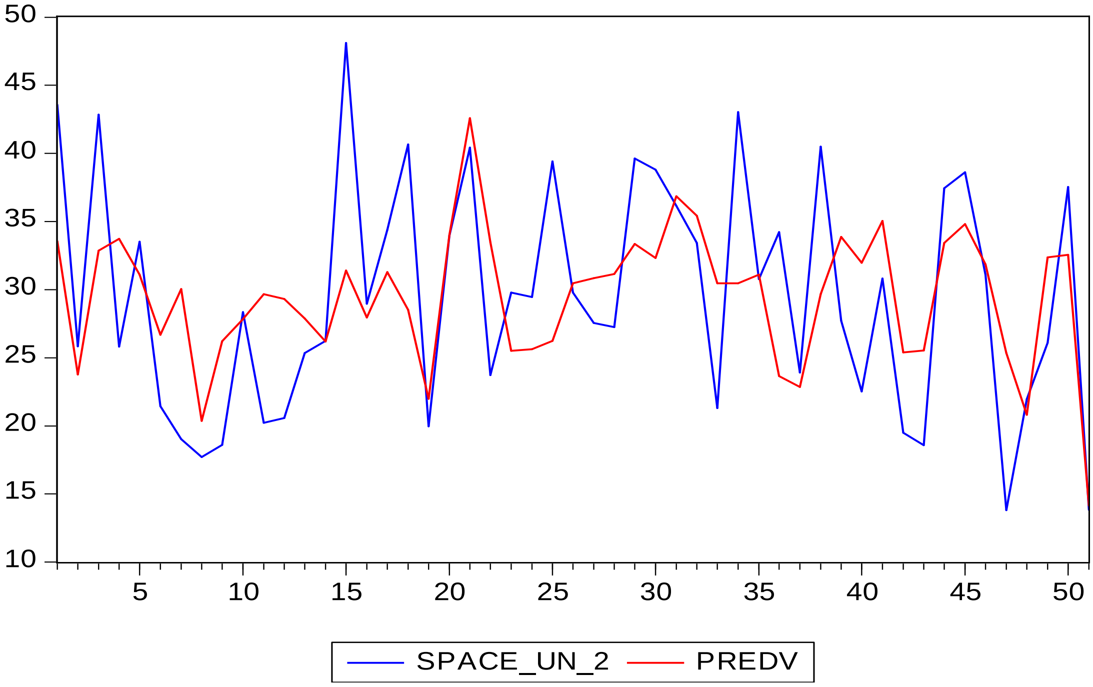
| Model | Spatial Lag | Spatial Lag | Spatial Lag |
|---|---|---|---|
| t | 1996 | 2004 | 2013 |
| | | | |
| Constant | 67.14598*** | 84.40864*** | 72.56218*** |
| GDP | -0.00000198685*** | -0.0000008064*** | -0.0000003987174*** |
| Population | 0.000120985*** | 0.00007676827*** | 0.00003420169*** |
| Literacy | -0.795929*** | -0.8155645*** | -0.6503559 *** |
| Weight Matrix (p) | -0.06284855 | -0.07077141 | 0.01936055 |
| R-squared | 0.57 | 0.44 | 0.33 |
| Log likelihood | -187.29 | -184.16 | -171.00 |
| Akaike info criterion | 384.577 | 378.31 | 352.004 |
| Schwarz criterion | 394.236 | 387.969 | 361.664 |
| Breusch-Pagan test | 3.29 | 2.81 | 3.05 |
| Likelihood Ratio Test | 0.13 | 0.14 | 0.01 |

Comparing the Log likelihood statistics between the classical model (-187 in 1996, -184 in 2004 and-171 in 2013) and the spatial lag model (-187 in 1996, -184 in 2004 and -171 in 2013) (table 7 and to table 9) indicates that including a spatial lag specification did not improve or worsen the fit. On the other hand the Akaike info criterion (383 in 1996, 376 in 2004 and 350 in 2013) using the classical model compared to (385 in 1996, 378 in 2004 and 352 in 2013) using the spatial lag model indicates some worsening of fit for the spatial lag specification.

