# Regional growth in Western Europe: detecting spatial misspecification using the R environment*

Roger Bivand† Rolf Brunstad‡

16th January 2006

**Abstract**

The work discussed in Bivand and Brunstad (2003) was an attempt to throw light on apparent variability in regional convergence in relation to agriculture as a sector subject to powerful political measures, in Western Europe, 1989–1999. The present study takes up a number of points made in conclusion in that paper. Since it is possible that the non-stationarity found there is related to further missing variables, including the inadequacy of the way in which agricultural subsidies are represented, we attempt to replace the agriculture variables with better estimates of producer subsidy equivalents. It is also sensible to check that agricultural support is not masking or masked by other variables, for example human capital. The paper is also an account of the development of software contributed to the R project (R Development Core Team, 2005) as packages, in particular the **spdep** package for spatial econometrics. New functions generously contributed by researchers will be presented and compared. We find that agricultural support does impact regional economic growth after human capital is taken into consideration, and that we can show that apparent non-stationarity is alleviated by adding these variables. We further find that the moderated remaining spatial autocorrelation can best be represented by a substantive spatial lag model.

**JEL classification: C13, C80, C88, Q18, R11**

**Keywords:** regional growth, agricultural policy, spatial econometrics, open source software, statistical computing.

## 1 Introduction

A striking feature of the rich and dynamic field of study of growth models and regional convergence is the wide range of empirical methods used. The range is

*Paper presented at: Workshop on Spatial Econometrics, Kiel Institute for World Economics, Kiel, Germany, April 8–9, 2005, and Special session on: Regional Institutions and Growth; 45th Congress of the European Regional Science Association, Amsterdam, 23–27 August, 2005

†Economic Geography Section, Department of Economics, Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway; Roger.Bivand@nhh.no

‡Department of Economics, Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway; Rolf.Brunstad@nhh.no

now so wide that the field itself is splitting into clusters, within each of which debate takes place around separate approaches, even though the underlying hypotheses may be similar. We will argue here that shared analytical platforms should be of assistance in enhancing inter-cluster communication, and will exemplify this using functions written in R (R Development Core Team, 2005).

We concluded our earlier analysis of interactions between agricultural policy and regional growth in Western Europe (Bivand and Brunstad, 2003) by observing that, while we found support for the role of agricultural subsidies in accounting for variations in regional growth, it would be important to try to replace the agriculture variables with better estimates of producer subsidy equivalents. These are now available, albeit not for the same regions as in the earlier study. Our attention has also been drawn to the desirability of checking whether the interaction between agricultural transfers and regional growth is not masking or masked by other variables, such as human capital, and this will also be undertaken here. We will also examine whether the choice of neighbour representation used, and/or breaks of series in regional GVA, impact our conclusions.

Our specific concerns in Bivand and Brunstad (2003) were focused on the technical question of spatial non-stationarity, a question which remains relevant, but which has been posed in rather sharper form by McMillen (2003). He argues that spatial econometrics methods are commonly used when in fact the fitted model is misspecified. We would argue that in order to move ahead in debating such issues, we need to be able to access shared toolboxes of functions allowing adequate comparisons of methods to be made. Implementations of methods can differ in detail, and when this is the case, only open source solutions can give researchers the insight needed to understand how different methods perform (or mis-perform) in relation to data sets of interest.

Rather than work with other available software libraries using Matlab[1] or Python[2], we have chosen to work using R, not least because of the willingness of colleagues and authors of methods and/or their implementations to contribute functions to the R **spdep** package and other spatial data analysis packages[3]. Some aspects of the use of functions in the **spdep** package are covered in Bivand (2002, 2006), and details of spatial weights matrix construction are given in Bivand and Portnov (2004). These papers also cover the dynamics of software development and distribution in the R environment.

We will first set the scene, introducing the use of spatial econometrics methods in the study of regional convergence mainly by reference to published reviews. Next, McMillen's criticism will be presented, and challenges drawn from this will be related to regional convergence results. The core of the paper is composed of the fitting of a sequence of more elaborate models using a variety of methods, including least squares, spatial lag and error simultaneous autoregressive models using maximum likelihood and generalised moments, semi-parametric spatial filtering, and geographically weighted regression. Weaknesses and strengths of the chosen

[1] such as James P. LeSage's Spatial econometrics library, `http://www.spatial-econometrics.com/`

[2] such as PySAL - A Python Library for Spatial Analytical Functions, `http://sal.uiuc.edu/projects_pysal.php`

[3] details of packages are available from the R-Geo website: `http://sal.uiuc.edu/csiss/Rgeo/index.html`

methods will be presented and compared, often using standard R functions available within the chosen computing environment. Where relevant, code snippets will be presented; this paper is also written as "reproducible statistical research" (Leisch and Rossini, 2003), with text interspersed with the R code needed to reproduce the results.

# 2 Convergence and spatial econometrics

There is now a broad convergence literature, reviewed elsewhere in this special number (see, for example, Fingleton and López-Bazo, 2006), in Fingleton (2003), and in the context of inequality by Rey and Janikas (2005). We will not repeat the general thrust of arguments about convergence, moving straight to the way in which β-convergence is typically operationalised.

In empirical studies following the tradition of Barro and Sala-i-Martin (1992), β-convergence is represented in the following way:

$$\frac{1}{T}\log(\frac{y_{i,T}}{y_{i,0}}) = \alpha + \theta\log(y_{i,0}) + u_i, \quad (1)$$

where $\alpha$ and $\theta$ are coefficients and $u_i$ is an independent and identically distributed disturbance term. Given an estimate of $\theta$, the speed of convergence may be represented as: $\beta = -\log(1+T\theta)/T$, with $100\beta$ expressing this speed in percentage points (Barro and Sala-i-Martin, 1992, p. 230). A positive value of $\beta$ indicates convergence of GVA per capita across territorial units of analysis, whereas a negative value would indicate divergence. The underlying regularity in this representation is that the rate of growth $y_{i,T}/y_{i,0}$ of a regional economy $i$ in the period up to $T$ is related to its initial condition in period 0 for some measure $y_{i,0}$.

Apart from other issues, the prime focus of spatial econometrics studies has been related to the disturbance term $u_i$. In some cases, dynamics have been examined just in relationship to the rate of change itself, when the growth rates of regions have been plotted against the average growth rates of sets of neighbours, and explored without taking account of other variables reflecting differences in initial conditions. In parentheses, a difference in statistical practice between spatial econometrics and spatial epidemiology is that while the latter usually fit models using case or observation weights to capture difference in impact for regions of greatly differing size (see, for example, Waller and Gotway, 2004), such weights are almost never used with economic data. Using regional population size weights in the work presented below does change the results, often boosting residual spatial autocorrelation — very possibly spatial dependence in the values of the weights feeds through into the fit. Further consideration of the use of case weights will however be left for elaboration in future research (for another paper taking up the issue of case weights, see Petrakos, Rodríguez-Pose and Rovolis, 2005).

