On proximity and hierarchy: exploring and modelling space using multilevel modelling and spatial econometrics

Thomas Vanoutrive
Ghent University, Department of Geography
University of Antwerp, Department of Transport and Regional Economics

Angela Parenti
University of Pisa, Dipartimento di Scienze Economiche

Abstract

Spatial econometrics and also multilevel modelling techniques are increasingly part of the regional scientists' toolbox. Both approaches are used to model spatial autocorrelation in a wide variety of applications. However, it is not always clear on which basis researchers make a choice between spatial econometrics and spatial multilevel modelling. Therefore it is useful to compare both techniques. Spatial econometrics incorporates neighbouring areas into the model design; and thus interprets spatial proximity as defined in Tobler's first law of geography. On the other hand, multilevel modelling using geographical units takes a more hierarchical approach. In this case the first law of geography can be rephrased as 'everything is related to everything else, but things in the same region are more related than things in different regions'. The hierarchy (multilevel) and the proximity (spatial econometrics) approach are illustrated using Belgian mobility data and productivity data of European regions. One of the advantages of a multilevel model is that it can incorporate more than two levels (spatial scales). Another advantage is that a multilevel structure can easily reflect an administrative structure with different government levels. Spatial econometrics on the other hand works with a unique set of neighbours which has the advantage that there still is a relation between neighbouring municipalities separated by a regional boundary. The concept of distance can also more easily be incorporated in a spatial econometrics setting. Both spatial econometrics and spatial multilevel modelling proved to be valuable techniques in spatial research but more attention should go to the rationale why one of the two approaches is chosen. We conclude with some comments on models which make a combination of both techniques.

Key words: spatial econometrics; multilevel modelling

1 Introduction

Numerous lists and classifications of geographic and spatial concepts exist. Table 1 gives the list of Fellman et al. (1999 p.7-18). Researchers implicitly and explicitly incorporate these concepts in different types of models, like gravity, spatial lag and spatial multilevel models. In what follows, we will focus on hierarchy, which is related to 'Place Similarity and Regions', and proximity, which is part of 'Interaction Among Places'. Since concepts are related, other concepts than proximity and hierarchy will be mentioned, like scale.

-Location, Direction and Distance
-Size and Scale
-Physical and Cultural Attributes
-Changing Attributes of Place (Time)
-Interaction Among Places
-The Structured Content of Place
-Place Similarity and Regions

Table 1: geographic concepts (Fellman et al., 1999)

To understand 'Interaction Among Places', one should first understand the first item showed in Table 1, i.e. the location of a place. Location becomes meaningful when considering the position of a place in relation to other places or activities. Indeed, the relative location relates to spatial interconnection and interdependence, and is as a consequence related to the concept of 'Interaction Among Places'. Just as for location, not the absolute, but the relative distance matters (e.g. time distance). Hereby, distance is a measure for nearness, like in Tobler's First Law of geography: "everything is related to everything else, but near things are more related than distant things" (Anselin, 2002; Miller, 2004). In short, things in each others proximity are more related.

Next to proximity, we will focus on hierarchy, which is related to 'Place Similarity and Regions'. The division of the earth's surface into regions is the basic tool of spatial (areal) generalisation. This generalisation is necessary to understand the overwhelming diversity and complexity of the earth's surface. The concept of 'Place similarity' indicates that a region contains significant elements of internal uniformity (spatial similarity) and external difference from surrounding territories. A region thus contains a less or more uniform set of physical, cultural and/or organisational features. The basic characteristics of regions are their location, spatial extent, boundaries and hierarchical arrangement (nesting of regions). Especially this hierarchical arrangement of regions is of our interest.

In this paper, we will investigate how proximity and hierarchy are incorporated in spatial regression models. As Anselin (2002) points out, a fundamental problem in the analysis of spatial correlation is the lack of identification of the parameters of the complete covariance matrix. Therefore, it is necessary to impose a structure on the variance–covariance. Three main approaches can be considered in spatial regression analysis. The first is a geostatistics approach, which defines space as a continuous surface. In regional science however, observations are usually discrete objects. As a consequence, we will not focus on geostatistical models or on techniques like point pattern analysis. The second approach uses an object view and the corresponding lattice model. In this case, a spatial structure is imposed using a spatial weights matrix that underlies a spatial process model. The third way to impose a spatial structure is the spatial error components approach, as used in spatial multilevel modelling (Anselin, 2002).

