Discussion Paper No. 02-49

Explaining the Level of Relative Investment Specialisation: A Spatial Econometric Analysis of EU Regions

Claudia Stirböck

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 02-49

Explaining the Level of Relative Investment Specialisation: A Spatial Econometric Analysis of EU Regions

Claudia Stirböck

Non-technical summary

The establishment of the EMU was accompanied by a broad discussion of potential core-periphery tendencies and asymmetric regional developments due to possible stronger sectoral concentration and regional specialisation. Up to now, we have no clear indication about the level of regional specialisation in the EU, specifically not about the specialisation tendencies to be expected due to increased factor mobility and market integration. In an analysis of the determinants of the level of relative investment specialisation in EU regions, Stirboeck (2002) provided evidence on the importance of regional size, gross domestic product, population density, the number of patents, economic openness, capital market integration, and the peripheral or central location of the region in the explanation of the even or uneven sectoral allocation of gross fixed capital formation. In this study, the sensitivity of the results in Stirboeck (2002) is tested with respect to the influence of differing formulations of specialisation measures. In addition, the spatial pattern of the data and the potential spatial interdependence between EU regions are taken into account. Besides classical econometric techniques, spatial econometric procedures are thus applied.

In order to test the robustness of the potential impacts, a Gini-coefficient, a coefficient of variation, as well as a Finger-Kreinin-index, are calculated to measure the level of relative regional specialisation of 56 NUTS 2- and 33 NUTS 1-regions. The respective sectoral allocation is measured in relation to the average EU investment structure. We consistently get the same results for all three indicators in GLS estimates, estimates in logit terms, instrumental variable estimates to control for potential effects of reverse causation, and a dynamic specification capturing possible first-order correlation effects. We find a bigger market size as well as a larger size of the region to reduce the level of regional specialisation. A higher unemployment rate, increasing economic openness, higher population density, the fact of being a central region, and the distance from this central region, instead, increase regional specialisation.

Independent of the specification of the spatial dependence, the spatial econometric regressions display negative spatial dependence or correlation. The OLS test diagnostics on spatial dependence point to a negative spatial error correlation, what is in line with the information criteria which indicate a better performance of the spatial error model compared to the spatial lag model. This negative spatial correlation of the error terms might simply be a result of measurement errors. Data inconsistencies, the regional databases' shortcomings as well as incompatibilities of the units of observation with actual economic regions may be one reason for such spatial nuisances in the data. Further improvements of the spatial econometric estimates can be obtained by adding spatially lagged external variables. This, however, leads to rather inconsistent results, largely depending on the specific model and spatial weights matrix chosen. However, independent of the spatial dependence model accounting for regional interactions and spatial correlation in the regional data, we can confirm the determinants of relative regional specialisation identified with the classical econometric methods.

Explaining the Level of Relative Investment Specialisation:

A Spatial Econometric Analysis of EU Regions

Claudia Stirboeck

Centre for European Economic Research (ZEW), Mannheim, Germany

July 2002

Abstract

This paper analyses the level of relative specialisation in terms of gross fixed capital formation in EU regions. Larger market and regional sizes diminish; a higher unemployment rate, population density, the fact of being a central region, the distance to the economic centre, and economic liberalisation increase the level of specialisation. These results are not sensitive to differing formulations of the specialisation indicator. Accounting for spatial dependence by the use of spatial econometric tools, negative spatial interactions probably due to data inconsistencies are present. However, the results of classical econometric estimates are robust.

Acknowledgements: Financial support by the Volkswagen-Stiftung within the project No. II/76547 is gratefully acknowledged. I would like to thank Laura Manthey, Paula Montoya, Jennifer Säwe, Elena Todorova, and especially Simone Giloth and Felix Morck for their enduring help in setting up the extensive regional data sets. I am indebted to Herbert S. Buscher, Jürgen Kaehler and Michael Schröder for helpful comments. All remaining errors, however, are my own.

JEL classification: C30, F15, F2, R12

Keywords: Economic Geography, Regional Economics, Capital Allocation, Spatial Econometric Analysis

Claudia Stirboeck
ZEW
PO Box 10 34 43
D-68034 Mannheim
Tel.: +49-621-1235-218
Fax: +49-621-1235-223
E-Mail: stirboeck@zew.de

I Motivation

The establishment of the EMU was accompanied by a broad discussion of potential core-periphery tendencies and asymmetric regional developments due to possible stronger sectoral concentration and regional specialisation. Sectoral concentration or regional specialisation enables the exploitation of industry-level economies of scale and knowledge spillovers. However, the absorption of asymmetric shocks or growth imbalances by national monetary policies is reduced. Up to date, we have no clear indication about the level of regional specialisation in the EU and especially not about the specialisation tendencies to be expected due to increased factor mobility and market integration.

In traditional trade theory, regional specialisation across sectors is assumed to be in accordance with comparative advantages. Market integration increases regional specialisation in line with trade expansion. Agglomeration tendencies such as a high density of population, capital or economic activity in only one regional area and a disequilibrium in economic developments are, however, not to be expected. Gravity models in international economics (Tinbergen, 1962; Linnemann, 1966) explain economic flows between regions through gravitational and resistance forces such as market size or market potential, distance, barriers to international activity etc. The spatial concentration of e.g. investments can thus be the result of gravitational forces which become stronger as soon as resistance forces, such as transport costs or imperfect integration, weaken.

The new economic geography has sharply increased the importance of regional economic theory in the 1990s. It has induced a new wave of attention to concentration and specialisation patterns. However, already before the 1990s, polarisation theories in the framework of regional economics have provided explanations for circular and cumulative agglomeration tendencies due to “forward and backward linkages” (Hirschman, 1958) or “backwash-effects” (Myrdal, 1957) which are unfortunate for peripheral regions. Since Krugman (1991), the new economic geography has gained a special focus of attention as according to these models, specialisation need not – like in the neo-classical world – develop according to the comparative advantage of regions, but can be the result of historical conditions and macroeconomic (partly random) processes. Thus, even similar regions can develop differently and the resulting patterns of specialisation are ex ante unpredictable. Due to the existence of economies of scale at the plant level (further increased by economies of localisation at the industrial level), firms do not produce at each single place of local demand. Instead, the production of each differentiated good is locally concentrated and close to large markets. The core thus specialises in scale-intensive economic activity, the periphery in agriculture or industries with constant or decreasing economies of scale. Transport costs are a centrifugal force working against the concentration of production. In case of high transport costs, the allocation of production over space is rather persistent. Decreasing transport costs, though, strengthen the centripetal force of economies of scale and might thus be a trigger for the concentration of production.

A number of empirical studies on sectoral agglomeration tendencies as well as regional specialisation have emerged in the last years. Though, only few provide econometric evidence on the determinants of sectoral concentration or regional specialisation. In addition, the regional (not national) focus as well as the analysis of capital (not only production, trade or employment) patterns, is neglected. An overview on recent descriptive and econometric studies on the named topics is given by Stirboeck (2001, 2002).

In an analysis of the determinants of the level of relative investment specialisation, Stirboeck (2002) provided evidence on the importance of regional size, gross domestic product (GDP), population density, the number of patents, economic openness, capital market integration, and the peripheral or central location of the region in the explanation of the even or uneven sectoral allocation of gross fixed capital formation. These analyses are now extended in this study testing for the sensitivity of the results with respect to the influence of different specialisation measures. In addition the spatial pattern of the data and potential spatial interdependence is accounted for. Besides classical econometric techniques, spatial econometric procedures are thus applied in the following.

II Determinants of relative regional specialisation levels

Based on the above referred different theoretical approaches, the level of regional specialisation in terms of gross fixed capital formation is to be explained by a large number of economic variables. However, explanatory variables added in this analysis are to some extent limited by data availability. Including the core variables mentioned above, the following specification was tested in Stirboeck (2002)[1] to explain the level of relative regional specialisation (LEVSPEC):

$$LEVSPEC_j = \beta_0 + \beta_1 MAR_j + \beta_2 AREA_j + \beta_3 PODEN_j + \beta_4 UEWP_j + \beta_5 DIST_j + \beta_6 INT_j + \beta_7 CENTR_j$$

The market size (MAR) of region j is approximated by gross domestic product (GDP). Additional important exogenous variables are the size of a region (AREA), population density (PODEN), unemployment rate in percent of working population (UEWP) which are all taken from the REGIO database. The distance of the region to the economic centre (DIST) capturing peripherality effects, an index of economic openness (QUINN_OPENN) reflecting market integration (INT)[2] as well as an indicator variable for central, economically most important regions (CENTR) are added. Details on all these variables are given in the appendix A. We now extend the analysis to test for the sensitivity of the results with respect to different estimation techniques and regional specialisation indicators.

In order to abstract from size and classification effects (i.e. the differing importance of sectors and the possibly inadequate disaggregation of economic activity in subsectors) in the calculation of regional specialisation, we refer to relative specialisation measures, i.e. specialisation in relation to an economy of reference. This is important since the absolute allocation of economic activity across sectors does not tell us anything about a particularly high level of sectoral engagement of a region while this is what we focus on: relative allocation and hence, relative specialisation. It is the unequal size of regions or sectors which generally causes the difference between the absolute and the relative specialisation index[3].

Relative regional specialisation in the EU can, of course, be analysed applying two different perspectives. First, it is possible to investigate the regional investment structure in relation to the national one which would be a national perspective. Second, the regional investment structure can be compared to the average EU structure, what means adopting a European per-

[1] We exclude the number of patents in the following estimates as regional data on this variables are only available since 1989.