## 2.1 The McMillen misspecification challenge

McMillen (2003) points up the ease with which we can test simple models of spatial data for spatial autocorrelation, and interpret the rejection of the null hypothesis of no dependence as meaning that dependence is present rather than some other

misspecification. This is close in sense to the warning bells rung by Fingleton (1999) about the way in which Moran's $I$ can respond to a number of misspecifications. McMillen's paper has not yet been cited often in the Regional Science literature, but has been lauded by Schabenberger and Gotway (2005, pp. 22–23). They write: "... the impact of heterogeneous means and variances on the interpretation of Moran's $I$ is both widely ignored and completely confused throughout the literature. McMillen (2003) offers perhaps the first correct assessment of the problem (calling it misspecification)" (page 23).

McMillen notes that "tests for spatial autocorrelation also detect functional form misspecification, heteroskedasticity, and *the effects of missing variables that are correlated over space*" (our italics, pages 208–209). Consequently, when studying convergence in the form discussed here, it will be of key importance to see whether the disturbance term $u_i$ not only contains information about possible spatial dependence, but also whether additional variables do not reduce such dependence. This is relevant in this case, for which no clear behavioural model for information spillover is proposed, but where spatial dependence is often attributed to a mismatch between the functional regions constituting the economy and the administrative boundaries used for data collection and aggregation. When a behavioural model for information spillover is available, that is when the observational units are active polities and their mutual interaction can be modelled, these concerns are perhaps less pressing (see, for example, Bivand and Szymanski, 2000).

# 3 Estimating convergence

In this section, we will present the source of the EU GVA data used, and apply a number of model fitting methods and specification tests. In the course of examining the results of fitting the simple growth model, it will become apparent that there are spatial regimes in the data, that groups of regions appear to behave in different ways. In the next section, additional variables will be included in the model, and a wider range of model fitting methods will be used.

## 3.1 Data sources for EU NUTS 1/2, 1989–1999

The data used for EU NUTS2 gross value added for 1989 and 1999 were extracted from Cambridge Econometrics' European Regional Databank, and are measured in million 1990 ECU and 1000 persons[4]. The start and finish years were chosen in the earlier paper (Bivand and Brunstad, 2003) to accommodate available agricultural data, for which major breaks in series occurred when revisions of the System of National Accounts, knocked on to EAA 97, the European system of agricultural accounts. Because the data source we are now using to measure agricultural policy impacts is also available for Greece, we have extended our regional coverage here, but still do not include Denmark, Ireland, UK, Berlin and former East Germany, or overseas dependencies and Atlantic islands. Our agricultural policy data source also aggregates NUTS2 regions to NUTS1 in Belgium, Netherlands and Germany, leaving us with 93 units of observation, partly at NUTS1 and partly at NUTS2.

[4]We would like to thank Cambridge Econometrics and Sasha Thomas for their help.

Fitting the simple growth model in R is done with the `lm()` function, which takes as its first argument a `formula` object. The left and right hand sides are separated by a tilde, and the variables used here are `growth` ($\frac{1}{10}\log(\frac{y_{1999}}{y_{1989}})$), and `gva89` ($y_{1989}$), where $y$ is GVA in million EUR divided by population figures for the same year in thousands, that is GVA per capita in thousand EUR. The `conv.lm` object contains the results of fitting the formula to the data using a least squares linear model with default arguments not given explicitly; arguments do not need to be named, and can also be given by position.

```
> conv.lm <- lm(formula = growth ~ log(gva89))
```

Table 1: Results and specification tests for the simple growth model.

| | value | std. error | p-value |
|---|---|---|---|
| Constant | -0.01123 | 0.00578 | 0.05498 |
| $\theta$ | 0.00967 | 0.00238 | 0.00011 |
| % speed of convergence | -0.923 | | |
| $\sigma$ | 0.0114 | | |
| adjusted $R^2$ | 0.144 | | |
| Studentized Breusch-Pagan test | 25.0 | | 5.6e-07 |
| RESET test | 3.14 | | 0.048 |
| Moran's $I$ | 0.656 | | <1e-08 |
| LM error | 78.8 | | <1e-08 |
| LM lag | 81.7 | | <1e-08 |
| DLM error | 21.6 | | 3.4e-06 |

Table 1 presents an extract of the results for this simple model. We see that the $\theta$ coefficient is significant, and has the opposite sign to that expected, with a consequent negative percentage convergence rate. Tests for heteroskedasticity and functional form indicate problems with the specification. In order to test for spatial dependence, we need a suitable weights matrix, and choose here, as in Bivand and Brunstad (2003), to use a sphere of influence graph definition of neighbours; row standardised spatial weights are used throughout. This definition is derived from work in computational geometry described in Avis and Horton (1985); the points representing the regions are GISCO NUTS2 or NUTS1 label points projected to Lambert Azimuthal Equal Area using EU standard parameters. This representation gives two subgraphs, there being no link between the Greek regions and the others. The spatial distribution of the standard residuals for this fit are shown in Figure 1, with a perhaps unexpected marked difference in sign between regions in Greece and in Spain and Portugal. The results of Moran's $I$ and the Lagrange Multiplier tests for spatial dependence are all highly significant.

The final line in Table 1, for the differenced model LM error test, is drawn from Kosfeld and Lauridsen (2004). It is calculated from the same model after the variables have been spatially differenced, $\Delta\mathbf{y} = \Delta\mathbf{X}\beta + \mathbf{u}$, where $\Delta = (\mathbf{I} - \mathbf{W})$, and $\mathbf{W}$ is a matrix of spatial weights. Kosfeld and Lauridsen (2004, page 714) suggest that if the LM error test is positive but the differenced model LM error test is zero, the regression is nonstationary and spurious, if both are positive, we have stationary spatial autocorrelation, and if the LM error test is zero but the differenced model LM error test is positive, autocorrelation is absent. If we ignore the other serious

misspecifi cations — something we should certainly not do — then their test strategy would suggest that spatial autocorrelation is present.

