# Further explorations of interactions between agricultural policy and regional growth in Western Europe: approaches to nonstationarity in spatial econometrics*

Roger Bivand[†] Rolf Brunstad[‡]

15th June 2005

**Abstract**

**Keywords:** spatial data analysis, open source software, regional growth, statistical computing.

## 1 Introduction

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. We tried to explore the possibility that some of the observed specification issues in current results are rooted in neglecting agricultural policy interventions, within the limitations imposed by data available. We also attempted to use this as a case setting for evaluating the appropriateness of geographically weighted regression (GWR) as a technique for assessing coefficient variability, over and above for instance country dummies, but possibly reflecting missing variables or other specification problems.

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 for the base year. We also look at ways of handling changes in agricultural policy regime occurring between years 0 and T. This raises the further challenge of looking at both spatial and temporal dimensions at the same time, which we will discuss, but are not likely to resolve satisfactorily.

*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 300, 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

On the technical side, the tests on GWR estimates also need to be more firmly established. The GWR results also need to be tested for spatial autocorrelation, and re-worked in an adaptive weighting framework, although GWR does already involve a spatial weighting of the observations themselves. The paper is therefore 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, and the **spgwr** package for GWR fitting. In particular, specific issues regarding the handling of the Jacobian in fitting spatial simultaneous autoregressive (SAR) models, and in interpreting GWR output will be discussed. Concentrating on implementations in R is justified by the preliminary nature of many of these methods requiring open source and replicable statistical research approaches, so that others can, if they wish, see how results were calculated.

One such technical issue is the representation of neighbours in the various approaches, and of the impact of symmetry requirements in conditional autoregressive (CAR) models typically used in MCMC estimation using OpenBUGS and elsewhere. Indeed, in many SAR models, symmetry is also required, or at least underlying symmetry, with the weights matrix in the row-standardised weighting scheme typically being similar to a symmetric matrix. Using the Western European regional growth data augmented with agricultural policy variables, we will try to explore how far some as-yet unresolved technical questions impede progress with substantive interpretation.

The paper has two threads, one focussing on the analysis of the relationships between regional growth and agricultural policy, generating models needing testing, while the other attempts to meet the software demands generated in the first thread, and to incorporate on-going research in spatial data-analytic methods to respond adequately to the potential importance of the substantive research question.

# 2 Convergence, agriculture and agricultural policy

Rather than review the convergence literature broadly, we prefer to focus on suggestions pointing up issues to be explored here. We will therefore be taking some positions as given, and will only mention them briefly for clarity. We will be concerned with β-convergence as represented in empirical studies in the following way:

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

where $\alpha$ and $\beta$ are coefficients and $u_i$ is a disturbance term (Paci, 1997, p. 617). Given an estimate of $\beta$, the speed of convergence may be represented as: $\theta = -\log(1+T\beta)/T$, with $100\theta$ expressing this speed in percentage points (Baumont et al. 2001, p.8).

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 $y_{i,0}$. The measure to be used here is gross value added per capita measured in EUR 1000 at constant 1990 prices. Figure 1 shows the regional distribution of this variable for 1989, the initial period to be used here except where stated. Details of the choice of period and regions will be given below in section 2.2. The contrast between lower values in most of the Iberian peninsula and southern

Italy and the rest of the study area is familiar, as are the effects of regional definition artifacts in Benelux and western Germany, with high urban values contrasting with apparently lower surrounding rural values despite commuting.

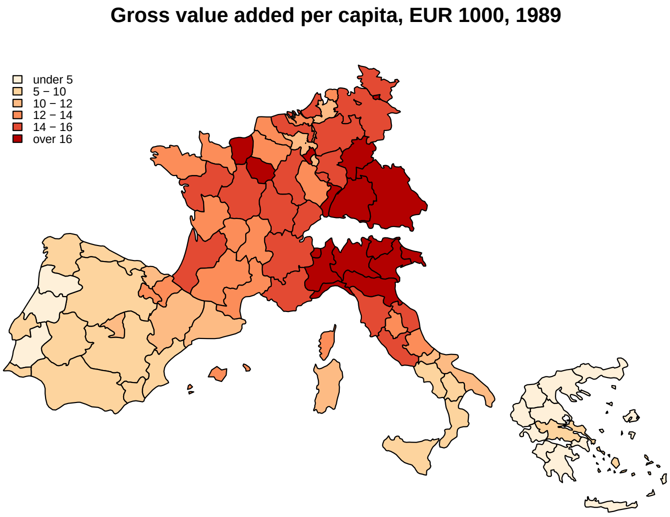


Figure 1: Gross value added per capita, EUR 1000, 1989 ($y_{i,0}$).

Figure 2 appears to fit well to the initial conditions: regions such as those in Spain and Portugal with low initial condition values have high growth rates, while French regions have low growth rates and medium initial conditions. However, closer inspection suggests that the stories of particular regions, or clubs of regions, are more complex than our simple convergence model indicates. This is strengthened by an examination of Table 1, and by the insignificance of a $\chi^2$ test (25.776, df = 25, p-value = 0.42) on the relationship.

Table 1: Contingency table of initial conditions (1989) by growth rates (1989-1999), both variables cut at figure class intervals.

| | [4,10] | (10,12] | (12,14] | (14,15] | (15,16] | (16,27] |
|---|---|---|---|---|---|---|
| [0.99,1] | 0 | 0 | 1 | 0 | 0 | 0 |
| (1,1.11] | 3 | 2 | 8 | 7 | 2 | 8 |
| (1.11,1.16] | 2 | 4 | 7 | 4 | 3 | 4 |
| (1.16,1.21] | 4 | 2 | 4 | 4 | 2 | 6 |
| (1.21,1.26] | 4 | 5 | 0 | 1 | 4 | 5 |
| (1.26,1.5] | 5 | 3 | 6 | 3 | 0 | 2 |

Among others, Paci notes that "the observed process of aggregate convergence can hide important structural change phenomena" (1997, p. 617). In analysing sectoral labour productivity, Paci finds that without using a "Southern" dummy or national dummies, the convergence relationship for agriculture is not significant (1997, p. 627). Others, including Fagerberg et al. (1997), Pons-Novell and Viladecans-Marsal (1999), Paci and Pigliaru (1999), and López-Bazo et al. (1999),