[2] In Stirboeck (2002) an indicator of capital market integration was additionally tested and confirmed the results for the economic openness indicator.

[3] While measures of absolute allocation are influenced by regional size and sectoral classification, measures of relative allocation are influenced by the sectoral patterns of either the economy of reference or the average pattern of the group of countries included. In case of a very special pattern of the reference economy, the relative specialisation pattern of the economic entities analysed can be biased. Further details on the construction of different relative and absolute concentration and specialisation indices can be found in Stirboeck (2001) as well as Krieger-Boden (1999).

spective. Both perspectives lead to slightly different specialisation patterns. These differing perspectives, however, appear to influence the regression results only to a negligible extent (Stirboeck, 2002). In this analysis, we thus focus on the European perspective. Up to 17 differentiated sectors – consistent to the industrial classification of Nace Rev. 1 - Nomenclature des activités économiques dans les Communautés Européennes - are available in the REGIO database and are included in this study.

The regional economic literature suggests a number of measures of regional specialisation. The studies of Krugman (1991), Brülhart (1998), Klüver and Rübel (1998), and Amiti (1999) refer to a Gini-coefficient. However, Sapir (1996) analysing absolute country specialisation with export data made use of the Herfindahl index instead, Greenaway and Hine (1991) as well as Kalemli-Ozcan, Sorensen and Yosha (1999) apply the Finger-Kreinin index.

In order to test the robustness of the explanatory variables, three different formulations of relative regional specialisation levels are included in the classical econometric analysis:

1. A coefficient of variation[4] (VCCFEU) is calculated as follows:

$$c = \frac{\sqrt{\frac{1}{n}\sum(\mathrm{B}_{\mathrm{sr}} - \overline{B_{sr}})^2}}{\overline{B_{sr}}}$$

It thus measures the variation of the relative investment shares B_{sr}[5] and captures their degree of homogeneity. This coefficient ranges in the interval $[0, \sqrt{n-1}]$. It is standardised by dividing by $\frac{1}{\sqrt{n-1}}$ to lie in the [0,1] interval.

2. A Gini-coefficient[6] (GCCFEU) of the region in focus is calculated as suggested by Krugman (1991) and applied in most recent empirical work. The cumulative sums of sectoral

---

[4] The coefficient of variation stresses changes at the outer sides of the distribution of relative sectoral shares which contrasts to the weighting of the Gini-coefficient described in footnote 5.

[5] Relative investment indices have been constructed by dividing the sectoral investment share of the respective region s_{ij} by the average investment share of the sector in the reference economy r_i:

$$B_{ij} = \tfrac{s_{ij}}{r_i} = (x_{ij} / \sum_i x_{ij}) / (\sum_j x_{ij} / \sum_i \sum_j x_{ij})$$

with i (j) as the sectoral (regional) index. As a result, this adapted „Balassa-index“ reflects the relative investment performance of a region in a sector. If the region’s investment in one sector is relatively strong (low) compared to the other regions, the index is higher (smaller) than 1.

As national GFCF data are not in all cases as complete as we wish them to be, we had to use adequate but different data representing the economic extent or importance of the different sectors in any country or region to calculate sectoral specialisation indices. Therefore it is referred to data of gross value added at factor costs when calculating EU sectoral shares. Eurostat (2000b) similarly uses the regional contributions of national gross value added as distributional weights when dividing the national values of GDP among the regions.

[6] The Gini-coefficient gives strong weight to the middle parts of the distribution of relative sectoral shares. As a consequence, changes in industrial sectors with relative shares similar to the median structure have a larger effect on the value of the Gini-coefficient than changes in industrial sectors at the outer sides of the distribution (Cowell, 1995). However, the coefficient’s range between 0 (low concentration) and (N-1)/N (high concentration) usually reflects well differences in the level of concentration. Therefore, the Gini-coefficient is

shares of the given regions are thus plotted against those of the reference economy ranked according to the Balassa-indices. Both sectoral structures are thus compared. In case of equal sectoral structures, we get a Gini-coefficient of zero representing a low level of relative investment specialisation. The Gini-coefficient ranges between 0 and (N-1)/N. The standardised Gini-coefficient G*N/(N-1) – ranging between 0 and 1 – is referred to as the Lorenz-Münzner-coefficient.

The outcome of the Gini-coefficient is slightly different from the one of the coefficient of variation. Both sectoral shares, the one of the respective region as well as the one of the reference economy, influence the value of the Gini-coefficient. If a large sectoral share in e.g. the reference economy is confronted with an even larger sectoral share in the region in focus, the value of the Gini-coefficient is largely determined by this economically important sector. This is the case for a number of regions.

3. The Finger-Kreinin-index (FKCFEU) is a bilateral comparison of sectoral shares of the region in focus and the average EU pattern. It is defined as:

$$FK_j = \sum_i^n \min(s_{ij}, r_i)$$

with s_{ij} as the share of region j and r_i as the share of the reference economy in sector i.

The first two indices are standardised to range between 0 (low specialisation) and 1 (high specialisation) while The Finger-Kreinin-index per definition lies in the interval [0,1]. Its interpretation, though, is inverse to the other two indices with 0 as perfect specialisation and 1 reflecting sectors of equal relative importance. A graphical comparison of all three indicators shows a similar development over time, the main difference is a generally lower value of the coefficient of variation which, however, does not influence its course. We therefore do not expect sharply differing results when using these alternative measures.

The estimates for the NUTS 2-level are displayed in Table 1, those for the higher aggregated NUTS 1-regions in the appendix in Table B1. For all three indicators measuring the regional level of specialisation, the estimates, indeed, mostly demonstrate the same significant determinants. We consistently get the same impact for the significant variables, i.e. an opposite sign of the explanatory variables in the analysis of the Finger-Kreinin index compared to the analyses of the coefficient of variation as well as the Gini-coefficient.

As the indicators are restricted to lie in the range between 0 and 1, we are possibly confronted with an econometric modelling problem in a simple linear specification. The specification was therefore additionally formulated in logit terms, i.e. a logistic transformation of the estimation model: z = ln [y/(1-y)]. This leads to the same significant determinants (see Table 1 and B1), since none of our specialisation indicators lies near the extremes of the given range.

We find a bigger market size (except for the Finger-Kreinin index at NUTS 2-level) as well as a larger size of the region to reduce the level of regional specialisation. A higher unemployment rate, increasing economic openness[7], the fact of being a central region as well as the distance from this central region, instead, increase regional specialisation. The population density is not consistently significant in the estimates. When running separate estimates in-

the most widely used inequality measure in the analysis of the spatial allocation of sectors or sectoral allocation of regional economic activity.

[7] This variable is not significant for the Gini-coefficient at NUTS 2-level. However, for the other two coefficients as well as at NUTS 1-level, we get a significant and positive impact on the level of specialisation.

cluding either population density or the central region dummy as an explanatory variable at NUTS 2-level, we find a significant increasing impact on specialisation of both variables which is in line with the results at NUTS 1-level.

The results of the estimates at NUTS 1-level are mostly comparable to those at NUTS 2-level. Only the variables DIST as well as AREA are not significant for two of the three indicators. In addition, the unemployment rate is insignificant for the Finger-Kreinin-index. For the respectively significant coefficients, we get consistent signs. The insignificance of the regional size variable at NUTS 1-level might simply be due to the fact that NUTS 1-regions are less varying in their size than NUTS 2-regions.

Table 1: Estimation Results for the Determinants of Relative Regional Investment Specialisation, NUTS 2-level, 1985-1994

	VCCFEU	GCCFEU	FKCFEU	VCCFEU	GCCFEU	FKCFEU
	level	level	level	logit	logit	logit
Constant	0.1083	0.3322	0.8012	-1.9003	-0.7137	1.3482
	2.86	10.80	25.70	-10.69	-5.24	9.41
GDP	-0.0007	-0.0004	-0.00005	-0.0037	-0.0019	-0.0001
	-4.96	-4.00	-0.24	-5.8	-3.95	-0.14
CENTR	0.1656	0.1654	-0.1277	0.9647	0.7001	-0.5850
	5.75	7.06	-7.16	7.13	6.76	-7.14
UEWP	0.0048	0.0065	-0.0061	0.0272	0.0266	-0.0280
	4.68	7.78	-8.02	5.67	7.25	-7.96
PODEN	0.0044	0.0015	0.1043	-0.0269	0.0144	0.4456
	0.63	0.26	1.89	-0.82	0.57	1.76
AREA	-0.0012	-0.0015	0.0018	-0.0074	-0.0060	0.0085
	-2.25	-3.28	4.18	-2.88	-3.03	4.17
QUINN_OPENN	0.0055	0.0026	-0.0075	0.0288	0.0123	-0.0361
	2.01	1.18	-3.30	2.26	1.26	-3.47
DIST	0.0598	0.0678	-0.0463	0.3583	0.2881	-0.2187
	3.08	4.30	-3.81	3.93	4.13	-3.91
DUM_FRA	0.0056	-0.0295	0.0292	0.0026	-0.1202	0.1308
	0.51	-3.33	4.35	0.05	-3.07	4.25
DUM_BEL	0.1596	0.1141	--	0.8008	0.4872	--
	10.94	9.63		11.70	9.29	
DUM_IRE	0.0537	0.0003	-0.3303	0.1737	-0.0139	-1.5011
	0.99	0.01	-9.77	0.68	-0.07	-9.66
DUM_LUX	-0.1065	-0.1939	-0.0541	-0.5969	-0.8364	-0.2202
	-2.49	-5.58	-2.05	-2.97	-5.44	-1.82
DUM_DEN	0.0300	-0.1177	-0.3818	0.1420	-0.4982	-1.6813
	0.69	-3.35	-14.11	0.70	-3.21	-13.52
no. obs.	487	487	377	487	487	377
SSR	3.6470	2.4063	1.0514	80.2778	47.1407	22.2063
Log Likelihood	500.75	602.00	573.84	-252.05	-122.42	-274.34
Prob Chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AIC	-2.003	-2.419	-2.975	1.088	0.556	0.075

Note: Lines below the coefficients report the z-values of the GLS estimates. SSR displays the sum of squared residuals. The probability of the Chi²-test gives the overall fit of the model. Since sectoral sums for gross fixed capital formation are not available for Belgian regions, we cannot calculate Finger-Kreinin indices for the 110 NUTS 2-Belgian regions.