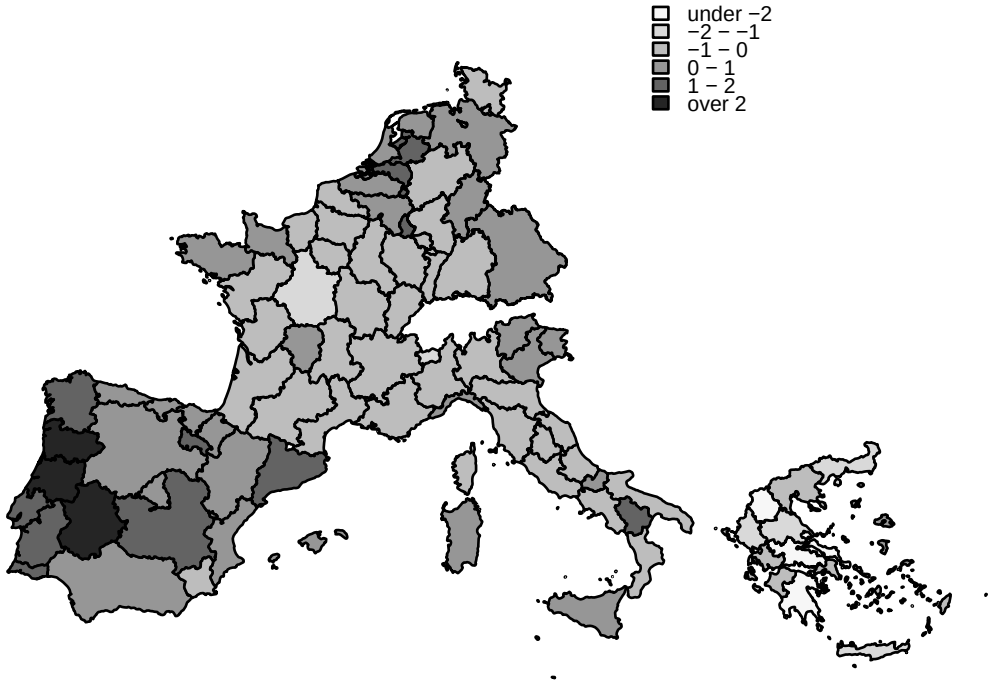


Figure 1: Simple model standardised residuals

Table 2: Summary measures for six weights matrices: number of regions, total number of links, minimum, maximum and mean number of links, and number of subgraphs, sorted by numbers of links.

| | n | links | min | max | mean | subgraphs |
|---|---|---|---|---|---|---|
| MST | 93 | 184 | 1 | 3 | 1.978 | 1 |
| Relative | 93 | 234 | 1 | 4 | 2.516 | 1 |
| Gabriel | 93 | 354 | 1 | 6 | 3.806 | 1 |
| SOI | 93 | 398 | 1 | 8 | 4.280 | 2 |
| D350 | 93 | 898 | 1 | 19 | 9.656 | 1 |
| ID350 | 93 | 898 | 1 | 19 | 9.656 | 1 |

It may be of interest to check whether the choice of defi nition of neighbours affects our results. In order to do this, a number of extra spatial weights were prepared, using the same input coordinates to represent the 93 regions. Table 2 shows summary measures for the six defi nitions compared, the sphere of influence neighbours, minimum spanning tree neighbours, Gabriel and relative neighbour graph neighbours, and distance neighbours with a threshold of 350km, all as row-standardised binary weights. Finally, the distance-based neighbours using inverse distance weights were added. The table shows the defi nitions sorted by their numbers of links, from the most to least sparse. Only the sphere of influence neighbours do not connect the regions of Greece to the rest of the map.

Table 3 shows the values of Moran's $I$ using different spatial weights for the residuals of the simple model sorted by probability value. The value for our chosen representation of neighbours, sphere of influence neighbours, appears to be very similar to the others calculated using the alternative neighbour definitions and weights. There are some differences in values and probability values, as would be expected, but the inference drawn would be the same,

Table 3: Observed values of Moran's $I$ from the residuals of the simple model for six different weights matrices (row-standardised), sorted by p-values.

| | Observed Moran's $I$ | p-value |
|---|---|---|
| D350 | 0.6216 | 2.086e-40 |
| ID350 | 0.6512 | 2.359e-38 |
| SOI | 0.6557 | 1.011e-21 |
| Gabriel | 0.6217 | 7.134e-18 |
| Relative | 0.6720 | 1.165e-13 |
| MST | 0.7115 | 2.986e-12 |

## 3.2 Spatial regimes

From our earlier work, we are aware that SW Iberia appears to behave differently in the simple model. On the basis of a scatter plot of the variables, in which the currency crises in Greece force down the EUR-denominated GVA per capita in 1999, and hence the growth rate, it seemed logical to include some special treatment for Greek regions too. Among others, Boldrin and Canova (2001) note that artefacts may be introduced into comparative series by deflators for price levels and/or exchange rates. The Cambridge Econometrics data for GVA used here are converted to ECU/Euro using tables from *International Financial Statistics Yearbook* published by the IMF (Sasha Thomas, personal communication, 2002). So regions in countries with unusual exchange rate series will appear to perform differently from regions in countries with typical exchange rate series. The left frame of Figure 2 shows exchange rate series deflated to purchasing power from the data set used by Boldrin and Canova (2001)[5], which show why the negative performance of Greek regions is an artefact of the conversion of Drachma GVA values to ECU/Euro.

Other issues connected to the very aggregated and manipulated nature of NUTS2 GVA values will be addressed below, but to examine the impact of different exchange rate impacts here, a factor `fGRSWIBa` was constructed with three values: `c("Other", "SW Iberia", "Greece")`, and a nested interaction model was estimated. The formula now expands into the main effects for the spatial regime factor, and then the initial condition variable `log(gva89)` for each level of the spatial regime factor, omitting the global constant (Chambers and Hastie, 1992, page 27). An analysis of variance between the simple model and this nested interaction model is the Chow test for the simple model for the chosen spatial regimes[6].

```
> conv1.lm <- lm(formula = growth ~ fGRSWIBa/(log(gva89)) - 1)
```

[5] `http://www.econ.umn.edu/~mboldrin/Papers/regio-eu.zip`

[6] Thanks to Achim Zeilis for clarifying the use of formula syntax for this case

We can see from the result of the Chow test: 35.9, probability value <1e-08, that taking account of spatial regimes is clearly justified. In addition, the slope signs are all negative as expected, with relatively high percentage convergence rates for two of the distinguished spatial regimes. Frame b) of Figure 2 shows the regression line fitted for all regions, and for each of the three spatial regimes in turn.