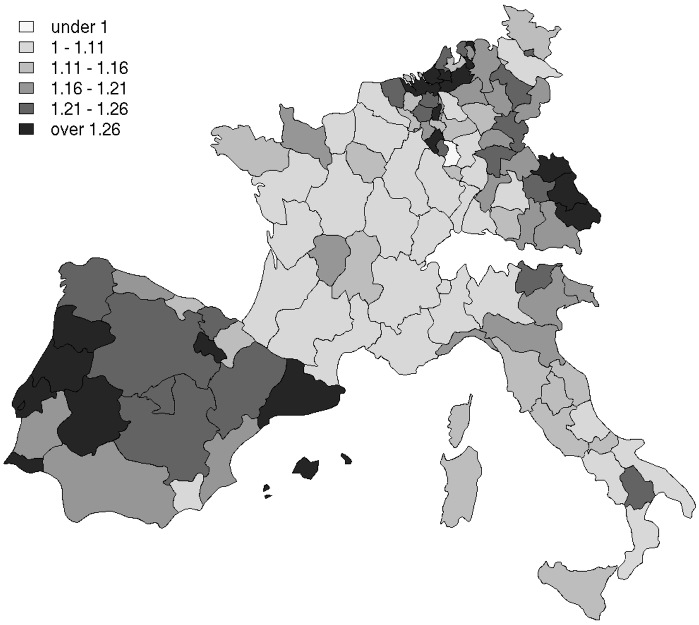


Figure 2: Regional economic change: GVA per capita 1999 as proportion of GVA per capita 1989 ($y_{i,1999}/y_{i,1989}$).

also draw attention to the specific structural role of agriculture in empirical analyses of convergence in European regions. There are of course also other structural phenomena of interest, but here we will concentrate on agriculture.

## 2.1 The impact of agricultural policy

The potential impacts of agricultural policy in the European Union on cohesion have been central in changes in the measures and component parts of the Common Agricultural Policy over the past decade. Cohesion is understood as accelerating regional economic growth in those parts of the EU with region GVA per capita markedly below that of developed regions, and thus most regions receiving cohesion support are agricultural regions, although not all agricultural regions are cohesion regions. The basic features of these measures, and changes taking place, including the Mac Sharry reforms and Agenda 2000, have recently been surveyed by Colman (2001), and links to regional policy are covered by Tondl (2001).

The second EU report on economic and social cohesion stresses the role of convergence, and concludes that specific measures will be needed to eliminate regional disparities (DG REGIO, 2001b). These will address differences in underlying conditions and factor endowments, among which labour force skills are seen as central. In addition, a preliminary study was devoted to the impact of community agricultural policies on cohesion (DG REGIO, 2001a). In this study, and more generally in the economics of tariff system structures, attention is drawn to difficulties in adequately measuring economic assistance.

Agricultural policy could be expected to interact with convergence in two ways. Firstly, and for practical reasons the only relationships to be explored empirically, one could expect the proportion of subsidised agriculture in a regional economy, and the intensity of the support, to influence the region's growth rate negatively; for present purposes we assume that agriculture may be treated as though it were uniformly subsidised. The reasons for these negative relationships are that subsidies attenuate the movement of labour and capital to other sectors (and/or regions) with higher returns, conserving structures of factor allocation at the cost of those paying for the subsidies. The subsidies may also also be expected to reduce or to distort incentives to farmers to change their mixes of products and/or methods of production.

In this sense the subsidies are counterproductive as they hamper the growth in GVA. However the recent discussion about the so-called multifunctionality of agriculture may indicate that agricultural activities produce benefits over and above the market value of agricultural production. In economic terms agricultural production may have positive external effects on perceived public goods like the amenity value of the cultural landscape. If this is the case, and if agricultural subsidies are used as a means to internalise these externalities, growth is reduced only because we are measuring the wrong thing, traditional GVA instead of an extended GVA including the willingness to pay for such amenities. Whether or not this is the case is of course of vital importance for the policy implications of a negative relation between agricultural support and regional growth. For a recent paper addressing this question see Brunstad et al. (1999).

The second approach not followed up here, would be to consider the impact on speed of convergence of changes in agricultural policy regimes; for such an approach to be considered, regionalised agricultural accounts would be needed. Since they are not available in a systematic form at scales and levels of detail needed for Europe-wide analysis, micro-level studies would probably be required, such as panel studies, to cast light on the detailed relationships between different subsidy regimes and the embedding of agriculture in regional economies.

While governments granting subsidies to producers can account for them from public expenditure, other forms of subsidy cannot be as readily measured. In particular this applies to subsidies based on tariffs, quota systems, import bans, etc., where the transfer is carried out with consumers within the trade barrier system subsidising producers by an amount equivalent to the difference between the local price and the price of the same good delivered to that market from an external source at world market prices. Measurement complications and production distortions here also affect markets in intermediate goods.

The most commonly adopted approach at the national scale is to estimate agricultural assistance using producer subsidy equivalents, an approach used in a number of international organisations, and described in detail by Cahill and Legg (1989). The data requirements are however substantial, not just for price and quantity series for the chosen commodities, but also insight into the intermediate agricultural goods involved in regionally varying production processes. Some of the work required to estimate regional PSE series has been carried out in the DG REGIO study (2001a), yet more in Heckelei and Britz (2000). Since the present study is only intended to flag the importance of the agricultural sector, it was found more appropriate to use less adequate but more accessible data, rather than further reduce the number of regions under consideration or increase uncertainties associated with estimating or interpolating variables.

## 2.2 Sources of data in agriculture

European agricultural accounts are available in two versions, EAA 89/92 and EAA 97 REV.1.1, and regional versions at the level of NUTS level 2 are available from Eurostat. The accounting data used here is for agriculture gross value added at market prices, subsidies, taxes linked to production (including VAT balance), and gross value added at factor costs, for the period 1988-1998, in million ECU. From the middle of the period, there are very many missing values, and some countries either only report at NUTS level 1, or level 0. For this reason, The study area does not include Ireland, Denmark, Greece, UK, Berlin and former East Germany, and also drops overseas dependencies and Atlantic islands. It was further found necessary to aggregate three regions in Belgium (BE1 Brussels, BE24 Vlaams Brabant, and BE31 Brabant Wallon) in order to maintain the spatial series. The number of regions included in the analysis is consequently 115; not covering Ireland and Greece is unfortunate, because — had data been available – they would have cast additional light on the issues under study. In order to smooth the agricultural accounts data somewhat, and to accommodate further missing data problems, the variables to be used below are averages of values reported during 1988-1990, in most cases but not all, the averages of three values.[1]

[1]It is unfortunate that the data are not more complete, because the choice of initial conditions means that some older members will appear to have higher support than newer members, and also

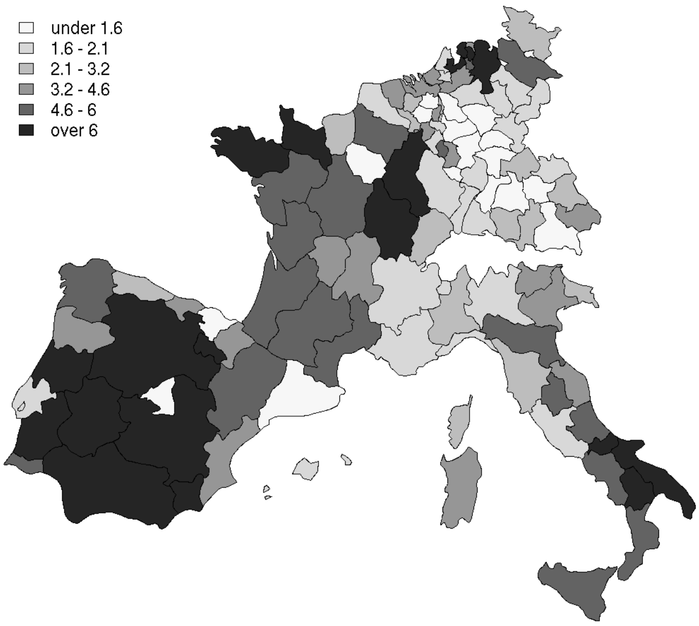


Figure 3: GVA in agriculture at factor costs (average 1988-90) as percentage of total GVA 1989.