In the specification used so far, we cannot exclude a potential problem of reverse causation between the level of capital concentration and regional gross domestic product or the regional rate of unemployment[8]. In order to control for such problems of endogeneity, instrumental variable regressions have been run additionally by the use of lagged values of the unemployment rate as well as of gross domestic product. Results presented in Table B2 are very similar, since most coefficients are nearly identical[9].

In order to take account of potential time correlation effects, a simple dynamic specification capturing first-order serial correlation was tested additionally. Results are also presented in Table B2. Besides the size of the region (at NUTS 2-level) and the distance to the centre, the impact of the explanatory determinants on regional investment specialisation patterns is mostly confirmed[10].

Our results are thus neither sensitive to the estimation method nor to the formulation of the indicator of regional specialisation. When controlling for spatial interdependencies in the following, we restrict our analysis of potential spatial autocorrelation to the coefficient of variation.

III Spatial econometric estimates

Regional economic developments and specialisation tendencies potentially underlie spatial dependence or interaction. If economic events in neighbouring regions are not independent, but influence each other, there is spatial dependence. In addition, regional data might have shortcomings such as a bad quality due to measurement problems or inadequately defined regional units, what is reflected in spatial autocorrelation. Standard regressions do not account for spatial dependence or autocorrelation thus leading to inefficient inference or even biased estimates in case of significant spatial processes. In addition to the use of classical econometric methods presented so far, we therefore refer to models of spatial econometrics in this section, which explicitly take account of spatial interaction (see e.g. Anselin, 1988). The structure of spatial interconnectedness is usually imposed by so-called spatial weights matrices (W). Wy, e.g., thus displays the spatially weighted average of y in nearby regions. A number of different spatial econometric models – as well as combinations of those – can be formulated: spatial correlation of the error term in e.g. a spatial autoregressive error model, of the endogenous variable itself in a spatial lag model as well as of explanatory variables in a spatial cross-regressive model[11].

[8] It is possible that the level of regional GDP (UEWP) increases (decreases) due to a strong specialisation in a sector with high growth-potential. On the other hand, a strong specialisation in the "wrong" sectors can also weaken GDP or augment UEWP. We thus see no obvious and clear reverse causation.

[9] We restricted these robustness tests to the Gini-coefficient and the coefficient of variation since only these two include all observations (resp. all regions).

[10] When excluding the British observations from the regressions at NUTS 1-level, we can also confirm the impact of the distance to the centre. Since we generally have information on only few sectors for the British regions, the British specialisation indices are less reliable than the other indices.

[11] In addition to spatial error models with a spatial autoregressive error term, the disturbance term can also follow a spatial moving-average process. For a discussion of different first or higher order spatial processes combining spatial autoregressive dependent variables or error terms, spatial moving-average error terms as well as spatially lagged external variables in such called "SARMA"- or "SARMAX"-models, see e.g. Anselin/Bera (1998: 251f).

In a **spatial autoregressive error model**, λ[12] captures the spatial autoregression of the error term ε while u is the independently and normally distributed error term with constant variance:

$$Y = X\beta + \varepsilon , \qquad \varepsilon = \lambda W\varepsilon + u = (I - \lambda W)^{-1}u, \qquad u \sim N(0, \sigma^2_u I). \qquad (M1)$$

The spatial autoregressive parameter λ thus displays the strength of correlation between the disturbance term ε and the weighted average of the disturbance terms of neighbouring regions $W\varepsilon$.

In a **spatial lag model**, ρ is the spatial autoregressive parameter which measures the reaction of Y to economic developments in surrounding areas, i.e. spatial spillovers or the influence exerted by a change in the specialisation level of a neighbouring region on the level of specialisation in region j:

$$Y = \rho WY + X\beta + v , \qquad v \sim N(0, \sigma^2_u I) . \qquad (M2)$$

Besides spatially interdependent endogenous variables, spatially interdependent economic events might affect neighbouring regions through spatially lagged exogenous variables, thus giving rise to a **“ spatial cross-regressive model”** with spatial regressive explanatory variables:

$$Y = X_1\beta_1 + WX_2\beta_2 + w , \qquad w \sim N(0, \sigma^2_u I) . \qquad (M3)$$

A model specification with spatially lagged explanatory variables only can be estimated by simple OLS estimates[13]. However, to prevent inefficient or even biased estimates in case of spatial lag or error dependence, we have to refer to different estimation methods. A standard technique to deal with spatial lag dependence or spatial autoregressive error terms is to conduct Maximum-Likelihood (ML) estimates[14].

The standard software (SpaceStat1.90) for this kind of analysis does not capture time-space-models[15]. However, as explained above, the simple dynamic relationship (presented in Table B2) provides evidence for significant serial first-order correlation. But this dynamic specification also confirms our results for the main determinants of the level of regional specialisation. Due to this, we might now limit our estimations to pure space models. Since the tests on spatial dependence require a normal distribution in the errors of the models estimated, we eliminate outlying observations from the datasets so that the hypothesis of the non-normality

[12] In spatial processes, the spatial autoregressive parameters are not restricted to the usual interval (–1,+1). The parameter space is instead restricted by $1/\omega_{min}$ and $1/\omega_{max}$ with ω_{min} and ω_{max} as the smallest and largest eigenvalues of the spatial weights matrix implemented in the regression. In case of row-standardised weights matrices ω_{max} =1, but ω_{min} > -1. Thus, the spatial autoregressive parameters of the spatial autoregressive error model as well as the spatial lag model can be smaller than –1 (Anselin, 2001: 321).

[13] See e.g. Haining (1990: 344-50). A problem of multicollinearity (between X and WX), however, arises in case of spatially autocorrelated external variables. In this case, estimated parameters have to be interpreted carefully.

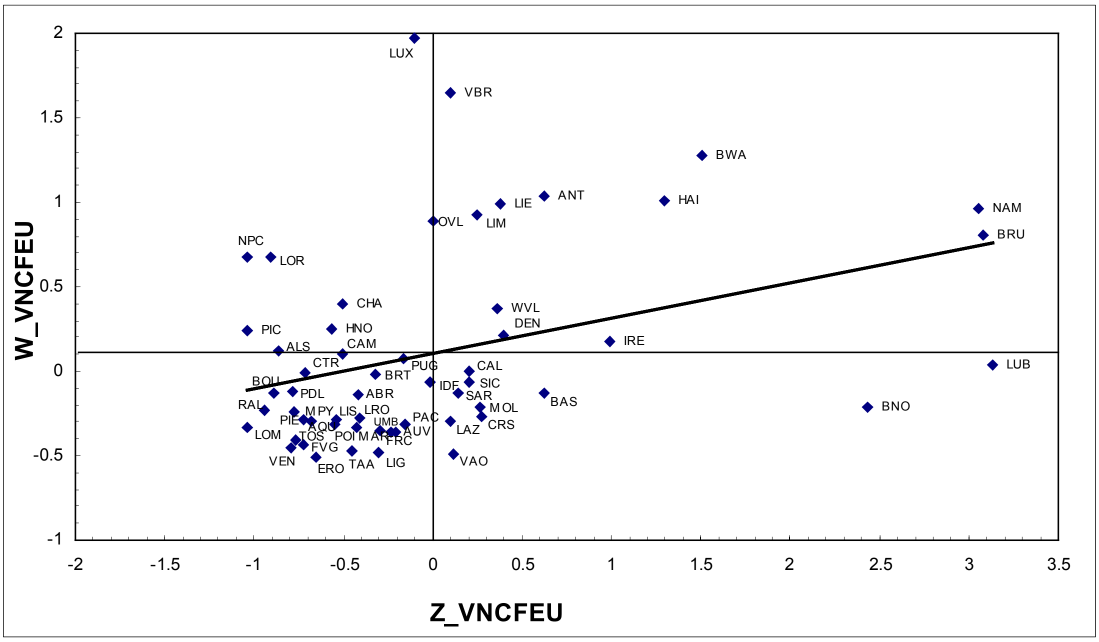
[14] While OLS provides biased estimates in case of spatial lag dependence, it leads to unbiased, but inefficient estimates in case of spatial autocorrelation of the error terms. Since the autocorrelation parameter λ is unknown, we cannot simply conduct weighted least squares estimates, however, and have to refer to maximum-likelihood estimates as well. For further details on this topic, see e.g. Anselin (1988, 1999a).