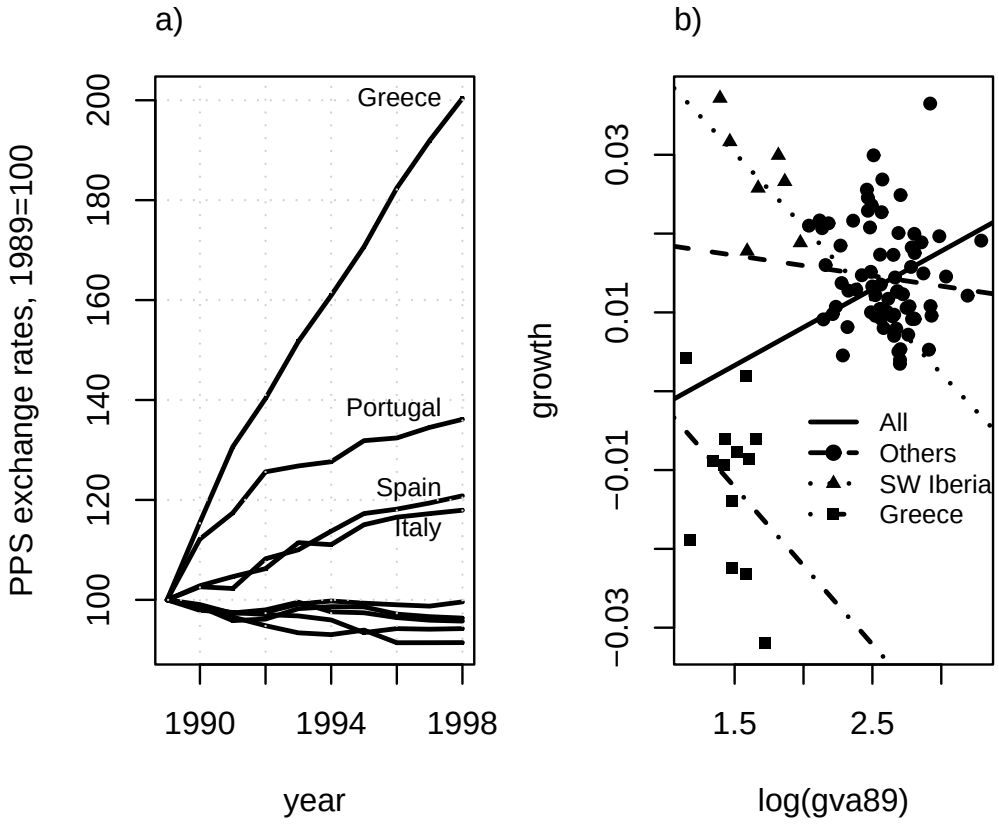


Figure 2: a) Purchasing power standard exchange rates to ECU/Euro, 1989–1998; b) Growth model fits with three spatial regimes, SW Iberia, Greece, and other regions.

Carrying nested interactions, implying separate convergence models for each of the spatial regimes, forward into the next section, in which additional variables will be considered to help account for variance in growth rates, will be cumbersome, and a potential simplification seems worth trying. Since we can see from the right frame of Figure 2 that the slopes of the regression lines for Greek regions and regions in SW Iberia are similar, and we know from Bivand and Brunstad (2003) that added variables account for growth differences between SW Iberia and the remaining regions, we will here try just using a factor (dummy) for the Greek regions, shifting the intercept for those regions to take account of the currency crises and shift in the Drachma/EUR exchange rate during the study period. The model is them:

```
> convGR.lm <- lm(growth ~ log(gva89) + fGR)
```

An analysis of variance between the nested interaction model and the simplified model with just a factor for Greek regions gives an $F$ value of 2.40, which with the appropriate degrees of freedom has a probability value of 0.074. Since the simplified spatial regimes model gives only a slightly worse fit, we will proceed

using a factor for Greek regions in the following section. The full results for this model are shown in Table 4, and now have a θ estimate with the expected sign, and which is clearly significant, and a positive percentage convergence rate. The model now accounts for roughly two-thirds of the variance in the growth variable, and the significance of all the presented specification tests has fallen substantially. The differenced model LM error test is not calculated, because the spatial difference of the Greek regions factor and its spatial lag is zero, this being caused by the fact that no Greek region has a neighbour outside Greece in the neighbourhood definition used here.

Table 4: Results and specification tests for the simplified spatial regimes model.

| | value | std. error | p-value |
|---|---|---|---|
| Constant | 0.03862 | 0.00573 | < 1e-08 |
| θ | -0.00920 | 0.00225 | 9.7e-05 |
| Greece factor | -0.03667 | 0.00321 | < 1e-08 |
| % speed of convergence | 0.965 | | |
| σ | 0.0073 | | |
| adjusted $R^2$ | 0.646 | | |
| Studentized Breusch-Pagan test | 3.82 | | 0.15 |
| RESET test | 4.32 | | 0.016 |
| Moran's *I* | 0.227 | | 0.00012 |
| LM error | 9.41 | | 0.0022 |
| LM lag | 10.3 | | 0.0013 |

# 4 Estimating convergence with additional variables

In Bivand and Brunstad (2003), our immediate concern was to relate agricultural variables to change, after taking initial GVA per capita into account. Having further reviewed the recent literature on regional convergence, we are aware that our tentative conclusions are not only subject to questioning with regard to the operationalisation of agricultural policy impacts, but also to the possibility that other variables with a similar spatial pattern may be at least as important as agricultural policy support in helping explain regional growth rates. We will take up one such variable, human capital, first, before going on to revise our operationalisation of agricultural policy impacts.

In terms of our initial, unconditional, convergence model (equation 1), positive values of β would indicate so-called absolute β-convergence. Allowing for differences in steady state across regions, conditional β-convergence can be represented in the following way (see for example Fingleton and López-Bazo, 2006, in this volume, equation 1):

$$\frac{1}{T}\log(\frac{y_{i,T}}{y_{i,0}}) = \alpha + \theta\log(y_{i,0}) + \delta'\mathbf{x}_i + u_i, \quad (2)$$

where δ is a vector of coefficients and $\mathbf{x}_i$ is a vector of variables controlling for differences across regions. In our case, these variables will include the dummy for Greek regions (shifting the intercept), a human capital variable, and an agricultural support variable.

## 4.1 Human capital 1996

We have operationalised human capital by using Eurostat tables for the percentage of the active population by NUTS2 region constituting human resources in science and technology (HRST) in 1996[7]. We are aware that 1996 is very late in our period, but earlier data with sufficient regional coverage have not been identified. The HRST definition is "persons who successfully completed education at the third level in a S&T field of study, or not formally qualified as above but employed in a S&T occupation where the above qualifications are normally required'[8]. We would expect regions with relatively higher levels of human capital defined in this way to experience higher rates of growth. The human capital variable `hc96` varies considerably between regions; in order to reduce numerical problems, the variable has been rescaled to thousands of HRST persons per 10,000 active.

## 4.2 Net agricultural support 1996

Although not ideal, some estimates of producer subsidy equivalents for agricultural policy support have been made available in a report prepared for the European Commission prior to the completion of the second cohesion report[9]. Although the report mentions numerous difficulties in collating a usable data set, including non-reporting by many countries, and non-reporting at the NUTS2 level by others, which is why we have been forced to aggregate NUTS2 regions to NUTS1 in Belgium, Netherlands, and Germany, we have chosen to use the data denoted Finpc1996 in the report, here `nett.pc.1996`. These are among the more complete, and give for 1996 a value per capita net transfer value imposed by the Common Agricultural Policy on the regions. The value is the result of subtracting market transfers from taxpayers and market transfers from consumers, from market transfers to producers, which constitutes the gross transfer value. As with the human capital variable, we are assuming that the 1996 values are highly correlated with the start of period values we would have preferred to use.

As we argued in our earlier work, we expect the level of agricultural policy support to be negatively related to regional growth, because higher levels of support are likely to slow the reallocation of labour and capital to non-agricultural sectors. We feel that using per capita values is not unreasonable in the present context, but continue to be aware that changes in CAP during the study period, and the fact that support levels for Greece, Spain and Portugal ramped up during our chosen period, may weaken any substantive conclusions we may reach. Our least squares model, including both human capital and agricultural support, is now:

```
> convhc2.lm <- lm(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR)
```

[7] The 1998 value for Portugal has been used for all Portuguese regions for 1996; Italian regions were matched to their earlier NUTS2 nomenclature.