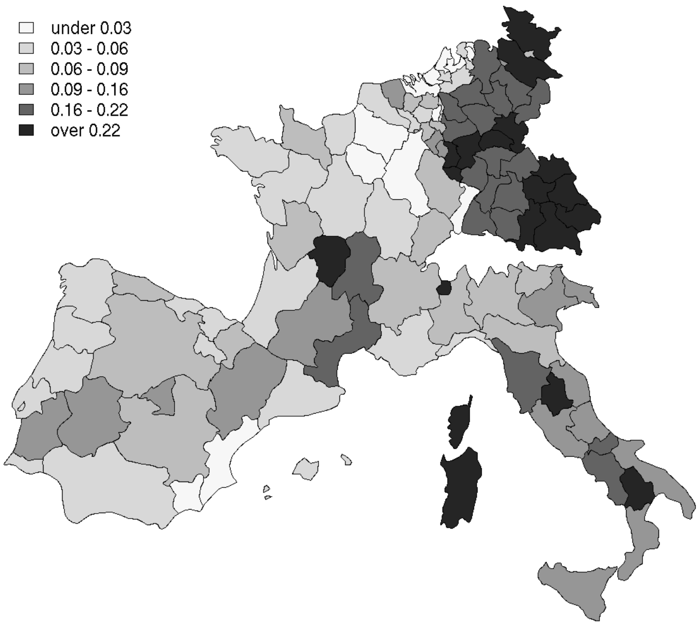


Figure 4: Relative subsidy levels 1988-90: agricultural subsidies (average 1988-90) as proportion of GVA in agriculture at market prices (average 1988-90).

Figure 3 shows average agricultural GVA measured at factor costs 1988-90 (in EAA 89/92 nomenclature, GVA at factor costs is GVA at market prices plus subsidies minus taxes linked to production) as a percentage of total GVA for 1989. The underlying total GVA and population datasets have been taken from Cambridge Econometrics' European Regional Databank, and are measured in 1990m ECU and 1000 persons. The two datasets have been merged after dropping NUTS level 2 regions that could not be used because of missing agricultural accounts data. There is a potential problem of double-counting involved in using regional GVA series, because they can and most often do include subsidies as a component of gross value added.[2]

Figure 4 displays the other variable to be used to represent the influence of agricultural policy, agricultural subsidies as a proportion of agricultural GVA at market prices. This is necessarily a very inadequate substitute for a properly constructed measure of producer subsidy equivalents, because it does not encompass the effects of tariff structures, which are not simply proportional to subsidy payments, but vary regionally with produced quantities of commodities. It has been chosen to restrict the agricultural variables to the 1988-90 period, partly because of missing data in the mid-1990's in agricultural accounts, but also because the base year date for the convergence model is 1989 - also chosen to match the years with least missing agricultural data.

# 3 Estimating convergence: specification issues

Attention has been drawn in a series of studies to specification problems found in estimating the standard convergence model using OLS. The roots of these problems are partly related to substantive spatial relationships, such as spillovers (Vayá et al., 1998), but may also involve missing variables, structural differences across the chosen study area, and functional form. Fingleton and McCombie (1998) and Fingleton (1999a, 2001) draw attention to the clear need to pay attention to specification, over and above the introduction of spatial econometric techniques also made by Vayá et al. (1998), Baumont et al. (2001), and in the North American context by Rey and Montouri (1999) and Rey (2001). Details of the estimation methods and tests are not repeated here for brevity, and may be found in the cited articles.

Results from the estimation of the standard convergence model, with the annual log GVA per capita growth rate (1989-1999) as the dependent variable and log GVA per capita in 1989 as the independent variable, are presented in Table 2. The table also shows the results of estimating the same model including the spatially lagged dependent variable using maximum likelihood, and of estimating an augmented model including the two agriculture variables defined above in section 2.2 for both estimation methods. The percentage speed of convergence for the OLS standard model, 0.75%, is comparable with that reported in Baumont et al. (2001, p. 26) of

because support to the older members declined during the chosen period while support to newer members probably rose.

[2] In addition, changes introduced in revisions of the System of National Accounts, knocked on to EAA 97, moving some subsidies into the definition of basic prices (OECD 2000); consequently the regional total GVA series may include varying amounts of subsidies depending on whether they are based on market prices or basic prices.

Table 2: Modelling convergence 1989-1999 (t-values or z-values in parentheses)

| | OLS | ML lag | OLS | ML lag |
|---|---|---|---|---|
| 1 Intercept | 0.0346 | 0.0147 | 0.0400 | 0.0203 |
| 2 | (6.56) | (2.86) | (5.52) | (2.91) |
| 3 log GVA pc 1989 | −0.00725 | −0.00287 | −0.00987 | −0.00498 |
| 4 | (−3.55) | (−1.61) | (−3.99) | (−2.24) |
| 5 % speed of convergence | 0.753 | 0.292 | 1.04 | 0.511 |
| 6 log % agriculture 1988-90 | | | −0.002 | −0.00133 |
| 7 | | | (−2.19) | (−1.68) |
| 8 log subsidy/GVA 1988-90 | | | −0.00138 | −0.000699 |
| 9 | | | (−1.9) | (−1.11) |
| 10 spatially lagged dependent variable | | 0.546 | | 0.512 |
| 11 | | (5.99) | | (5.41) |
| 12 $\sigma$ | 0.00704 | 0.00597 | 0.00687 | 0.00591 |
| 13 log likelihood | 407.8 | 421.1 | 411.7 | 422.8 |
| 14 AIC | −809.6 | −834.1 | −813.3 | −833.7 |
| 15 Chow | 3.772 | 6.225 | 1.284 | 5.808 |
| 16 p-value | 0.026 | 0.044 | 0.28 | 0.21 |
| 17 RESET F | 3.821 | | 1.067 | |
| 18 p-value | 0.0248 | | 0.387 | |
| 19 Breusch-Pagan | 0.0486 | 0.0012 | 2.09 | 0.908 |
| 20 p-value | 0.826 | 0.972 | 0.555 | 0.824 |
| 21 Moran's I | 0.332 | | 0.308 | |
| 22 Z | 5.6 | | 5.38 | |
| 23 Robust LM (err) | 0.444 | | 0.0915 | |
| 24 p-value | 0.505 | | 0.762 | |
| 25 Robust LM (lag) | 3.86 | | 2.41 | |
| 26 p-value | 0.0495 | | 0.120 | |

0.84% for 1980-1995 and 138 regions.