[15] The implementation of simultaneous time-space effects would have high computational costs and be rather complicated. In the past, different solutions to this problem have been suggested. Schulze (1982) e.g., first, eliminates serial correlation and, second, models the spatial dependence what he refers to as a “four-step-Aitken procedure”.

of errors can be rejected for each model presented in the following, i.e. we can assume a normal distribution of the error terms[16].

In order to prevent that our findings are due to the formulation of spatial dependence imposed by the spatial weights matrix, we control the sensitivity of our results with respect to different weights matrices[17]. First, we include two inverse distance matrices and, second, a neighbourhood contiguity matrix. The distance matrices are based on Euclidean distances between administrative centres of the regions as well as between regional centres as provided by the ArcView software. We use the squared inverse of both distance matrices which thus reflects a decreasing intensity of influence of nearby regions with increasing distance[18]. In addition, we use a neighbourhood contiguity matrix, with the element $w_{ij} = 1$ in case of a common border of the regions i and j, and 0 otherwise, while the diagonal is set to 0.

Graph 1: Moran Scatterplot for NUTS2-regions with bivariate regression line, average level of relative specialisation for 1985-94, squared inverse distance between regional capitals



Note: Z_VNCFEU displays the deviation of the different VNCFEU from their mean while W_VNCFEU represents the spatially weighted average of the neighbouring values.

16 Tests on normality of errors are based on skewness tests as well as Jarque-Bera tests – both are applied in such a way that the non-normality of errors cannot even be assumed at the 10% level of significance. By this, we can suppose a normal distribution of the residuals.

17 However, a common procedure is also to test a variety of slightly differing distance matrices in order to find the spatial weights matrix that best fits the underlying process of spatial dependence like e.g. in Molho (1995) or Niebuhr (2001).

18 We additionally tested for the potential influence of inverse distance matrices with constant influence. In many cases, we got similar results, though the estimates' fit was generally not as good as the one for the preferred weights matrices capturing a decreasing strength of the influence of neighbourhood economic activity.

The Moran scatterplot was introduced by Anselin (1995)[19] and is used to visualise the patterns of spatial association between neighbouring regions. It thus gives a description of the spatial distribution of the variable observed. The Moran scatterplot in Graph 1 displays the spatial association between the 56 regions (Anselin, 1996) with respect to their average level of specialisation and the spatially weighted average of the neighbouring values (W_VNCFEU). The levels of specialisation are taken as deviations from their means (Z_VNCFEU), the scatterplot is thus centred around [0,0]. By this, different scatterplots are comparable. In the upper right and the lower left quadrant, those regions are displayed which are surrounded by similarly specialised regions and are thus marked by positive spatial association. Regions with dissembling neighbours are located in the upper left (regions with low specialisation surrounded by highly specialised regions) and the lower right quadrants (vice versa). Using the weights matrix of the squared inverse distances of regional capitals we find four outlying regions[20]: Brussels (BRU), Namur (NAM), Luxembourg/Belgium (LUB)[21] and Basse-Normandie (BNO) all with a strongly uneven allocation of relative investment shares. The two former, however, are surrounded by similarly specialised regions while the two latter are surrounded by dissimilar regions. The degree of linear association between the vectors Z_VNCFEU and W_VNCFEU is displayed by the linear regression line superimposed in Graph 1. The linear association between the average specialisation levels of the 56 regions and the spatially weighted average of the neighbouring regions' specialisation levels is thus positive.

In addition to the visualisation of the linear association by use of the bivariate regression line in a Moran scatterplot, its degree, i.e. the slope of the bivariate regression line, is also formally indicated by the Moran I statistic. Moran I is defined as $I=y'Wy/y'y$ with W as the row-standardised weights matrix (Anselin, 1995: 105)[22]. The Moran I coefficient is centred around its theoretical expected mean which is $-1/(N-1)$. Values larger than its expected mean display positive spatial autocorrelation. Referring to all 442 observations[23] of the 56 regions we analyse, and not the regional averages like in the graph above, this expected mean amounts to -0.0023. For the two squared inverse distance matrices, the different significant Moran I-values are 0.2477 (distance between regional capitals) and 0.2445 (distance between regional centres), and 0.2452 for the neighbourhood-contiguity matrix – at a 1%-level of significance respectively. The Moran coefficient thus points to a significant positive spatial autocorrelation of the level of specialisation, i.e. regions with similar levels of specialisation are more spatially clustered than in the case of random patterns. In other words, regions with a high (low) level of specialisation are likely to be surrounded by highly (low) specialised regions.

However, from this kind of analysis, we only get information about spatial associations, i.e. the spatial clustering of similar or dissimilar regions. Evidence on spatial dependencies or even causal interactions have to be derived from spatial regression analyses.

[19] The exploratory spatial data analysis tools rely on the methods of exploratory data analysis following e.g. Tukey (1977).

[20] In the standardised Moran scatterplot, those values further than two units away from the origin are ususally treated as outliers according to the two-sigma rule (Anselin, 1995: 45).

[21] However, we have to note than in contrast to the Italian and French regions, specialisation indices for Belgian regions only base on 11 of the 17 sectors.

[22] For further details on the Moran I coefficient see Anselin (1996: 115ff) and Anselin (1992: 132f).

[23] We now refer to the 442 observations of the restricted dataset which lead to normally distributed residuals in the OLS estimates of the given specification.

III.1 Spatial lag and error dependence models

Table 2 displays the diagnostics on the potential spatial structure to be found in the disturbance terms of simple OLS estimates. Evidence on spatial dependence in our analysis of relative regional specialisation levels is provided by a number of tests. The Moran I test[24] is a general test on spatial correlation without giving precise information on the particular spatial structure. It provides evidence for a negative spatial dependence as the two significant Moran I values are smaller than its expected value of –0.0023. In addition to the Moran I test, we refer to two tests basing on the Lagrange-Multiplier (LM) principle which both test for a specific form of spatial dependence. The LM-error test, suggested by Burridge (1980), tests the null hypotheses of no spatial correlation of the residuals: H_0: $\lambda = 0$. The LM-lag-test, introduced by Anselin (1988), tests the one of absence of spatial dependence of the endogenous variable: H_0: $\rho = 0$ [25]. In all cases, the Lagrange Multiplier (LM)-error test has a higher value than the LM-lag-test thus pointing to a model of spatial error dependence rather than of spatial lag dependence[26].

The significant spatial structure in the residuals thus provides evidence that the GLS estimates presented above suffer from a misspecification. Our aim is now to investigate these effects in more detail in order to check for robustness of the determinants and efficiency of the coefficient tests identified by use of classical econometric techniques.

Table 2: Diagnostics on Spatial Dependence, OLS estimates

	ID_K2		ID_Z2		NGH_CM	
Moran's I (error)	-9.6033	***	-9.4960	***	-7.9972	***
Lagrange Multiplier (error)	45.4157	***	42.9428	***	43.2672	***
Lagrange Multiplier (lag)	20.8389	***	17.6116	***	18.3823	***

Note: Spatial weights matrices are defined as follows: ID_K2: squared inverse of distance between regional capitals; ID_Z2: squared inverse of distance between regional centres; NGH_CM: neighbourhood contiguity matrix with the element $w_{jk} = 1$ in case of a common border of the regions j and k, and 0 otherwise.

With respect to the spatial lag specification presented in Table 3, we get a significant spatial lag (W_VNCFEU) with ρ differing between –0.15 and –0.34 for the three weights matrices. A higher level of specialisation in one region significantly reduces specialisation in the neighbouring regions. Therefore, we seem to be confronted with a negative spatial interaction between the levels of regional specialisation of surrounding areas. Regarding the other explanatory variables, the model's results coincide with the above found results. While the LR-tests confirm the spatial lag dependence, the test diagnostics for further spatial error dependence provide evidence for a still existing spatial structure in the error terms.

[24] The same test statistic, we used earlier on to analyse the spatial association of the VNCFEU variable, is now applied to the residuals of the OLS regression. For further details, see Anselin and Bera (1998: 265ff).

[25] For detailed information on both tests, see Anselin and Bera (1998).

[26] In addition, we also pay attention to two further modifications of these spatial LM-tests, the robust LM-lag and LM-error tests which control for a joint significance of both, spatial lag and error dependence. These are developed by Bera and Yoon (1993) and are discussed in detail in Anselin et al. (1996). However, throughout the study they do not provide other information than the "simple" LM-error and LM–lag tests.

Table 3: Maximum-Likelihood Estimates of Spatial Lag Models (442 obs.)

VARIABLE \ weights matrix	ID_K2	z-value	ID_Z2	z-value	NGH_CM	z-value
W_VNCFEU	-0.3360	-3.71	-0.1918	-3.59	-0.1518	-3.48
CONSTANT	0.1551	6.11	0.1292	6.18	0.1286	6.06
GDP	-0.0005	-7.58	-0.0005	-8.01	-0.0005	-8.21
CENTR	0.1252	9.76	0.1293	10.23	0.1227	9.41
UEWP	0.0054	10.90	0.0049	10.60	0.0045	9.85
PODEN	0.0341	10.31	0.0335	10.07	0.0363	10.94
AREA	-0.0010	-3.93	-0.0009	-3.72	-0.0008	-3.54
QUINN_OPENN	0.0054	4.27	0.0053	4.22	0.0052	4.15
DIST	0.0623	7.12	0.0648	7.51	0.0474	4.34
DUM_FRA	-0.0087	-1.78	-0.0108	-2.24	-0.0129	-2.68
DUM_BEL	0.1281	11.06	0.1157	12.44	0.0980	13.92
DUM_IRE	-0.0748	-2.52	-0.0823	-2.80	-0.1157	-3.88
DUM_LUX	-0.0344	-1.76	-0.0442	-2.31	-0.0610	-3.30
DUM_DEN	0.0639	3.26	0.0558	2.93	0.0185	0.92
Breusch-Pagan test	111.6672	***	102.6882	***	98.6269	***
LR-Test on spatial lag dependence	17.3318	***	14.8472	***	15.3765	***
LM-Test on spatial error dependence	7.4510	***	5.2042	**	3.8535	**
Log Likelihood	827.64		826.40		826.66	
AIC	-3.6816		-3.6760		-3.6772	

Note: For spatial weights matrices see Table 2.