[8] Science and technology in Europe Statistical pocketbook, Eurostat, 2005, page 144.

[9] The report is available from Inforegio: `http://europa.eu.int/comm/regional_policy/sources/docgener/studies/pac_en.htm`, and the calculations made are described in Chapter 5, methodology in Section 5.1, and tables in appendices to that chapter. The data coverage, both by region and by year, is limited

Table 5: Results and specification tests for the model including human capital and agricultural policy support variables.

| | value | std. error | p-value |
|---|---|---|---|
| Constant | 0.043655 | 0.005839 | <1e-08 |
| θ | -0.015000 | 0.002532 | 6e-08 |
| hc96 | 0.003156 | 0.001092 | 0.0048 |
| nett.pc.1996 | -0.001231 | 0.000655 | 0.0634 |
| Greece factor | -0.036537 | 0.003028 | <1e-08 |
| % speed of convergence | 1.63 | | |
| σ | 0.00677 | | |
| adjusted $R^2$ | 0.696 | | |
| Studentized Breusch-Pagan test | 3.46 | | 0.48 |
| RESET test | 1.36 | | 0.24 |
| Moran's $I$ | 0.108 | | 0.017 |
| LM error | 2.15 | | 0.14 |
| LM lag | 4.16 | | 0.041 |

Table 5 demonstrates clearly that adding a factor for Greek regions, and variables expressing regional differences in human capital and agricultural support, has alleviated most of the specification problems. The agricultural support variable has the expected sign, but is not highly significant. Although it is a borderline case, we feel that it does add extra information to account for differences in regional growth. Regions with lower net transfers to agriculture experience slightly faster growth than regions with larger net transfers, even when differences in human capital are taken into account.

Since we can assume that we have now reached a moderately acceptable model with regard to variables related to differences between regions, we can return to the question of the extent to which conclusions drawn may be influenced by breaks in series and other measurement shifts in the per capita GVA series for NUTS2 regions. If we replace 1999 as year $T$ by the years 1994–1998, we can see from Table 6 that there are some changes in coefficient values.

Table 6: Coefficient estimates and convergence estimates for years $T$ from 1994 to 1999

| | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 |
|---|---|---|---|---|---|---|
| (Intercept) | 0.03205 | 0.03438 | 0.03644 | 0.04066 | 0.04267 | 0.04366 |
| log(gva89) | -0.01100 | -0.01089 | -0.01339 | -0.01467 | -0.01527 | -0.01500 |
| hc96 | 0.00267 | 0.00229 | 0.00317 | 0.00323 | 0.00351 | 0.00316 |
| nett.pc.1996 | -0.00201 | -0.00197 | -0.00186 | -0.00177 | -0.00147 | -0.00123 |
| fGRTRUE | -0.02928 | -0.02726 | -0.02361 | -0.02159 | -0.03290 | -0.03654 |
| β | 0.01131 | 0.01127 | 0.01406 | 0.01560 | 0.01643 | 0.01625 |
| CI 2.5% β | 0.01874 | 0.01793 | 0.02095 | 0.02213 | 0.02263 | 0.02235 |
| CI 97.5% β | 0.00414 | 0.00486 | 0.00748 | 0.00940 | 0.01055 | 0.01050 |

If we look at our convergence estimates, both point estimates and similarly transformed 95% confidence interval estimates for the years 1994-1999 as shown in Table 6, there are no cases in which the point estimates of β fall outside the confidence intervals of the other years. On the other hand, the point estimates do differ,

with 1994 and 1995 appearing to belong to a different period than the later years. Tightening the confidence intervals for 1994 and 1995 to 85%, the $\beta$ values for 1998 and 1999 become borderline, but no clear evidence of major impacts of breaks of series is found.

Table 7: Observed values of Moran's $I$ from the residuals of the model including human capital and agricultural policy support variables for six different weights matrices (row-standardised), sorted by p-values.

| | Observed Moran's $I$ | p-value |
|---|---|---|
| D350 | 0.1105 | 0.0004738 |
| ID350 | 0.1166 | 0.0007382 |
| Gabriel | 0.1459 | 0.0064393 |
| MST | 0.2152 | 0.0086831 |
| Relative | 0.1690 | 0.0140957 |
| SOI | 0.1084 | 0.0167113 |

## 4.3 Convergence with spatial dependence

The values of both Moran's $I$ and the two Lagrange Multiplier tests shown in Table 5 are much less significant than before. If the results from applying alternative neighbour definitions do not differ from our chosen spatial weights, the remaining spatial patterning of the disturbance term may be attributed to arbitrary regional boundaries, or unspecified spillover, and either ignored or removed by using alternative spatial model fitting techniques. However, before drawing such a conclusion, we check the values of Moran's $I$ shown in Table 7. It appears that our chosen neighbour definition is the most conservative, but similar to the relative graph neighbours. The distance-based neighbour definitions indicate significant residual autocorrelation, as, to a lesser extent, do the Gabriel graph and minimum spanning tree definitions. Revisiting the Kosfeld and Lauridsen differenced model test and using the Gabriel graph neighbour definition, we find that the LM error test statistic is 3.531 with probability value 0.0602, while the differenced model statistic is 14.97 with probability value 0.000109. This suggests that we could conclude that the remaining spatial dependence is not very strong.

Fingleton and López-Bazo (2006) discuss in detail how spatial effects may be understood, and restate the argument, presented above, that a spatial error formulation of spatial effects may be "a manifestation of the omission of one or more spatially autocorrelated variables" from the vector $\mathbf{x}_i$ in equation 2. They also point out that the unconstrained spatial Durbin model nests both the spatial lag and spatial error specifications of spatial econometrics models (see Anselin, 2002, for further details and a review of these methods). Rewriting equation 2 in vector form with $g_{y_i} = \frac{1}{T}\log(\frac{y_{i,T}}{y_{i,0}})$ and reorganising the right hand side as $\mathbf{z}_i = [1, y_{i,0}, \mathbf{x}_i]$, we have:

$$\mathbf{g_y} = \mathbf{Z}\zeta + \mathbf{u} \tag{3}$$

where $\zeta = [\alpha, \theta, \delta]$. Extending this to the unconstrained spatial Durbin form can be achieved by adding spatially lagged $\mathbf{g_y}$ and $\mathbf{Z}$ variables to the right hand side:

$$\mathbf{g_y} = \rho\mathbf{W}\mathbf{g_y} + \mathbf{Z}\zeta + \mathbf{W}\mathbf{Z}\gamma + \mathbf{u} \quad (4)$$

In the spatial lag model, $\gamma = 0$, in the spatial error model, the Common Factor constraint $\gamma = -\rho\zeta$ is satisfied. As Fingleton and López-Bazo (2006) point out, quite often studies of convergence in Europe have ended up with a spatial error specification, despite the fact that this indicates that spatial externalities are not substantive phenomena, but rather random shocks diffusing through space. Table 5 suggests that the spatial lag specification may be better suited to our data, but there are still a number of steps to be taken before we can conclude that we are looking at substantive spatial externalities, that high growth rates spill over from region to neighbouring region.