The spatial autoregressive coefficient is estimated as -0.06 in 1996, -0.07 in 2004 and 0.02 in 2013 and is highly insignificant. The Breusch-Pagan test for heteroskedasticity in the error terms (p values are > 0.05) suggests that heteroskedasticity is not a serious problem. The second test is an alternative to the asymptotic significance test on the spatial autoregressive coefficient, it is not a test on remaining spatial autocorrelation. The Likelihood Ratio Test is one of the three classic specification tests comparing the null model (the classic regression specification) to the alternative spatial lag model. The values of (p values are > 0.05) confirms the weak significance of the spatial autoregressive coefficient.

The below graph (actual vs predicted) also suggest that including the spatial lag specification does not create a good fit.

**Graph 22: Actual vs Spatial Lag Modelled Regional Unemployment Rate (%, 2013)**



## Estimating the Spatial Errol Model

Anselin (2005) states that while it is tempting to focus on traditional measures, such as the R2, this is not appropriate in a spatial regression model. The value listed in the spatial lag output is not a real R2, but a so-called pseudo-R2, which is not directly comparable with the measure given for OLS results. The proper measures of fit are the Log-Likelihood, AIC and SC.

Comparing the Log likelihood statistics between the classical model (-187 in 1996, -184 in 2004 and-171 in 2013) and the spatial lag model (-186 in 1996, -184 in 2004 and -170 in 2013) (table 7 and to table 9) indicates that including a spatial error specification did neither improve nor worsen the fit. On the other hand the Akaike info criterion (383 in 1996, 376 in 2004 and 350 in 2013) using the classical model compared to (381 in 1996, 376 in 2004 and 349 in 2013) using the spatial error model indicates a very slight improvement of fit for the spatial lag specification.

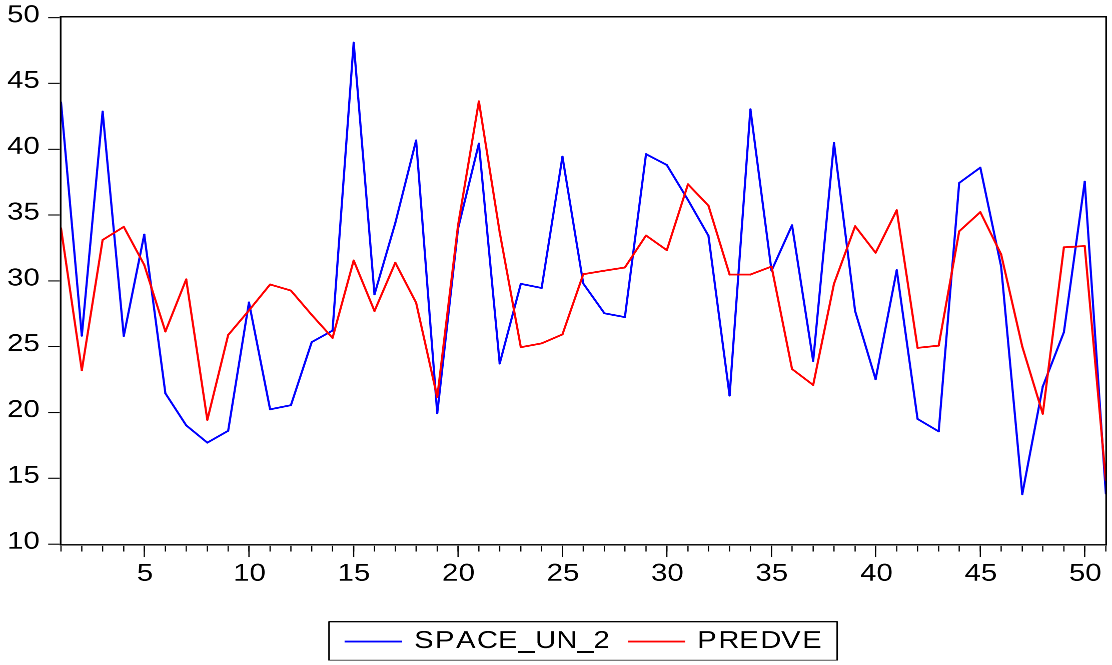
The spatial autoregressive coefficient is estimated as 0.26 in 1996, 0.14 in 2004 and 0.21 in 2013 and is highly insignificant. The Breusch-Pagan test for heteroskedasticity in the error terms (p values are > 0.05) suggests that heteroskedasticity is not a serious problem. The second test is an alternative to the asymptotic significance test on the spatial autoregressive coefficient, it is not a test on remaining spatial autocorrelation. The Likelihood Ratio Test is one of the three classic specification tests comparing the null model (the classic regression specification) to the alternative spatial lag model. The values of (p values are > 0.05) confirms the weak significance of the spatial autoregressive coefficient.