Table 4: Maximum-Likelihood Estimates of Spatial Error Models (442 obs.)

VARIABLE \ weights matrix	ID_K2	z-value	ID_Z2	z-value	NGH_CM	z-value
CONSTANT	0.0685	4.48	0.0806	5.05	0.0731	4.49
GDP	-0.0005	-8.72	-0.0005	-8.60	-0.0005	-9.45
CENTR	0.1338	10.93	0.1289	10.03	0.1306	10.06
UEWP	0.0055	19.96	0.0054	17.40	0.0056	16.89
PODEN	0.0460	13.42	0.0367	11.19	0.0339	10.34
AREA	-0.0002	-0.73	-0.0005	-2.38	-0.0005	-2.07
QUINN_OPENN	0.0054	4.64	0.0054	4.48	0.0055	4.54
DIST	0.0635	9.87	0.0597	8.56	0.0637	8.47
DUM_FRA	-0.0153	-4.54	-0.0148	-3.99	-0.0128	-3.72
DUM_BEL	0.0858	17.38	0.0864	16.12	0.0918	17.46
DUM_IRE	-0.1348	-5.05	-0.1118	-4.00	-0.1181	-4.28
DUM_LUX	-0.0335	-2.14	-0.0530	-2.92	-0.0600	-3.44
DUM_DEN	0.0225	1.29	0.0353	1.93	0.0385	2.14
Lambda	-1.8751	-15.05	-1.2889	-11.35	-0.8625	-6.29
Breusch-Pagan test	121.3978	***	102.9594	***	110.3036	***
LR-Test on spatial error dependence	57.8458	***	37.5543	***	35.0633	***
LM-Test on spatial lag dependence	0.1184		2.1527		0.0233	
Log Likelihood	847.90		837.75		836.51	
AIC	-3.7778		-3.7319		-3.7263	

Note: For spatial weights matrices see Table 2.

ML estimates of the spatial autoregressive error specification are presented in Table 4. Like in the spatial lag specification, all explanatory determinants remain significant and do not change their sign. Again, we find a negative spatial correlation which is displayed in the negative spatial autocorrelation coefficients for the error terms[27]. The LR-tests on spatial error

[27] As explained above, spatial correlation coefficients are restricted to the range between $1/\omega_{min}$ and $1/\omega_{max}$. Calculating the smallest eigenvalues for the different weights matrices, we get the following minimal values of the spatial error correlation coefficient: -2.17 using ID_K2 (with ω_{min} = -0.461), -1.477 in case of ID_Z2

correlation confirm the spatial autoregressive error dependence and the tests on further spatial lag dependence are not significant.

Both spatial models (spatial lag as well as spatial autoregressive error model) provide evidence for the same empirical relationships between the level of specialisation and the explanatory variables we identified by the use of classical econometric approaches without taking account of spatial dependence[28]. In addition, both models appear to be better specified than the OLS estimates of the non-spatial model which has an AIC of –3.647[29]. Comparing the two specifications, we consistently get better (i.e. lower) information criteria for the spatial error specification what is in line with the spatial dependence diagnostics for the OLS estimates. However, it is possible that we are confronted with a misspecification of the spatial error model since the additionally conducted two tests on the common factor hypothesis are significant, thus pointing to an "inherent inconsistency" of the spatial error model. This might be caused by a further influence of spatially lagged explanatory variables (see Anselin, 1992: 212). This is worth being analysed in the following section.

III.2 Estimates including spatially lagged external variables

Searching for the optimal specification capturing the underlying spatial dependence, we therefore also tested for possible spatial dependence of spatially lagged external variables. The economic situation of the nearby regions might be of importance for the regional level of specialisation. As explained above, a specification with spatial regressive components, i.e. spatially lagged exogenous variables, can be estimated by use of OLS. Results of this "spatial cross-regressive model" specified by equation M3 are displayed in Table 5. The fit of the model compared to the spatial error model, however, does not convincingly improve[30].

With respect to the spatial regressive components, we can find that they are significant in many cases and consistently show the same sign for different spatial weights matrices. This means that our results do not dependent on the choice of the distance matrix. In most cases, higher unemployment, a bigger market size (gross domestic product), a bigger geographical size as well as a lower population density of the neighbouring regions have a significant increasing influence on the level of regional specialisation. A larger size, a higher gross domestic product, and a lower population density of a region also lower its level of specialisation. Except for the spatially lagged unemployment rate, the analysis of the spatially lagged external variables thus displays the same impact of neighbouring regions as the analysis of the level of specialisation of nearby regions itself. This means that the results of the spatial cross-regressive model in Table 5 focusing on the spatially lagged external variables size, gross domestic product and population density largely confirm the negative spatial lag dependence identified above. The influence of the spatial lag of the unemployment rate, instead, is pointing to a positive spatial autocorrelation.

(with ω_{min} = -0.677), and –1.41 by use of the neighbourhood contiguity matrix (with ω_{min} = -0.711). For our three estimates, the spatial error correlation coefficient is thus within the necessary range.

[28] Population density is insignificant in the estimates at NUTS 2-level with all 487 observations, but significant and positive in the estimates with the restricted dataset of 442 observations and in the spatial dependence models as well as in the estimates at NUTS 1-level.

[29] These values refer to the estimates only including 442 observations.

[30] The fit of the model is slightly better compared to the spatial error specification in case of the weights matrix ID_Z2 according to the AIC.

Table 5: Spatial cross-regressive model (442 obs.)

VARIABLE \ weights matrix	ID_K2	t-value	ID_Z2	t-value	NGH_CM	t-value
Constant	-0.0162	-0.70	0.0060	0.26	0.0880	4.11
GDP	-0.0004	-6.31	-0.0004	-6.64	-0.0005	-8.19
CENTR	0.1219	9.64	0.1224	9.59	0.1283	9.80
UEWP	0.0036	5.47	0.0033	5.24	0.0043	7.06
PODEN	0.0407	11.80	0.0397	11.55	0.0368	10.81
AREA	-0.0010	-4.19	-0.0011	-4.33	-0.0009	-3.86
QUINN_OPENN	0.0046	3.73	0.0046	3.70	0.0052	4.09
DIST	0.0567	6.35	0.0560	6.08	0.0679	6.70
W_UEWP	0.0063	4.25	0.0061	4.38	0.0008	1.11
W_GDP	0.0006	2.38	0.0007	2.64	-0.00005	-0.38
W_PODEN	0.0085	0.90	-0.0094	-2.14	-0.0539	-3.06
W_AREA	0.0028	4.33	0.0019	2.93	0.0007	2.10
DUM_FRA	-0.0166	-3.10	-0.0137	-2.63	-0.0201	-3.61
DUM_BEL	0.1265	11.00	0.1327	13.73	0.1128	11.56
DUM_IRE	-0.0611	-2.08	-0.0577	-1.95	-0.0705	-2.33
DUM_LUX	-0.0390	-2.11	-0.0390	-2.08	-0.0689	-3.64
DUM_DEN	0.0627	3.35	0.0669	3.54	0.0593	2.97
Jarque-Bera	19.3293	***	13.8962	***	7.6363	**
Heteroscedasticity test	105.2997	***	98.4490	***	102.2501	***
Moran's I (error)	-8.8796	***	-4.6882	***	-8.7775	***
Lagrange Multiplier (error)	36.6833	***	11.0126	***	45.1469	***
Lagrange Multiplier (lag)	33.7476	***	24.4012	***	84.6808	***
Log Likelihood	849.4870		842.7620		831.9830	
AIC	-3.7669		-3.7365		-3.6877	
R^2 adj.	0.8093		0.8034		0.7935	
Prob F	0.0000		0.0000		0.0000	

Note: For spatial weights matrices see Table 2. SpaceStat carried out the Breusch-Pagan test for ID_K2, and the Koenker-Basset test for the other two weights matrices to test heteroscedasticity.

The diagnostics for spatial dependence point to an additional spatial dependence in all three cases. We therefore also checked a combination of spatial lagged exogenous and endogenous variables (combination of models M2 and M3), i.e. a “mixed regressive-spatial regressive model” (Florax und Folmer, 1992), as well as a combination of spatial lagged exogenous variables and spatial error autocorrelation (combining M1 and M3). The estimation results for the combination of spatial lagged exogenous and endogenous variables are given in Table B3. However, the spatially lagged external variables size and GDP become insignificant. The coefficients of the significant spatially lagged external variables PODEN and UEWP now both point again to a positive spatial correlation, in addition to the significant negative spatial lag dependence. With respect to the information criteria, all estimates display a better fit than the simple OLS estimates of spatially lagged external variables (specification M3) as well as the ML estimates of the spatial error model (specification M2).