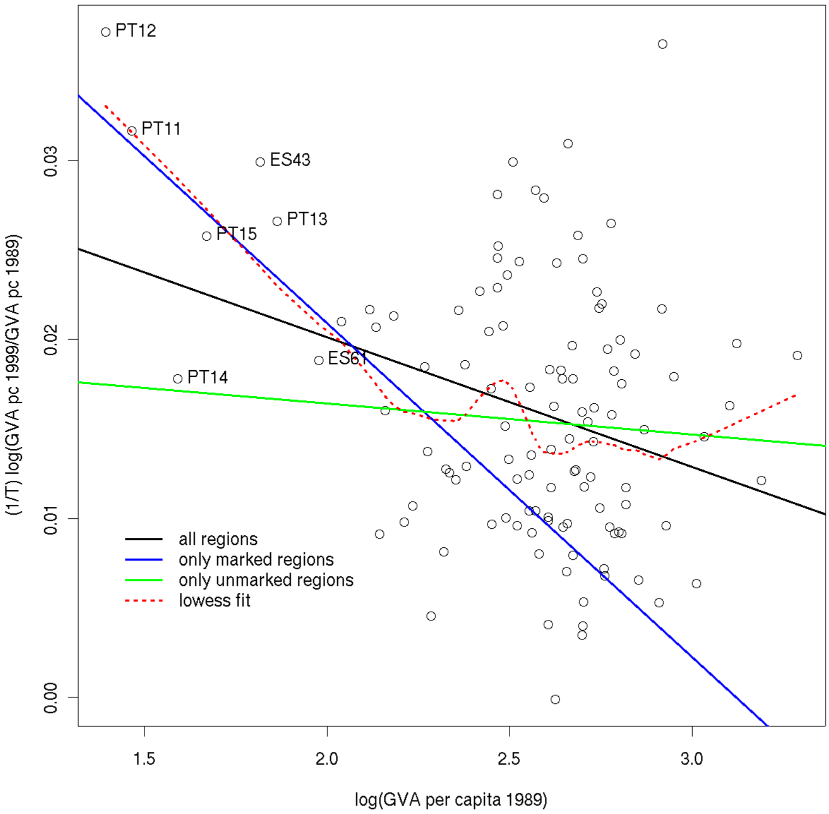


Figure 5: Plot of standard OLS convergence model (1989-1999) with regression fitted lines for all regions, only SW Iberia, all without SW Iberia, and a Lowess fit.

Concentrating first on the OLS estimates of the standard model, it is worth noting that while the Breusch-Pagan test does not reject homoskedasticity, the RESET test (Johnson and DiNardo, 1997, p. 121) using the second and third powers of the fitted values as extra variables indicates the presence of some specification error (the same result is found for all T=1990-1999 for base year 1989, and often for other base years). Figure 5 gives us a view of the relationship: it seems that the regions belonging to South West Iberia (the Portuguese regions and the Spanish regions of Extramadura and Andalucia) differ structurally from the remainder. They have much lower initial condition values and much higher growth rates, and the slopes of regression lines for all regions, just SW Iberia, and all without SW Iberia are quite different. The plotted Lowess fit (a robust local regression fit passing a moving window across the data set, see Cleveland, 1979) confirms that there are structural differences in the data set, which could be modelled by including the square of the independent variable, but may better be considered in a Chow test context (Johnson and DiNardo, 1997, p. 113-116). Subsetting the data set into SW Iberia and not SW Iberia, we can conduct a Chow test for structural difference in both slope and intercept coefficients, differences found to be significant as shown in Table 2.

The fact that the SW Iberia constitutes a block of outliers is also seen in Figure 7, a Moran scatterplot of the annual log growth rates 1989-1999 (see also Rey and Montouri, 1999, p.150). This plot of the values of the variable in question on the x-

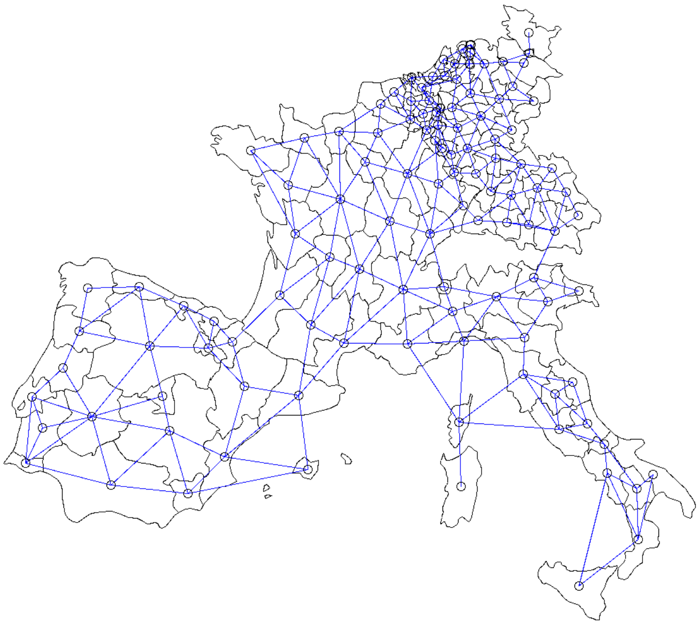

Figure 6: Sphere of influence neighbours for NUTS2 regions

axis against its spatially lagged values on the y-axis lets us see how the inclusion of the spatially lagged variable might give more insight into the convergence process, where growth rates would depend on the average of growth rates across neighbours of regions, possibly representing spillover or perhaps underbounding of functional regions. Here neighbours have been defined using a sphere of influence graph (see Figure 6) rather than boundary contiguity (Vayá et al. 1998) or distance (Baumont et al. 2001). Given a set of points representing the regions, a row standardised weights matrix was constructed and used in all analyses presented here.

The tests for spatial specification problems in the standard model, Moran's I (Cliff and Ord, 1981), and LM tests not reported here, are all highly significant — Figure 8 shows visually the spatial patterning of the residuals of the initial model. In addition, the robust LM test for a missing spatially lagged dependent variable in the presence of spatial error dependence is significant, but not the other way round (Anselin et al. 1996, Anselin and Bera 1998). It is not impossible that this lack of clarity in the results of the spatial specification tests is related to Fingleton's finding, that Moran's I may also detect spatial non-stationarity (1999b).