Here we will use three maximum likelihood methods, the unconstrained spatial Durbin model including the spatially lagged dependent and independent variables, spatial lag simultaneous autoregression — including the lagged dependent variable on the right hand side with coefficient $\rho$ — and spatial error simultaneous autoregression — placing the spatial dependence in the disturbance term. Functions for fitting using these methods have been present in the R **spdep** package for some time, `lagsarlm()` and `errorsarlm()`.

Since the outcome of testing for Moran's $I$ in the current model was not clear, and alternative neighbour definitions seemed to give clearer views of the possible residual autocorrelation, we first estimate spatial lag, spatial Durbin, and spatial error models for the different weights definitions, and compare the autoregressive coefficient estimates, the probability values of likelihood ratio tests against the hypothesis that $\lambda = 0$ or $\rho = 0$, and the AIC values of the fits, shown in Table 8.

Table 8: Estimated values of $\rho$, likelihood ratio test p-values and AIC values from the spatial lag model, the spatial Durbin model, and $\lambda$, likelihood ratio test p-values and AIC values from the spatial error model for six different weights matrices (row-standardised).

| | Lag | | | Durbin | | | Error | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\rho$ | p-value | AIC | $\rho$ | p-value | AIC | $\lambda$ | p-value | AIC |
| SOI | 0.235 | 0.0527 | -660.0 | 0.173 | 0.2036 | -655.4 | 0.206 | 0.1453 | -658.4 |
| Gabriel | 0.293 | 0.0111 | -662.7 | 0.236 | 0.0762 | -655.9 | 0.259 | 0.0566 | -659.9 |
| Relative | 0.240 | 0.0172 | -661.9 | 0.200 | 0.0754 | -655.5 | 0.214 | 0.0638 | -659.7 |
| MST | 0.225 | 0.0164 | -662.0 | 0.218 | 0.0327 | -655.5 | 0.222 | 0.0326 | -660.8 |
| D350 | 0.331 | 0.0326 | -660.8 | 0.368 | 0.0549 | -653.7 | 0.386 | 0.0395 | -660.5 |
| ID350 | 0.357 | 0.0192 | -661.8 | 0.334 | 0.0603 | -654.2 | 0.368 | 0.0392 | -660.5 |

All of the spatial lag models estimated with alternative neighbour definitions perform better, understood as capturing more variance associated with spatial pattern, than the sphere of influence definition, with the Gabriel graph slightly ahead of the minimum spanning tree. In the spatial Durbin and spatial error model case, again all the alternative definitions outperform the sphere of influence definition, with the minimum spanning tree definition just best. On the basis of this comparison, maintaining the choice of sphere of influence neighbours does not seem justified, and Gabriel graph neighbours are used for the spatial lag model and minimum spanning tree neighbours otherwise.

Insignificant likelihood ratio and Wald tests suggest that the spatial Durbin model with minimum spanning tree weights does not perform better than the spatial error model with minimum spanning tree weights, and that the Common Factor constraints cannot be rejected. However, none of the fitted $\gamma$ coefficients (the lagged intercept is omitted because it is collinear with the intercept for row-standardised spatial weights) are significantly different from zero, pointing us towards a spatial lag specification. Comparing the Akaike AIC values shows that fitting coefficients on lagged **Z** variables costs more in terms of lost degrees of freedom than it helps in accounting for variance in the dependent variable.

Because R is naturally extensible, it attracts collaboration between scholars, and function `GMerrorsar()` has been contributed by Luc Anselin. It uses the generalised moments approach to fitting the spatial error model proposed by Kelejian and Prucha (1999), in this version using sparse matrix code from the **SparseM** package (Koenker and Ng, 2005) to calculate the log-likelihood of the fitted model for testing against the non-spatial least squares estimates. These functions need additional arguments, in particular the `listw` argument for the spatial weights.

Since the functions use numerical optimisation, they may need extra tuning in order to get coefficient estimates closer to the optimum on very flat surfaces. The ML functions are fit by line search for the spatial coefficient, while `GMerrorsar()` is fit by numerical optimisation on a surface of $\lambda$ and $\sigma^2$ values. This surface approaches its minimum very gradually, with quite a large range of values $\lambda$ corresponding to a narrow range of $\sigma^2$. The differences between the ML and GM implementations are very small, but are influenced by the choice of tolerance for the GM estimator numerical optimiser. One could argue, on the one hand, that users of such estimators should be shielded from arcane technical issues such as these, but, on the other hand, they do make a difference in practice, and access to such control arguments is a feature of mature statistical software.

The two ML functions are also subject to scaling issues in inverting the asymptotic variance matrix — the matrix can appear singular to numerical linear algebra functions when variable scaling is very uneven. Here, the ML functions use weights matrix eigenvalues in finding the objective function optimum, and full weights matrices for calculating the asymptotic variance matrix of the coefficient estimates; they can use sparse matrix techniques if desired.

```
> convhc2.lag <- lagsarlm(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, listw = nb2listw(gab_nb))
> convhc2.err <- errorsarlm(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, listw = nb2listw(mst_nb), tol.solve = 1e-16, tol.opt = 1e-14)
> convhc2.gm <- GMerrorsar(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, listw = nb2listw(mst_nb), control = list(reltol = 1e-14))
```

Table 9 shows the results of fitting the model including human capital and agricultural policy support variables using three SAR functions (the least squares results are included in the final column for comparative purposes). The similarity of coefficient estimates is obvious, and largely confirmed by the final line, which reports likelihood ratio tests been the SAR function results — including an extra spatial coefficient — and the least squares results. Akaike's An Information Criterion repeats this, with only minor differences between the least squares and error SAR fits. The largest differences are between the lag SAR results and the least squares fit, and

Table 9: Results for the model including human capital and agricultural policy support variables fitted with three SAR functions.

| | ML Lag SAR | (std.err) | ML Error SAR | GM Error SAR | OLS |
|---|---|---|---|---|---|
| Constant | 0.03275 | (0.006673) | 0.04116 | 0.04134 | 0.04366 |
| $\theta$ | -0.01159 | (0.002604) | -0.01376 | -0.01385 | -0.01500 |
| hc96 | 0.00248 | (0.001033) | 0.00295 | 0.00297 | 0.00316 |
| nett.pc.1996 | -0.00109 | (0.000608) | -0.00112 | -0.00113 | -0.00123 |
| Greece factor | -0.02692 | (0.004659) | -0.03565 | -0.03572 | -0.03654 |
| % convergence | 1.232 | | 1.480 | 1.4905 | 1.6252 |
| $\sigma$ | 0.00629 | | 0.00635 | 0.00636 | 0.00677 |
| AIC | -662.72 | | -660.84 | -660.82 | -658.27 |
| $\rho$ | 0.29286 | (0.111) | | | |
| $\lambda$ | | | 0.22249 | 0.20981 | |
| LR test p-value | 0.011 | | 0.033 | 0.033 | |

as can be seen from the standard errors, the spatial lag model performs quite adequately, being a somewhat better representation than either the spatial error model or the non-spatial model.