The results of the mixed regressive-spatial autoregressive error models are also displayed in Table B3. The test on further spatial lag dependence is highly significant for the estimates using the neighbourhood contiguity matrix, thus pointing to a misspecification. Only the use of the weights matrix on the basis of squared inverse distances between regional capitals (ID_K2) provides evidence of an improved fit for this specification. In this regression, however, we find a negative spatial dependence with respect to the spatial error term and the spatially lagged GDP and size of the region, but the contrary with respect to the spatially lagged population density and unemployment rate. Using ID_Z2 (NGH_CM), we only have two (three) significant spatially lagged external variables, also differing in their impact. Results are thus inconsistent and largely dependent on the choice of the spatial interaction matrix for this specification.

However, both extensions of the simple spatial cross-regressive model largely confirm the described impacts of economic variables on the regional level of specialisation while controlling for spatial interdependencies. Thus, the determinants identified by classical econometric methods remain robust throughout all the spatial econometric estimates.

It might be possible to further improve the spatial specification by the use of e.g. higher-order spatial models or models of spatial heterogeneity. In addition, the autocorrelation which is still obvious in the Breusch-Pagan tests might also be due to time correlation and not to further spatial variation. However, we restrict our analyses to the presented spatial specifications since more sophisticated spatial models are not implemented in the software so far.

IV Conclusion

This analysis further checked the results of an earlier study (Stirboeck, 2002) with respect to the use of different measures of relative investment specialisation and different estimation methods. In particular, potential regional interdependencies as well as spatial data correlation due to shortcomings of regional data or an inadequate definition of regional entities are taken into account by the use of spatial econometric tools. Independent of the estimation technique and the indicator used, we find similar results to Stirboeck (2002).

Table 6: Impact of economic variables on the level of relative specialisation

Economic variable	**Sign of impact on the level of investment specialisation**
Market potential (Gross domestic product)	-
Fact of being a central region	+
Unemployment rate	+
Population density	+
Size if a region	-
Economic openness	+
Distance to economic centre	+

Table 6 summarises the identified determinants of the level of relative specialisation. The bigger the size of a region is, the higher is the similarity of relative investments. Market size reflects the economic and demand potential of a region: The higher it is, the lower the relative specialisation in terms of investments tends to be. This contrasts to the results of recent empirical studies on sectoral agglomeration, which found market size to have an increasing influence on sectoral concentration across space. While firms tend to locate close to large markets and high demand (thus increasing the spatial concentration of sectors), regions with a large market seem to attract capital of all types of sectors with a rather even relative allocation (thus decreasing relative regional dissimilarities). This effect is counteracted by an apparently strong tendency towards high specialisation of central, economically most important regions who demonstrate to have a significantly higher level of relative investment specialisation. The unemployment rate, finally, reflects negative economic performance of a region (not accounting for migration effects etc.). The higher it is, the stronger the relative regional specialisation turns out to be.

The higher the distance of a region to its national economic centre is, the less similar are its investment shares to EU average. Peripheral regions are thus stronger specialised in terms of relative investments than regions closer to the centre. In addition, the extent of economic openness consistently seems to have a significant increasing impact on relative specialisation levels of gross fixed capital formation. This means that further market integration of EU countries can be assumed to lead to a lower diversification of sectoral investments in EU regions. Regional investment structures will thus become more dissimilar. Finally, the impacts of population density on the specialisation level are not significant throughout all estimates. However, we might still suppose an increasing impact on regional specialisation.

The level of regional specialisation might also influence the unemployment rate or gross domestic product respectively, what we controlled for in instrumental-variable estimates. However, we are confronted with highly specialised regions which are subject to a good economic performance as well as highly specialised regions which are not. Thus, not only the level of specialisation is essential, but also the sectoral pattern of specialisation. This is an important topic which is worth being analysed further in future research.

Regardless of the estimation method and the discussion about the best fit, the spatial econometric regressions display negative spatial dependence or interaction between the regions. The OLS test diagnostics on spatial dependence point to a spatial error correlation, what is in line with the information criteria which indicate a better performance of the spatial error model compared to the spatial lag model. We might thus interpret the spatial correlation in the data as nuisance manifested in negatively spatially correlated error terms. Data inconsistencies, shortcomings of the regional databases or incompatibilities of the units of observation with actual economic regions might be one reason for such spatial nuisances in the data.

Further improvements of the spatial econometric estimates can be obtained by adding spatially lagged external variables. However, results are rather inconsistent with respect to the identification of the nature of the spatial dependence, largely differing with respect to the specific model and spatial weights matrix chosen.

Bibliography

AMITI, M. (1999): Specialization Patterns in Europe, Weltwirtschaftliches Archiv 135 (4), 573-593.

ANSELIN, L. (1988): Spatial econometrics: methods and models, Dordrecht, Kluwer Academic Publishers.

ANSELIN, L. (1992): SpaceStat Tutorial, A Workbook for Using SpaceStat in the Analysis Spatial Data, available: http://www.spacestat.com/docs/spacestat_tutorial.pdf.

ANSELIN, L. (1995): SpaceStat Version 1.80 User's Guide, available: http://www.spacestat.com/docs/V180man.pdf.

ANSELIN, L. (1996): The Moran scatterplot as an ESDA tool to assess local instability in spatial association, in: Fisher, M. and D. Unwin (eds.): Spatial analytical perspectives on GIS, Taylor and Francis, London.

ANSELIN, L. (1999a): Spatial econometrics, mimeo, available: http://geog55.gis.uiuc.edu/~luc/papers/baltchap.pdf.

ANSELIN, L. (1999b): Spacestat Version 1.90, Ann Arbor, BioMedware.

ANSELIN, L. (2001): Spatial Econometrics, in: Baltagi, B (Hrsg.): A Companion to Theoretical Econometrics, Oxford, Basil Blackwell, 310-330.

ANSELIN, L. AND A. K. BERA (1998): Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in: Ullah, A. und D.E.A. Giles (Hrsg.): Handbook of Applied Economic Statistics, New York, 237-289.

ANSELIN, L., BERA, A.K., FLORAX, R. AND M.J. YOON (1996): Simple diagnostic tests for spatial dependence, Regional Science and Urban Economics 26 (1996), 77-104.
BERA, A.K. AND M.J. YOON (1993): Specification Testing with Misspecified Alternatives, Econometric Theory, 9, 649-658.
BRÜHLHART, M. (1998): Trading Places: Industrial Specialisation in the European Union, Journal of Common Market Studies 36 (3) 1998, 319-346.
BURRIDGE, P. (1980): On the Cliff-Ord Test for Spatial Autocorrelation, Journal of the Royal Statistical Society, 42, 107-108.
COWELL, F. A. (1995): Measuring Inequality, London, Prentice Hall.
ESRI (2001): ArcView 3.2, http://www.esri.com/software/arcview/index.html.
FLORAX, R. AND H. FOLMER (1992): Specification and estimation of spatial linear regression models – Monte Carlo evaluation of pre-test estimators, Regional Science and Urban Economics 22, 405-432.
GREENAWAY, D. AND R.C. HINE (1991): Intra-industry Specialization, Trade Expansion and Adjustment in the European Economic Space, Journal of Common Market Studies 29(6), 389-403.
HAINING, R. (1990): Spatial data analysis in the social and environmental sciences, Cambridge University Press, Cambridge.
HIRSCHMAN, A. (1958): The Strategy of Economic Development, Yale University Press, New Haven.
IMF (several years): Annual Report on Exchange Arrangements and Exchange Restrictions, Washington.
KALEMLI-OZCAN S., B.E. SORENSEN AND O. YOSHA (1999): Risk Sharing and Industrial Specialization: Regional and International Evidence, mimeo.
KLÜVER, A. AND G. RÜBEL (1998): Räumliche Industriekonzentration und die komparativen Vorteile von Ländern – eine empirische Studie der Europäischen Union, Passauer Diskussionspapiere, Diskussionsbeitrag No. 5-98, Passau.
KRIEGER-BODEN, C. (1999): Nationale und regionale Spezialisierungsmuster im europäischen Integrationsprozeß, Die Weltwirtschaft (2), 234-254.
KRUGMAN, P. (1991): Geography and Trade, Cambridge/Mass., MIT Press.
LINNEMANN, H. (1966): An Econometric Study of International Trade Flows, North-Holland, Amsterdam.
MOLHO, I. (1995): Spatial Autocorrelation in British Unemployment, Journal of Regional Science, Vol. 53, 4, 641-658.
MYRDAL, G. (1957): Economic Theory and Under-Developed Regions, London.
NIEBUHR, A. (2001): Convergence and the Effects of Spatial Interaction, Jahrbuch für Regionalwissenschaft, 21, 113-133.
QUINN, D. (1997): The Correlates of Change in International Financial Regulation, American Political Science Review, 91(3), 531-551.
QUINN, D. (2000): Democracy and International Financial Liberalization, mimeo.
SAPIR, A. (1996): The Effects of Europe's Internal Market Program on Production and Trade: A First Assessment, Weltwirtschaftliches Archiv 132 (3), 457-475.
SCHULZE, P.M. (1982): Zur Schätzung regionaler Konsumfunktionen, Jahrbuch für Regionalwissenschaft, 116-128.
STIRBOECK, C. (2001): Agglomeration Tendencies in EU Regions: Where Does Capital Go?, ZEW Discussion Paper No.01-28.
STIRBOECK, C. (2002): Relative Specialisation of EU Regions: An Econometric Analysis of Sectoral Gross Fixed Capital Formation, ZEW Discussion Paper No.02-36.
TINBERGEN, J. (1962): Shaping the World Economy, Suggestions for an International Economic Policy, Twentieth Century Fund, New York.
TUKEY, J.W. (1977): Exploratory Data Analysis, Reading MA, Addison-Wesley.