Following the augmentation of the standard model with the two agricultural variables, we can see that the non-spatial specification problems encountered in the standard model and reflected in the results of the Chow test and the RESET test are alleviated (column 3 in Table 2). Neither variable is strongly significant, but both have the expected signs, with both higher proportions of agriculture in regional GVA, and higher ratios of subsidies to agricultural regional GVA being associated

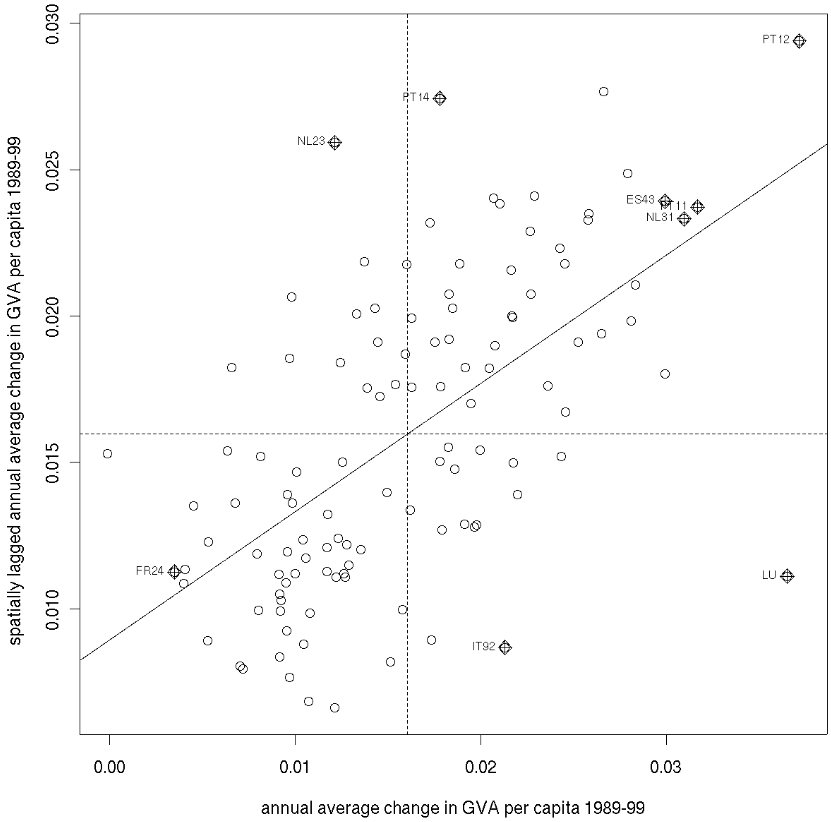


Figure 7: Moran scatterplot of convergence model dependent variable $1/T\,\log(y_{i,T}/y_{i,0})$

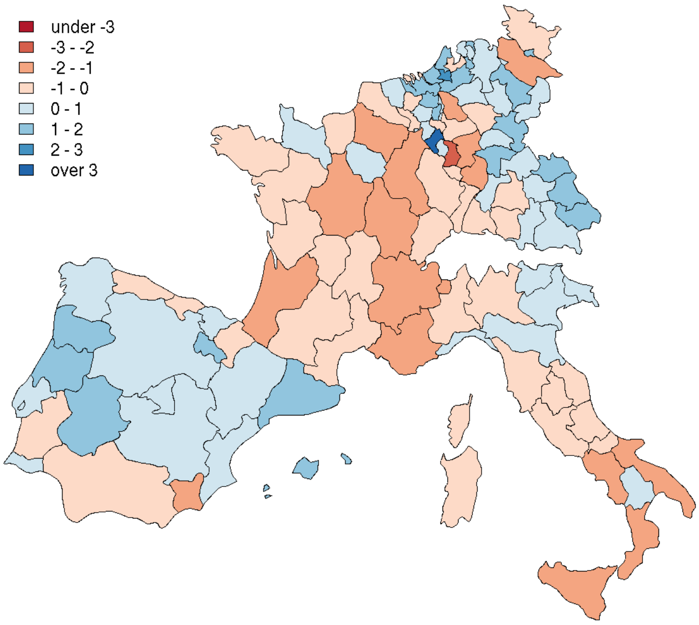


Figure 8: Standardised residuals of OLS model fit

with lower growth rates. The percentage speed of convergence is somewhat greater, at 1.04.

Stepping back to examine the ML lag estimates of the standard model, we can see that the inclusion of the spatially lagged growth rate reduces the percentage speed of convergence markedly, with the underlying coefficient value becoming much less significantly different from zero. The coefficient of the spatially lagged dependent variable is itself highly significant, and no residual spatial autocorrelation was found to remain in the model. Further details of estimation and testing methods may be found in Anselin (1988).

Introduction of the agricultural variables in the augmented ML lag model sees a further, but much less marked improvement in $\sigma$ and log likelihood; however the value of the Akaike Information Criterion (AIC) rises a little compared to the standard ML lag model, suggesting possible interaction between the spatially lagged growth rate and the added variables. Adding the spatially lagged independent variables did not help, with AIC rising further to -833.7 (AIC for the standard ML lag model with spatially lagged initial conditions was -834.1).

Before moving on, let us examine the local Moran's $I$ values for the residuals from the initial OLS model. The saddlepoint approximation has been shown by Tiefelsdorf (2002) to be better suited to local Moran's $I$ — which is the local expression for each spatial unit of the components summed to make global Moran's $I$ — than the Normal approximation. It may be noted that permutation tests are not necessarily suitable for local indicators of spatial association. Figure 9 shows the

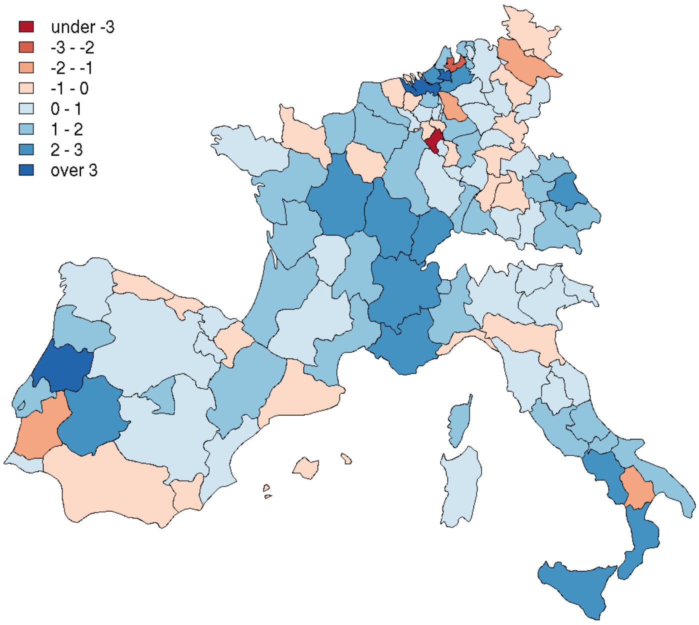


Figure 9: Saddlepoint approximation of local Moran's $I_i$ tests for OLS fit

values equivalent to the standard deviate of global Moran's $I$ for the residuals from the initial OLS model. Probability values are not presented, because here multiple tests would be performed on the same data.