Tiefelsdorf and Griffith (forthcoming) discuss in detail a semi-parametric approach to handling the misspecification challenge when known and available explanatory variables have been included. They argue that it is possible to include a set of spatial proxy variables which absorb remaining spatial pattern in the disturbances. Code implementing this approach has been contributed by Yongwan Chun and Michael Tiefelsdorf, and is included in **spdep** as function `SpatialFiltering()`. The function solves the eigenproblem of a quadratic form including the spatial weights matrix and the projection matrix of the regression, and searches over combinations of eigenvectors to find the smallest set best spatially "whitening" the disturbances of the original regression if they are added on the right hand side (for a full exposition please refer to Tiefelsdorf and Griffith, forthcoming). Once these are found, the search terminates and the spatial proxy variables, here a matrix with three columns, can be added to the model for estimation by least squares.

```
> res <- SpatialFiltering(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, nb = mst_nb, style = "W")
> SFilt <- res$dataset
> SF <- lm(growth ~ log(gva89) + hc96 + nett.pc.1996 + fGR + SFilt)
```

Table 10 reports the output from estimating the model including human capital and agricultural policy support variables with spatial proxy variables[10]. Akaike's AIC shows improvement on the models without the spatial proxy variables, as does the analysis of variance between these estimates and those without the proxy variables, reported as the $F$ test result on the final line of the table. This is clearly a satisfactory result, sharpening the precision of the coefficient estimates, with all having signs as expected, and all, including agricultural support, now significant at conventional levels of significance. None of the specification tests are now significant. Perhaps the only element of uncertainty not carried through is that associated with the choice of the proxies, of course over and above the worry that we could

[10]It appears to be the case that spatial proxy standard error estimates are equal in general.

Table 10: Results and specification tests for the model including human capital and agricultural policy support variables estimated using semi-parametric filtering of spatial autocorrelation.

| | value | std. error | p-value |
|---|---|---|---|
| Constant | 0.043655 | 0.004910 | < 1e-08 |
| $\theta$ | -0.015000 | 0.002130 | < 1e-08 |
| hc96 | 0.003156 | 0.000918 | 0.00093 |
| nett.pc.1996 | -0.001231 | 0.000551 | 0.02814 |
| Greece factor | -0.036537 | 0.002546 | < 1e-08 |
| SFiltvec5 | 0.016869 | 0.005695 | 0.00401 |
| SFiltvec3 | 0.015378 | 0.005695 | 0.00844 |
| SFiltvec13 | -0.013551 | 0.005695 | 0.01970 |
| SFiltvec20 | -0.013900 | 0.005695 | 0.01684 |
| SFiltvec31 | 0.017912 | 0.005695 | 0.00232 |
| SFiltvec7 | -0.010398 | 0.005695 | 0.07158 |
| SFiltvec6 | -0.009048 | 0.005695 | 0.11603 |
| % speed of convergence | 1.63 | | |
| $\sigma$ | 0.0057 | | |
| adjusted $R^2$ | 0.785 | | |
| AIC | -684.2 | | |
| Studentized Breusch-Pagan test | 10.3 | | 0.5 |
| RESET test | 1.46 | | 0.13 |
| Moran's $I$ | -0.104 | | 0.47 |
| F-test | 6.2 | | 7.9e-06 |

proxy away patterning in the residual that other diagnostic tools ought to have detected (for example outliers). Using semi-parametric filtering on the initial simple model, with nine eigenvectors selected, yields a significant estimate of the $\theta$ coefficient with the wrong sign: 0.00967, and a very satisfactory adjusted $R^2$: 0.766.

Finally, partly because we adopted this approach in our earlier work, we estimate the model including human capital and agricultural policy support variables by geographically weighted regression, to explore whether any traces of non-stationarity can be found. The use and development of geographically weighted regression is fully discussed in Fotheringham, Brunsdon and Charlton (2002), and discussion with the authors have been very helpful in writing a package for R. In this case, we use the same region label point coordinates as for constructing the sphere of influence graph neighbours, and find the adaptive bandwidth (percentage of regions included in each weighted regression) for a Gaussian kernel, by cross-validation.

The `print()` method for the object returned by the `gwr()` function provides a convenient summary of the results. The adaptive bandwidth includes about half of the regions in each regression, with weights declining with increasing distance in a Gaussian kernel. None of the variables displays a change of sign, although there is some variation in the local regression coefficient estimates. The Akaike AIC values are given as specified in Fotheringham, Brunsdon and Charlton (2002), with the reservations expressed there, and in the source code for this part of the function contributed by Danlin Yu[11]. If we use the corrected form `AICc`, the improvement on the least squares fit for the same model is not large, and suggests that no non-

[11] see the source package for further details, file `R/gwr.R`.

stationarity misspecification remains.

```
> library(spgwr)
> adapt.gauss.cv <- gwr.sel(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, coords = coords, adapt = TRUE, verbose = FALSE)
> gauss1.gwr <- gwr(growth ~ log(gva89) + hc96 + nett.pc.1996 +
+     fGR, coords = coords, adapt = adapt.gauss.cv, hatmatrix = TRUE)
> gauss1.gwr

Call:
gwr(formula = growth ~ log(gva89) + hc96 + nett.pc.1996 + fGR,
    coords = coords, adapt = adapt.gauss.cv, hatmatrix = TRUE)
Kernel function: gwr.gauss
Adaptive quantile: 0.5006546 (about 46 of 93)
Summary of GWR coefficient estimates:
                   Min.    1st Qu.     Median    3rd Qu.       Max. Global OLS
X.Intercept.  0.0146000  0.0254300  0.0315700  0.0480500  0.0513900     0.0437
log.gva89.   -0.0169300 -0.0160000 -0.0111200 -0.0084310 -0.0035510    -0.0150
hc96          0.0020080  0.0024740  0.0030860  0.0035630  0.0039240     0.0032
nett.pc.1996 -0.0016070 -0.0014160 -0.0010090 -0.0007013 -0.0004114    -0.0012
fGRTRUE      -0.0424300 -0.0403700 -0.0333100 -0.0314000 -0.0274500    -0.0365
Number of data points: 93
Effective number of parameters: 11.22109
Effective degrees of freedom: 81.77891
Sigma squared (ML): 3.776752e-05
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): -660.597
AIC (GWR p. 96, eq. 4.22): -674.2285
Residual sum of squares: 0.00351238
```

Table 11: Correlations between local GWR coefficient estimates.

| | X.Intercept. | log.gva89. | hc96 | nett.pc.1996 | fGRTRUE |
|---|---|---|---|---|---|
| X.Intercept. | 1.000 | | | | |
| log.gva89. | -0.987 | 1.000 | | | |
| hc96 | -0.471 | 0.323 | 1.000 | | |
| nett.pc.1996 | -0.921 | 0.911 | 0.414 | 1.000 | |
| fGRTRUE | -0.973 | 0.970 | 0.408 | 0.955 | 1.000 |

Table 11 does however reveal a problem with geographically weighted regression studied in detail by Wheeler and Tiefelsdorf (2005), that strong correlation can be found between the local coefficient estimates. We can see that there are strong negative correlations between the constant term and `log(gva89)`, and the constant term and the Greek regions factor, and strong positive correlation between `log(gva89)`, the Greek regions factor, and agricultural support. The human capital variable local coefficient estimates are not highly correlated with the other local estimates. These artefacts do, as Wheeler and Tiefelsdorf (2005) suggest, reduce confidence in the use of geographically weighted regression for more than exploratory purposes until their origins have been resolved satisfactorily.