Appendix

A Data description

The regional disaggregation of the data is given according to the Nomenclature of Territorial Units for Statistics (NUTS - Nomenclature des unités territoriales statistiques). The REGIO database disaggregates data for the three aggregation levels NUTS 1, 2 and 3. However, data for GFCF is not available further disaggregated than the NUTS 2-level. In addition, it is not complete (with regard to the regional and/or the sectoral disaggregation – the latter needed for the calculation of the specialisation indices). Data availability is sufficient for the seven countries given below. Here, the UK does not provide data disaggregated further than NUTS 1-level. Luxembourg, Denmark as well as Ireland are only regarded as one single region at the NUTS 1- as well as at the NUTS 2-level (=monoregional countries). The maximum number of regions available is therefore 33 at the NUTS 1-level and 56 at the NUTS 2-level.

Table A1: Regional data for GFCF from the REGIO database

Country	NUTS level	Respective national dis-aggregation level	Number of regions NUTS 1	Number of regions NUTS 2
UK	1	Groups of Counties or local authority regions	11 (with 3 n.a.)	n.a.
Belgium	2	Provinces	3	11
France	2	Régions	8	22
Italy	2	Regioni	11	20
Denmark	1&2	-	1	1
Ireland	1&2	-	1	1
Luxembourg	1&2	-	1	1
Total number of regions			33 (+ 3 n.a.)	56

Note: Version of *NUTS 1995*. French oversea departments (DOM – départements outre-mer) are not counted in total sums for France as well as for the EU. The three non-available British regions are: North, North-West as well as South East (including London).

Table A1 displays the availability and the number of regions for the two levels of aggregation. Table A2 presents the respective regions.

Table A2: Overview on NUTS 2- as well as NUTS 1-regions

France		**Italy**		**Belgium**		**United Kingdom**	
Nuts 2		**Nuts 2**		**Nuts 2**			
Alsace	ALS	Abruzzo	ABR	Antwerpen	ANT		
Aquitaine	AQU	Basilicata	BAS	Brabant Wallon	BWA		
Auvergne	AUV	Calabria	CAL	Bruxelles-capitale	BRU		
Basse-Normandie	BNO	Campania	CAM	Hainaut	HAI		
Bourgogne	BOU	Emilia-Romagna	ERO	Liège	LIE		
Bretagne	BRT	Friuli-Venezia Giulia	FVG	Limburg (B)	LIM		
Centre (F)	CTR	Lazio	LAZ	Luxembourg (B)	LUB		
Champagne-Ardenne	CHA	Liguria	LIG	Namur	NAM		
Corse	CRS	Lombardia	LOM	Oost-Vlaanderen	OVL		
Franche-Comté	FRC	Marche	MAR	Vlaams Brabant	VBR		
Haute-Normandie	HNO	Molise	MOL	West-Vlaanderen	WVL		
Ile de France	IDF	Piemonte	PIE				
Languedoc-Rousillon	LRO	Puglia	PUG			**Monoregional Countries (Nuts2, Nuts1)**	
Limousin	LIS	Sardegna	SAR			Denmark	DEN
Lorraine	LOR	Sicilia	SIC			Ireland	IRE
Midi-Pyrénées	MPY	Toscana	TOS			Luxembourg	LUX
Nord - Pas-de-Calais	NPC	Trentino-Alto Adige	TAA				
Pays de la Loire	PDL	Umbria	UMB				
Picardie	PIC	Valle d'Aosta	VAO				
Poitou-Charentes	POI	Veneto	VEN				
Provence-Alpes-Côte d'Azur	PAC						
Rhône-Alpes	RAL						
Nuts 1		**Nuts 1**		**Nuts 1**		**Nuts 1**	
Bassin Parisien	BPA	Abruzzo-Molise	ABM	Bruxelles-capitale	BRU	East Anglia	EAN
Centre-Est	CES	Campania	CAM	Région Wallonne	RWA	East Midlands	EMI
Est (F)	EST	Centro (I)	CEI	Vlaams Gewest	VLA	North	NOR
Ile de France	IDF	Emilia-Romagna	ERO			North West	NWE
Méditerranée	MED	Lazio	LAZ			Northern Ireland	NIR
Nord - Pas-de-Calais	NPC	Lombardia	LOM			Scotland	SCO
Ouest	OUE	Nord Est	NES			South East	SOE
Sud-Ouest	SOU	Nord Ovest	NOV			South West	SOW
		Sardegna	SAR			Wales	WAL
		Sicilia	SIC			West Midlands	WMI
		Sud	SUD			Yorkshire and the Humber	YOR

Data are taken from the Eurostat REGIO Database (yearbooks up to 2000) which – for gross fixed capital formation - comprises data for the years 1985 to 1994. All data included in the analysis are based on ESA79 (European System of Accounts, Version 1979).

Table A3: List of explanatory variables, REGIO Database

abbreviation	variable	unit
GFCF	Gross Fixed Capital Formation	Currency: Billions of ECU
GDP	Gross domestic product	Currency: Billions of ECU
PAT	European R&D patent applications	total number
UEWP	Total Unemployment rates	in % OF WORKING POPULATION
POP	Total annual average population	in Mio. PERSONS
PODEN	Population density	in 1000 INHABITANTS/KM2

In addition to the available national account data, a number of further variables has been used in the econometric analysis. Transport costs are proxied by a distance variable (**DIST**) which measures the optimal route distance to the centres of the respective countries which are Paris, Rome, London and Brussels. The distance is defined to be 1 for Denmark, Luxembourg as well as Ireland, and it is equally 1 for the regions containing the capital of the respective country. Central and economically important regions (**CENTR**) in the analysis are Île de France (France), Brussels (Belgium), and Lazio (Italy).

Table A4: List of further explanatory variables

abbreviation	variable	unit
DIST	distance to centre, index of peripherality	1000 km
CENTR	regional dummy set for central region	0 or 1
QUINN_OPENN	indicator of openness per country	0-14 (variation by 0.5)

Available indicators of liberalisation arising from official sources are mostly indicator variables being either 0 or 1. However, such indicator variables do not allow to differentiate the varying levels of control or to capture a decreasing level of control over time. Measuring a level of integration for each year is therefore a better solution from an econometric point of view. Quinn (1997, 2000) has constructed such a yearly index of openness on the basis of those restrictions published by the IMF since the 1950s. This index is scaled from 0 (highest degree of restrictions) up to 14 (highest degree of liberalisation) and aggregates the different indicators of liberalisation progress in seven specified fields (capital in – and outflows, im– and exports of goods and of services as well as international conventions of liberalisation) with a respective degree of liberalisation between 0.5 and 2. Quinn weighs quantitative restrictions of imports for example the highest (i.e. he attributes the lowest partial liberalisation index of 0 in case of full and 0.5 in case of partly quantitative restrictions), existence of laws requiring the approval of international transactions are scored 1, taxes 1.5 and finally free trade 2. With regard to capital account liberalisation, Quinn attributes 0 in case of required approval for capital transactions which are rarely granted, 0.5 (1) in case of occasional (frequent) approval and finally 1.5 in case of taxing measurements (without the need of an official approval).

These indicators, however, are only available at country, not regional, level, which has to be taken into account in econometric analysis. Detailed restrictions for Luxembourg are not available as Luxembourg and Belgium are part of a common monetary union since the 1950s. In our analysis the „Quinn-indicator“ for Luxembourg is therefore naturally set equal to the one of Belgium.

B Estimation Results

Table B1: Estimation Results for the Determinants of Relative Regional Investment Specialisation, NUTS 1-level, 1985-1994

	VCCFEU	GCCFEU	FKCFEU	VCCFEU	GCCFEU	FKCFEU
	level	level	level	logit	logit	logit
Constant	0.0498	0.2835	1.0009	-2.2550	-0.9265	2.3059
	1.41	8.84	21.48	-12.13	-6.70	10.33
GDP	-0.0004	-0.0004	0.0011	-0.0024	-0.0018	0.0051
	-3.80	-4.03	3.77	-4.05	-4.05	3.76
CENTR	0.1108	0.1251	-0.1283	0.7193	0.5244	-0.5924
	5.64	7.04	-6.16	6.98	6.85	-5.93
UEWP	0.0021	0.0050	0.0000	0.0171	0.0205	0.0016
	2.05	5.41	0.02	3.21	5.18	0.22
PODEN	0.0453	0.0326	-0.2501	0.1705	0.1469	-1.2060
	8.57	6.81	-3.09	6.14	7.12	-3.11
AREA	0.0002	-0.0001	-0.0008	0.0010	-0.0005	-0.0036
	1.08	-0.86	-2.3	1.15	-0.80	-2.27
QUINN_OPENN	0.0097	0.0067	-0.0230	0.0494	0.0303	-0.1118
	3.63	2.76	-6.33	3.52	2.90	-6.41
DIST	0.0270	0.0215	-0.0744	0.1684	0.0942	-0.3594
	1.48	1.3	-3.15	1.76	1.32	-3.17
DUM_FRA	-0.0151	-0.0211	0.0254	-0.1129	-0.0842	0.1157
	-1.24	-1.91	1.96	-1.76	-1.77	1.86
DUM_BEL	0.0268	-0.0020	--	0.1663	-0.0046	--
	1.70	-0.14		2.01	-0.07	
DUM_IRE	0.0499	-0.0313	-0.2813	0.0458	-0.1345	-1.2804
	1.67	-1.16	-8.95	0.29	-1.15	-8.5
DUM_LUX	-0.0499	-0.1589	-0.0305	-0.2827	-0.6839	-0.0970
	-1.60	-5.63	-0.90	-1.73	-5.62	-0.59
DUM_DEN	0.0183	-0.1346	-0.3821	0.0362	-0.5616	-1.6789
	0.68	-5.54	-12.88	0.26	-5.37	-11.81
DUM_UKD	0.1314	0.0080	-0.4464	0.7095	0.0352	-1.9785
	14.22	0.96	-41.98	14.63	0.98	-38.82
no. obs.	292	292	262	292	292	262
SSR	0.9453	0.7746	0.9269	26.0299	14.3837	21.2893
Log Likelihood	422.69	451.77	367.64	-61.37	25.2256	-42.93
Prob Chi²	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AIC	-2.799	-2.998	-2.700	0.516	-0.077	0.435

Note: Line below the coefficients presents the z-values of the GLS estimates. SSR displays the sum of squared residuals. The probability of the Chi²-test gives the joint significance of all coefficients. Since sectoral sums for gross fixed capital formation are not available for Belgian regions, we cannot calculate Finger-Kreinin indices for the 30 NUTS 1-Belgian regions.