**Table 10: Results of the Estimation using the Spatial Errol Model**

| Model | Spatial Lag | Spatial Lag | Spatial Lag |
|---|---|---|---|
| t | 1996 | 2004 | 2013 |
| | | | |
| Constant | 69.04407*** | 83.60888*** | 77.69864 *** |
| GDP | -0.00000193119*** | -0.00000062026*** | -0.0000040979884*** |
| Population | 0.0001198381*** | 0.0007622609*** | 0.00003574229*** |
| Literacy | -0.8727519*** | -0.8505819*** | -0.7171712*** |
| LAMBDA | 0.2642504 | 0.1369931 | 0.2068666 |
| R-squared | 0.59 | 0.439559 | 0.346624 |
| Log likelihood | -186.65 | -184.036775 | -170.539205 |
| Akaike info criterion | 381.311 | 376.074 | 349.078 |
| Schwarz criterion | 389.038 | 383.801 | 356.806 |
| Breusch-Pagan test | 2.93 | 3.09 | 3.52 |
| Likelihood Ratio Test | 1.39 | 0.38 | 0.94 |

The below graph (actual vs predicted) also suggest that including the spatial lag specification does not create a good fit.

**Graph 23: Actual vs Spatial Error Modelled Regional Unemployment Rate (%, 2013)**



## Summary and Conclusions

Social science regression models commonly applied to cross-section and panel data assume observations on decision-making units are independent of one another. This assumption is important to contemplate since violation results in regression estimates that are biased and inconsistent.

Spatial econometrics is a field whose analytical techniques are designed to incorporate dependence among observations (regions or points in space) that are in close geographical proximity. Extending the standard linear regression model, spatial methods identify cohorts of nearest neighbors and allow for dependence between these regions/observations.

The results suggest that the province has experienced very little (if any at all) spatial dependence from 1996 to 2013. The results strongly suggest spatial heterogeneity.

## REFERENCES

Anselin, L., (2005) "*Exploring Spatial Data with GeoDa: A Workbook"* http://sal.agecon.uiuc.edu/ Copyright c 2004-2005 Luc Anselin

Anselin, L., (2003) *"An Introduction to Spatial Autocorrelation Analysis with GeoDa"* http://sal.agecon.uiuc.edu/ Copyright c 2004-2005 Luc Anselin

Chen, Y., (2011) *"On the Four Types of Weight Functions for Spatial Contiguity Matrix"* http://arxiv.org/ftp/arxiv/papers/1106/1106.0824.pdf Accessed 2 February 2015

Katchova, A., (2013) *"Spatial Econometrics Example"* Accessed 2 February 2015

Katchova, A., (2013) *"Spatial Econometrics"* Accessed 2 February 2015

LeSage, J.P. & Pace, R.K., (2009) "*An Introduction to Spatial Econometrics*" Chapman and Hall/CRC, ISBN 9781420064247

LeSage, J.P., (1999) "*The Theory and Practice of Spatial Econometrics*" http://www.spatial-econometrics.com/../sbook.pdf Accessed 2 February 2015

Sawada, M., (2009) "*Global Spatial Autocorrelation Indices - Moran's I, Geary's C and the General Cross-Product Statistic"* http://www.lpc.uottawa.ca/publications/moransi/moran.htm Accessed 2 February 2015

Smith, T.E., (2011) "*Spatial Weight Matrices*" http://www.seas.upenn.edu/~ese502/lab-content/extra_materials/SPATIAL%20WEIGHT%20MATRICES.pdf Accessed 2 February 2015

Stieve, T., (2012) *"Moran's I and Spatial Regression"* http://sites.tufts.edu/../Morans-I-and-Spatial-Regression.docx Accessed 18 June 2015

Tobler, W.R., (1976) "*Spatial interaction patterns*" Journal of Environmental Systems 6:271–301

Ward, D.M. & Gleditsch, K.S., (2007) "*An Introduction to Spatial Regression*" https://web.duke.edu/methods/pdfs/SRMbook.pdf Accessed 2 February 2015

Ward, M.D. & Gleditsch, K.S., (2008). *Spatial Regression Models*. Thousand Oaks, CA: Sage.

Wikipedia (2015) "*Moran's I*" http://en.wikipedia.org/wiki/Moran's_I Accessed 2 February 2015