Figure 10 extends the treatment of local Moran's $I$ to the residuals of a spatial error model (the spatial lag model is not accommodated by this approach). If the global spatial dependence is not removed in this way, local Moran's $I$ may misleadingly indicate dependencies which come from the global autocorrelation (or even a global spatial trend). Using the same class intervals and symbology, the removal of the global process pales the general impression. Finally, Figure 11 shows the local Moran's $I$ values following the addition of the two agriculture variables fitted by OLS. It seems that there is still plenty of information left in the map, perhaps strengthening concerns about non-stationarity.

# 4 Estimating convergence: exploring non-stationarity

As pointed out, Fingleton (1999b) has indicated that Moran's I may detect spatial non-stationarity in addition to residual autocorrelation. Because Moran's I continues to be significant in the augmented model, perhaps implying non-stationarity, and in the poor improvement of the augmented ML lag model over the standard ML lag model, it seems appropriate to explore the standard and augmented models using geographically weighted regression. Having originally been developed as a tool to exploratory spatial data analysis, its status is under revision at present,

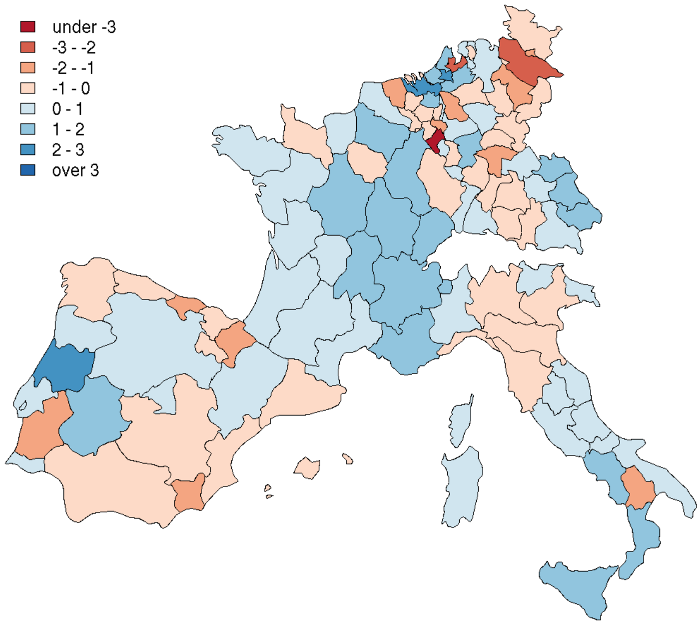


Figure 10: Saddlepoint approximation of local Moran's $I_i$ tests for SAR error fit

and it has been suggested that it may be used to indicate the presence of spatial non-stationarity in more formal ways. Exploratory methods have been used in the analysis of convergence models by Rey and Montouri (1999), Le Gallo and Ertur (2000), Rey (2001) and Arbia (2001).

## 4.1 Geographically weighted regression

Geographically weighted regression (GWR) is a technique for examining possible variability in coefficient estimates across study areas composed of regions represented by points, and was introduced by Brunsdon et al. (1996, a full description is found in Fotheringham et al. 2000). Instead of simply estimating global coefficient values over the whole data set, local parameter values are estimated for each region/point in the data set:

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \varepsilon_i$$

where $y_i, i = 1, n$ are observations of the dependent variable, $x_{ik}, i = 1, n, k = 1, p$ are observations of the $p$ independent variables, $\varepsilon_i$ are disturbance terms, and $\beta(i)$ are unknown parameter vectors which are functions of location $i$.

The local estimates, one regression for each region, are made using weighted regression, with the weights assigned to observations being a function of the distance between the region for which coefficient estimates are required and all the

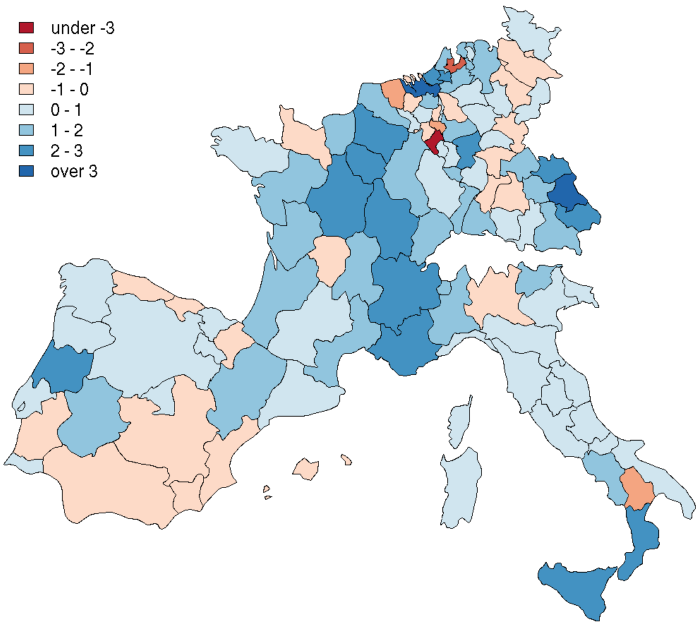


Figure 11: Saddlepoint approximation of local Moran's $I_i$ tests for augmented OLS fit

Table 3: GWR coefficient estimates and $R^2$ — fixed bandwidth

| | (Intercept) | log GVA pc 1989 | $R^2$ | Equivalent % speed of convergence |
|---|---|---|---|---|
| Minimum | −0.0286 | −0.04503 | 0.000 | 5.984 |
| Lower quartile | 0.0088 | −0.01086 | 0.071 | 1.150 |
| Median | 0.0196 | −0.00093 | 0.263 | 0.093 |
| Upper quartile | 0.0422 | 0.00298 | 0.676 | −0.294 |
| Maximum | 0.1278 | 0.02105 | 1.000 | −1.910 |
| Global OLS | 0.0346 | −0.00725 | 0.100 | 0.753 |

other regions. The distance function is directly comparable to those used in kernel density methods, and here use has been made of a bisquare function:

$$w(i)_{jj} = (1-(d_{ij}^2/d^2))^2, d_{ij} \leq d$$

where $w(i)_{jj} = 0$ when $d_{ij} > d$; $w(i)_{jj}$ is the weight assigned to observation $j$ in estimating the parameters for observation $i$. The parameter $d$ is termed the bandwidth, and for the bisquare function is the maximum distance for non-zero weights; it may be found by cross-validation and the fixed value used here to begin with is 375.8 km, found by cross-validation for the initial model. The bandwidth may also be adaptive, expressing the proportion of observations to retain within the weighting kernel "window". The parameters are estimated by weighted least squares:

$$\hat{\beta}(i) = [\mathbf{X}^T\mathbf{W}(i)\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}(i)\mathbf{y}$$

where $\mathbf{W}(i)$ is the diagonal weights matrix for observation $i$; $\mathbf{X}^T\mathbf{W}(i)\mathbf{X}$ is assumed to be invertible.