Table B2: Estimates checking for robustness, 1985-94

	NUTS 2	NUTS 1	NUTS 2	NUTS 1		NUTS 2	NUTS 1	NUTS 2	NUTS 1
IV 2SLS	VCCFEU	VCCFEU	GCCFEU	GCCFEU	dyn. model	VCCFEU	VCCFEU	GCCFEU	GCCFEU
Constant	0.1054	0.0460	0.3288	0.2784	Constant	0.0007	0.0111	0.1093	0.0645
	2.55	1.21	9.69	8.13		0.03	0.42	3.61	2.52
					End. lag AR(1)	0.7586	0.6529	0.6071	0.6857
						25.36	15.81	15.45	17.29
GDP (IV)	-0.0007	-0.0004	-0.0004	-0.0004	GDP	-0.0002	-0.0002	-0.0002	-0.0001
	-4.72	-3.85	-3.76	-3.85		-2.00	-2.08	-1.83	-1.74
CENTR	0.1683	0.1145	0.1703	0.1286	CENTR	0.0458	0.0394	0.0662	0.0404
	5.50	5.62	6.79	6.99		2.32	2.71	3.15	3.13
UEWP (IV)	0.0046	0.0018	0.0064	0.0048	UEWP	0.0009	0.0003	0.0023	0.0014
	3.97	1.59	6.81	4.85		1.33	0.46	3.02	2.12
PODEN	0.0028	0.0435	0.0004	0.0313	PODEN	0.0016	0.0161	0.0021	0.0105
	0.37	7.90	0.06	6.29		0.35	3.90	0.43	3.03
AREA	-0.0012	0.0002	-0.0014	-0.0001	AREA	-0.0002	0.0001	-0.0003	-0.00003
	-2.16	1.06	-3.1	-0.81		-0.66	0.44	-0.88	-0.27
QUINN_OPENN	0.0059	0.0103	0.0028	0.0071	QUINN_OPENN	0.0033	0.0043	0.0028	0.0042
	2.01	3.71	1.15	2.8		1.80	2.23	1.43	2.51
DIST	0.0576	0.0254	0.0666	0.0174	DIST	0.0205	0.0070	0.0200	0.0009
	2.86	1.35	4.04	1.03		1.56	0.54	1.44	0.08
DUM_FRA	0.0063	-0.0127	-0.0279	-0.0189	DUM_FRA	0.0011	-0.0050	-0.0154	-0.0081
	0.55	-1.03	-2.98	-1.7		0.15	-0.57	-1.98	-1.07
DUM_BEL	0.1594	0.0259	0.1183	0.0046	DUM_BEL	0.0388	0.0072	0.0455	0.0058
	10.33	1.58	9.36	0.31		3.56	0.64	4.05	0.60
DUM_IRE	0.0516	0.0487	-0.0043	-0.0327	DUM_IRE	0.0256	0.0316	-0.0035	0.0002
	0.91	1.61	-0.09	-1.19		0.69	1.44	-0.09	0.01
DUM_LUX	-0.1106	-0.0561	-0.1973	-0.1613	DUM_LUX	-0.0301	-0.0235	-0.0772	-0.0521
	-2.48	-1.74	-5.39	-5.55		-1.01	-1.03	-2.42	-2.54
DUM_DEN	0.0257	0.0150	-0.1225	-0.1377	DUM_DEN	0.0202	0.0206	-0.0609	-0.0499
	0.57	0.56	-3.34	-5.63		0.67	1.03	-1.92	-2.77
DUM_UKD	---	0.1444	---	0.0178	DUM_UKD	---	0.0608	---	0.0127
		14.81		2.02			7.21		2.22
no. obs.	457	270	457	270	no. obs.	428	256	428	256
SSR	3.4217	0.8065	2.2969	0.6582	SSR	1.2571	0.3673	1.3749	0.2750
Prob F	0.0000	0.0000	0.0000	0.0000	Prob Chi2	0.0000	0.0000	0.0000	0.0000

Note: GDP and UEWP have been instrumented by their first lag. The probability of the F-test and Chi²-test for IV and GLS estimates respectively display the joint significance of all coefficients. T-values are given in the IV estimates, z-values for the dynamic model.

Table B3: Maximum likelihood estimates of mixed regressive-spatial autoregressive lag/error models (442 observations)

	Spatial lag model with spatially lagged exog. var.							Spatial error model with spatially lagged exog. var.					
VARIABLE \ weights matrix	ID_K2	z-value	ID_Z2	z-value	NGH_CM	z-value		ID_K2	z-value	ID_Z2	z-value	NGH_CM	z-value
W_VNCFEU	-1.1626	-6.10	-1.0722	-6.95	-0.5843	-8.92							
CONSTANT	0.2259	4.86	0.2359	6.15	0.1680	7.90		-0.0321	-1.54	0.0457	2.24	0.0743	4.01
GDP	-0.0004	-7.06	-0.0004	-6.91	-0.0004	-7.63		-0.0004	-7.22	-0.0004	-7.13	-0.0005	-7.63
W_GDP	0.0000	-0.16	0.0000	-0.06	-0.0001	-0.87		0.0006	2.64	0.0000	0.00	0.0001	0.82
CENTR	0.1055	8.31	0.1039	8.38	0.0874	6.86		0.1339	11.24	0.1257	9.68	0.1306	9.34
UEWP	0.0043	6.81	0.0041	6.51	0.0028	4.88		0.0035	5.21	0.0029	4.43	0.0047	7.98
W_UEWP	0.0105	6.94	0.0102	7.31	0.0084	8.18		0.0072	5.11	0.0055	4.52	0.0020	2.98
PODEN	0.0477	14.00	0.0435	12.95	0.0466	14.44		0.0432	12.30	0.0308	6.02	0.0411	10.44
W_PODEN	0.0950	6.28	0.0626	6.01	-0.0161	-0.98		0.0308	3.35	0.0157	1.65	-0.0692	-3.27
AREA	-0.0008	-3.46	-0.0009	-3.92	-0.0009	-4.00		-0.0008	-3.00	-0.0009	-3.37	-0.0006	-2.20
W_AREA	-0.0001	-0.12	-0.0009	-1.28	0.0005	1.54		0.0030	4.73	0.0014	2.18	0.0000	0.09
QUINN_OPENN	0.0048	4.17	0.0048	4.05	0.0045	3.90		0.0045	3.96	0.0045	3.78	0.0052	4.32
DIST	0.0608	7.26	0.0594	6.85	0.0335	3.33		0.0600	8.81	0.0611	7.58	0.0688	7.91
DUM_FRA	-0.0099	-1.87	-0.0074	-1.47	-0.0253	-5.07		-0.0236	-5.96	-0.0173	-4.05	-0.0150	-3.94
DUM_BEL	0.1325	11.91	0.1376	14.67	0.1351	15.03		0.1033	11.10	0.0869	8.07	0.1164	12.57
DUM_IRE	-0.0534	-1.91	-0.0426	-1.52	-0.0823	-3.04		-0.0890	-3.23	-0.0722	-2.48	-0.0879	-2.99
DUM_LUX	0.0103	0.59	0.0031	0.16	-0.0585	-3.46		0.0280	1.61	-0.0554	-3.08	-0.0662	-3.87
DUM_DEN	0.0804	4.35	0.0874	4.76	0.0302	1.67		0.0418	2.39	0.0441	2.34	0.0528	2.79
Lambda								-1.7179	-10.36	-1.3164	-13.17	-0.8295	-5.84
Breusch-Pagan test	145.6835	***	146.1669	***	151.7797	***		157.1167	***	139.5096	***	116.9912	***
LR-Test on spatial lag/error dependence	32.4698	***	28.0814	***	78.6685	***		58.1719	***	22.4040	***	33.0700	***
LM-Test on spatial error/lag dependence	0.3817		0.4956		0.0099			0.0083		1.2940		33.3046	***
Log Likelihood	865.72		856.80		871.32			878.57		853.96		848.52	
AIC	-3.8358		-3.7955		-3.8612			-3.8985		-3.7872		-3.7625	

Note: For spatial weights matrices see Table 5. Lambda displays the spatial error autocorrelation.