In substantive terms, we can conclude that, once misspecification issues are addressed by adding relevant explanatory variables, the modelling of spatial dependence using the spatial lag form can be used to handle remaining spatial patterning. As we can see, little variation is observed in the coefficient estimates yielded by the various estimators, and the use of geographically weighted regression does not reveal any remaining marked non-stationarity. Consequently, the role of human capital and net support to agriculture can be accepted as providing useful additional information about rates of regional growth over and above the initial levels of gross value added.

# 5 Concluding remarks

One of our intentions in the work presented here has been to revisit our conclusions from earlier work, that there are interactions between regional growth and agricultural support in a conditional β-convergence context, and that higher levels of support are associated with lower levels of growth. We feel that this is certainly sustained, even after another explanatory variable operationalising human capital has been introduced. Beyond this, we feel that the use of a dummy variable for Greek regions can be justified with reference to the very different exchange rate conditions to which they were subject in this period. We also feel that the traces we have found of substantive spatial effects, in the form of a significant spatial lag model, fit well with the interpretations given by Fingleton and López-Bazo (2006), and that our change of neighbour representation has sharpened the resulting picture. We would also like to reiterate a point made in Bivand and Brunstad (2003), that our treatment of GVA here does not include amenity effects as a component of agricultural subsidies. Positive external effects such as the amenity value of the cultural landscape will be an important part of the Green Box, and are not currently acknowledged in our operationalisation. Taking proper account of amenity effects is of course of vital importance for the policy implications of a negative relation between agricultural support and regional growth.

The other intentions are related to the way in which different methods are made accessible to the research community, especially in the light not only of rapid progress in the elaboration of techniques used in spatial econometrics, but also in the face of criticism of some of our typical analytical procedures. In other fields of scientific enquiry, shared access to code and often data permits the reproduction of results, partly to check whether choices made by the analyst in terms of methods and their implementations have impacted conclusions drawn. Many of the functions needed for spatial econometrics are now available to researchers, and collaborative software development and distribution is progressing. In order to pay appropriate attention to critical views, such as those of McMillen (2003), or Wall (2004), it is very helpful to be able to reproduce their results in an open source environment, allowing in principle every facet of the problems exposed to be examined. The choice of methods available is also growing, and different research communities have their own preferences for different kinds of formulations, as for example Banerjee, Carlin and Gelfand (2004) make clear.

We have, then, attempted by example to show how many alternative implementations of estimators are being written using R, and are being made available to students and researchers. This collaborative process is enhanced by contacts between different disciplines and fields of study facing similar tasks of estimating models with data from spatial aggregates, be they in regional science, epidemiology or other fields. Ideally one would hope that conclusions about observed response variables were not too heavily conditioned by the procedures of analysis typically chosen in a given field of study, and that McMillen's precautionary warning (p. 215) will lead to greater care in the choice of explanatory variables, and of functional form, so as to reduce the possibility of misspecification being represented in our results as spatial autocorrelation.

# References


Anselin L (2002) Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27: 247–267

Avis D, Horton J (1985) Remarks on the sphere of influence graph. In Goodman JE et al. (eds) *Discrete Geometry and Convexity*, New York, New York Academy of Sciences 323–327

Banerjee S, Carlin BP, Gelfand AE (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC, Boca Raton

Barro RJ, Sala-i-Martin X (1992) Convergence. *Journal of Political Economy* 100: 223–251

Bivand RS (2002) Spatial econometrics functions in R: classes and methods. *Journal of Geographical Systems* 4: 405–421

Bivand RS (2006) Implementing spatial data analysis software tools in R. *Geographical Analysis* 38: 23–40

Bivand RS, Brunstad RJ (2003) Regional growth in Western Europe: an empirical exploration of interactions with agriculture and agricultural policy. In: Fingleton B (ed) *European Regional Growth*, Springer, Berlin, 351–373

Bivand RS, Portnov BA (2004) Exploring spatial data analysis techniques using R: the case of observations with no neighbors. In: Anselin L, Florax RJGM, Rey SJ (eds) *Advances in Spatial Econometrics*, Springer, Berlin, 121–142

Bivand RS, Szymanski S (2000) Modelling the spatial impact of the introduction of Compulsory Competitive Tendering. *Regional Science and Urban Economics* 30: 203–219

Boldrin M, Canova F (2001) Inequality and Convergence: Reconsidering European Regional Policies. *Economic Policy* 32: 207-253.

Chambers JM, Hastie TJ (1992) Statistical models. In: Chambers JM, Hastie TJ (eds) *Statistical Models in S*, Wadsworth & Brooks/Cole, Pacific Grove CA 13–44

Fingleton B (1999) Spurious spatial regression: some Monte Carlo results with spatial unit root and spatial cointegration. *Journal of Regional Science* 39: 1–19

Fingleton B (2003) *European Regional Growth*. Springer, Berlin

Fingleton B, López-Bazo E (2006) Empirical growth models with spatial effects. *Papers in Regional Science* (this number)

Fotheringham AS, Brunsdon C, Charlton M (2002) *Geographically Weighted Regression*. Wiley, Chichester

Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509–533

Koenker R, Ng P (2005) SparseM: Sparse Linear Algebra. R package version 0.61 `http://www.econ.uiuc.edu/~roger/research/sparse/sparse.html`

Kosfeld R, Lauridsen J (2004) Dynamic spatial modelling of regional convergence processes. *Empirical Economics* 29: 705–722

Leisch, F., Rossini, A., 2003. Reproducible statistical research. *Chance* 16: 46–50

McMillen DP (2003) Spatial autocorrelation or model misspecification? *International Regional Science Review* 26: 208–217

Petrakos G, Rodríguez-Pose A, Rovolis A (2005) Growth, integration, and regional disparities in the European Union. *Environment and Planning A* 37: 1837–1855

R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`

Rey SJ, Janikas MV (2005) Regional convergence, inequality, and space. *Journal of Economic Geography* 5: 155–176

Schabenberger O, Gotway CA (2005) *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, Boca Raton

Tiefelsdorf M, Griffith DA (forthcoming) Semi-parametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*

Wall MM (2004) A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* 121: 311–324

Waller LA, Gotway CA (2004) *Applied spatial statistics for public health data*. Wiley, Hoboken NJ

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7: 161–187