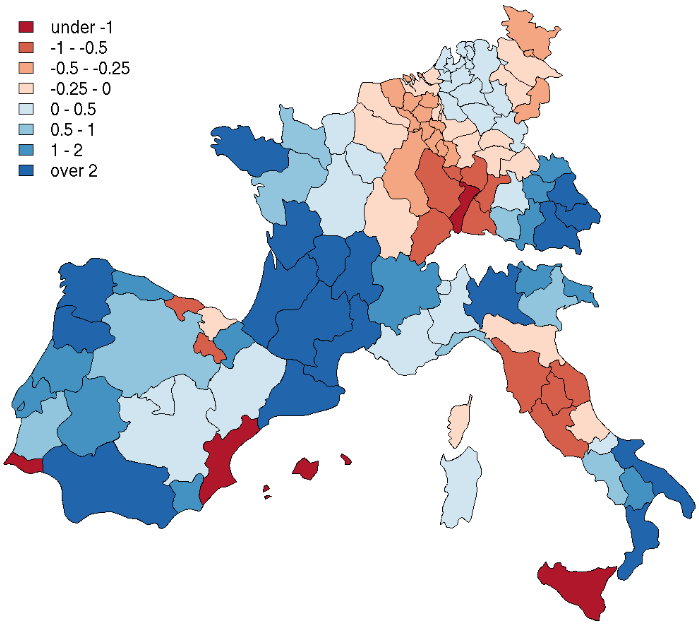


Figure 12: GWR percentage convergence estimates — fixed bandwidth

Table 3 shows the results of a geographically weighted regression for the initial model, using a bisquare kernel and a 375.8 km bandwidth. We see that there is considerable variation in the coefficient estimates, with both changing sign. The speeds of convergence equivalent to the coefficient values also cross zero; the coefficients of determination cover the full range between zero and unity. The percentage speeds of convergence by region are shown in Figure 12 but do not seem to convey a great

Table 4: GWR coefficient estimates and $R^2$ — adaptive bandwidth

| | (Intercept) | log GVA pc 1989 | $R^2$ | Equivalent % speed of convergence |
|---|---|---|---|---|
| Minimum | −0.0393 | −0.04127 | 0.001 | 5.322 |
| Lower quartile | 0.0047 | −0.01046 | 0.159 | 1.105 |
| Median | 0.0200 | −0.00185 | 0.352 | 0.186 |
| Upper quartile | 0.0440 | 0.00305 | 0.525 | −0.301 |
| Maximum | 0.1185 | 0.01879 | 0.859 | −1.722 |
| Global OLS | 0.0346 | −0.00725 | 0.100 | 0.753 |

deal, perhaps because of the very strong negative correlation, equal to -0.988, between the GWR coefficient estimates for the intercept and the explanatory variable. To check this, let us try the same analysis with adaptive bandwidths.

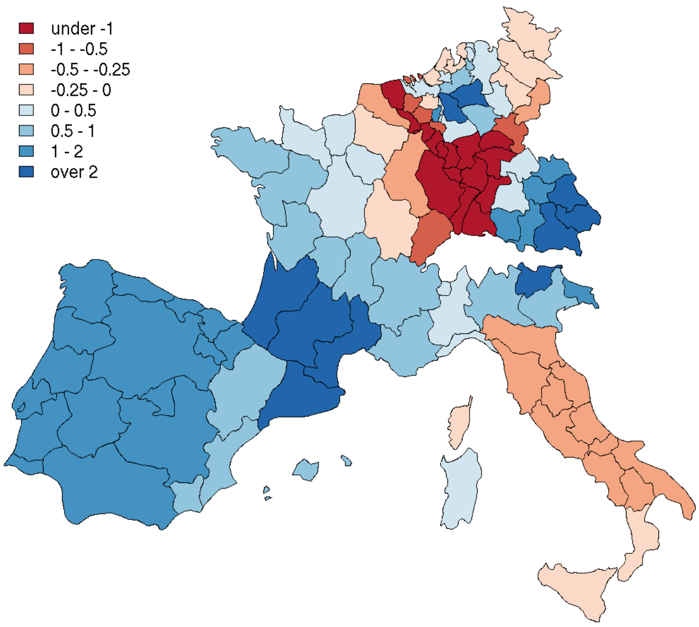


Figure 13: GWR percentage convergence estimates — adaptive bandwidth

Table 4 shows the results of a geographically weighted regression for the initial model, using a bisquare kernel and a 0.217 bandwidth, that is the proportion of observations that are included in the weighting window irrespective of distance. Figure 13 shows the percentage convergence values by region; the correlation between the GWR coeffcient estimates for the intercept and the explanatory variable is now -0.993.

When the GWR approach is applied to the model augmented with agriculture variables, with an adaptive bandwidth of 0.417 of the observations (found by cross-validation), the results show little change. Table 5 shows that all the explanatory

Table 5: GWR coefficient estimates and $R^2$ for augmented model — adaptive bandwidth

| | (Intercept) | log.GVApc89 | log.agric8890 | log.sub/GVA8890 | $R^2$ | % convergence |
|---|---|---|---|---|---|---|
| Minimum | −0.0490 | −0.02978 | −0.00421 | −0.00356 | 0.022 | 3.535 |
| Lower quartile | −0.0212 | −0.01061 | −0.00205 | −0.00227 | 0.159 | 1.121 |
| Median | 0.0029 | 0.00405 | 0.00161 | −0.00076 | 0.366 | −0.397 |
| Upper quartile | 0.0437 | 0.01139 | 0.00274 | 0.00072 | 0.498 | −1.079 |
| Maximum | 0.0915 | 0.02097 | 0.00380 | 0.00429 | 0.725 | −1.904 |
| Global OLS | 0.0400 | −0.00987 | −0.00200 | −0.00138 | 0.159 | 1.039 |

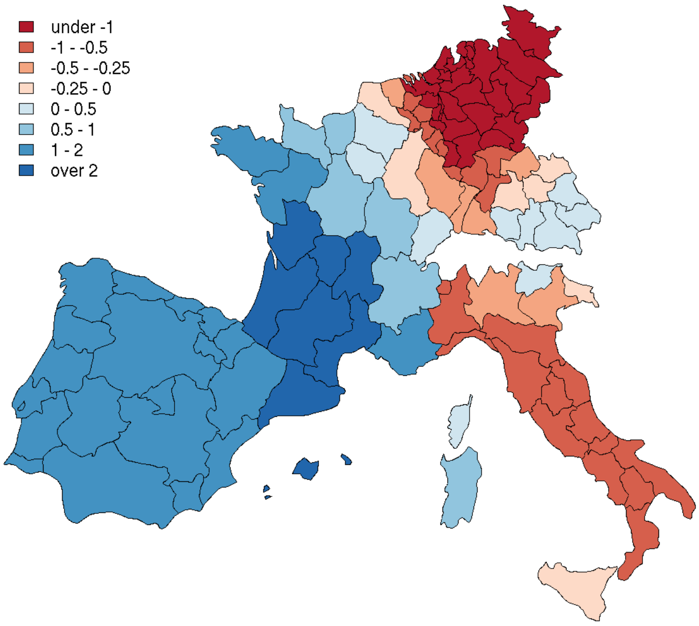


Figure 14: GWR percentage convergence estimates for augmented model — adaptive bandwidth

Table 6: Correlation matrix of explanatory variables

| | (Intercept) | log.GVApc89 | log(agproc89) |
|---|---|---|---|
| log.GVApc89 | −0.965 | | |
| log(agproc89) | −0.611 | 0.579 | |
| log(relsubs8890) | 0.289 | −0.066 | 0.099 |

Table 7: Correlation matrix of GWR coefficient estimates

| | (Intercept) | log.GVApc89 | log.agproc89. |
|---|---|---|---|
| log.GVApc89 | −0.996 | | |
| log.agproc89. | −0.900 | 0.911 | |
| log.relsubs8890. | 0.421 | −0.354 | −0.158 |

variables cross zero, and Figure 14 has a marked banded pattern, typical of GWR coefficient estimates maps. In a forthcoming paper, Wheeler and Tiefelsdorf (forthcoming) show that GWR coefficient estimates can be highly mutually correlated, in a negative sense. It appears that when one coefficient estimate moves in one direction, it seems to force at least one of the others in the other direction. In the present case as Table 6 shows, there is a good deal of correlation between the original explanatory variables anyway. Table 7 indicates that this is quite severe, and does raise doubts about the potential for using geographically weighted regression for exploring stationarity in spatial data.

# 5 Placeholder — conclusions

Since this paper is still being completed, a number of avenues for further work cannot be reported fully at present. Among these are attempts to get closer to a treatment of the producer subsidy equivalent rather than the surrogates used here. Another is to revisit the question of weighting in the simultaneous autoregressive context, or some form of detrending prior to modelling; this may extend to alternative GWR formulations. Finally, while the paper in its present form in large part reproduces Bivand and Brunstad (2003), it is now cast in a reproducible research form, so that other methods can be used directly, because all the figures and tables are now generated by pre-processing the paper in R.

# References


Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L, Bera AK (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Marcel Dekker, New York, pp. 237-289

Anselin L, Bera AK, Florax R, Yoon MJ (1996) Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26: 77-104

Arbia G (2001) The role of spatial effects in the empirical analysis of regional concentration. Journal of Geographical Systems 3: 271-281

Baumont C, Ertur C, Le Gallo J (2001) A spatial econometric analysis of geographic spillovers and growth for European regions, 1980-1995. (Unpublished working paper 2001-04, LATEC UMR, University of Burgundy)

Bivand R, 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, pp. 351–373

Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Analysis 28: 281-298

Brunstad RJ, Gaasland I, Vårdal E (1999) Agricultural production and the optimal level of landscape preservation. Land Economics 75: 538-546

Cahill C, Legg W (1989) Estimation of agricultural assistance using producer and consumer subsidy equivalents: theory and practice. OECD Economic Studies 13: 13-43

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74, 829-836.

Cliff AD, Ord JK (1981) Spatial processes. Pion, London

Colman D (2001) The Common Agricultural Policy. In: Artis M, Nixson F (eds) The economics of the European Union. Oxford University Press, Oxford, pp. 97-124

DG REGIO (2001a) Study on the impact of community agricultural policies on economic and social cohesion. (http://www.inforegio.cec.eu.int/wbdoc/docgener/-studies/pac_en.htm, accessed 2 Jan 2002)

DG REGIO (2001b) Unity, solidarity, diversity for Europe, its people and its territory: Second report on economic and social cohesion. (http://www.inforegio.-cec.eu.int/wbdoc/docoffic/official/reports/contentpdf_en.htm, accessed 2 Jan 2002)

Fagerberg J, Verspagen B, Caniels M (1997) Technology, growth and unemployment across European regions. Regional Studies 31: 457-466

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

Fotheringham AS, Brunsdon C, Charlton ME (2000) Quantitative geography: perspectives on spatial data analysis. Sage, London

Heckelei T, Britz W (2000) Concept and explorative application of an EU-wide, regional agricultural sector model (CAPRI project). (http://www.agp.uni-bonn.de/agpo/rsrch/capri/eaaecapri.pdf, accessed 2 Jan 2002)

Johnston J, DiNardo J (1997) Econometric methods. McGraw-Hill, New York

Le Gallo J, Ertur C (2000) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. (Unpublished working paper 2000-09, LATEC UMR, University of Burgundy)

López-Bazo E, Vayá E, Mora AJ, Suriñach J (1999) Regional economic dynamics and convergence in the European Union. Annals of Regional Science 33: 343-370

OECD (2000) Eurostat's EAA97: Main changes and the impact on the EAA-data. OECD Statistics Directorate STD/NA/AGR(2000)10 (http://www1.-oecd.org/std/AGRMeet2000/docs/naa00_10.pdf, accessed 2 Jan 2002)

Paci R (1997) More similar and less equal: economic growth in the European regions. Weltwirtschaftliches Archiv 133: 609-634

Paci R, Pigliaru F (1999) Is dualism still a source of convergence in Europe? Applied Economics 31: 1423-1436

Pons-Novell J, Viladecans-Marsal E (1999) Kaldor's Laws and spatial dependence: evidence for the European regions. Regional Studies 33: 443-451

Rey SJ (2001) Spatial empirics for economic growth and convergence. Geographical Analysis 33: 195-214

Rey SJ, Montouri BD (1999) US regional income convergence: a spatial econometric perspective. Regional Studies 33: 143-156

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`

Tiefelsdorf M (2002) The Saddlepoint approximation of Moran's $I$ and local Moran's $I_i$ reference distributions and their numerical evaluation. Geographical Analysis 34: 187-206

Tondl G (2001) Regional policy. In: Artis M, Nixson F (eds) The economics of the European Union. Oxford University Press, Oxford, pp. 180-212

Vayá E, López-Bazo E, Artís M (1998) Growth, convergence and (why not?) regional externalities. (Unpublished working paper E98/31, "Anàlisi Quantitativa Regional" Research Group, University of Barcelona)