1.1 Hierarchy

Hierarchy deals with the inequalities between geographical objects. These objects can be grouped or classified using different hierarchical levels. The most well-known example of hierarchy in geography is that of the differentiation between centres, like in Christaller's theory of central places. However, in the market and transport principles, lower level units are shared by several higher level units (Pumain and Saint-Julien, 2001). In this paper we use the term hierarchy only for lower level units which are just part of one higher level unit, like in Christaller's administration principle. A higher level unit is thus a group of lower level units.

In spatial multilevel models, levels are geographical units, like municipalities, districts, regions and countries. Multilevel modelling is increasingly used to incorporate contextual and compositional factors in regressions and to investigate the role of higher geographical scales. Examples can be found in health (Langford et al., 1998), housing market (Orford, 2000) and commuting research (Schwanen et al., 2004). Langford et al. (1998) state that '*including higher levels of geographical aggregation simultaneously in a model of smaller units is essential to draw useful conclusions from the data analysed.'* Next to this, data characteristics can motivate the choice for a hierarchical spatial structure. Data are often available at different geographical scales and aggregating the data at the highest level means information loss and the risk of ecological fallacy (Langford et al., 1998; Groenewegen et al., 1999), while Langford et al. (1998) also demonstrate '*the dangers of using aggregated data, taken from* *different sources, without specifying an appropriate hierarchy'*. Moreover, distributing neighbourhood characteristics over individual cases leads to underestimating the standard errors of the higher-level regression coefficients, increasing the probability of statistically significant, but non-existent relationships (Groenewegen et al., 1999; Luke, 2004).

Despite the rise of spatial multilevel modelling, the use of predefined spatial hierarchies is often contested (Brunsdon et al., 1998). Chaix et al. (2005b) propose the use of a continuous notion of space to avoid the dependence on a space fragmented into arbitrary administrative areas. Unlike this critique, they acknowledge that the hierarchical approach '*may be the appropriate choice when the context is defined in a way that is not strictly geographical (for example, workplaces or schools); when investigating processes that operate on the scale of administrative areas (for example, matters related to public policies); or when spatial correlation can be reduced to the correlation within areas'* (Chaix et al., 2005a; Chaix et al., 2005b). Groenewegen et al. (1999) link '*the usual assumption that spatial units are given, meaningful and fixed'* to the `modifiable areal unit problem' (MAUP). MAUP is the '*phenomenon whereby different results are obtained in analysis of the same data grouped into different sets of areal units'* (Manley et al., 2006). This problem can be subdivided in a scale and a zoning effect. The zonation issue concerns the effect of the arbitrary nature of the boundary division, while the scale issue covers the change of results when the level of analysis changes (Manley et al., 2006; Kwan and Weber, 2008). The zonation issue is not necessarily a problem if a variable is a real contextual variable in stead of the aggregation of individual characteristics, or so-called compositional factors (Groenewegen et al., 1999). This division between compositional and contextual data received much attention in literature (Duncan et al., 1998; Subramanian et al., 2001; Mohan et al., 2005; French and Jones, 2006). Indeed, a more nuanced view of geography questions if regional variation is 'real', or just an artefact of geography, i.e. a compositional effect. Thereby, a compositional effect is nothing more than the weighted sum of patterns at smaller scales, while a 'real' neighbourhood effect is separated from the individual characteristics of individuals in a region. In the latter case, places only differ because different kinds of people live in different places (Johnston et al., 2007). When these compositional effects are taken into account, one can investigate at which scale processes occur. The scale issue in MAUP changes from a problem to an opportunity and enables researchers to detect the spatial scales that matter. As a consequence, a hierarchical spatial analysis has the potential to detect the origin of spatial variation. Moreover, modelling different scales simultaneously, like in spatial multilevel modelling, allows overcoming both the atomistic fallacy of individual-based studies, and the ecological fallacy of aggregated research (Duncan et al., 1998; Langford et al., 1998; Subramanian et al., 2001; Mohan et al., 2005; French and Jones, 2006). However, a spatial pattern in general originates from several distinct processes operating at several different spatial scales, and these scales may vary over space as well (Langford et al., 1998; Manley et al., 2006). Spatial multilevel modelling is thus a useful tool, but not a 'wonder' technique which can solve all problems in spatial analyses.

1.2 Proximity

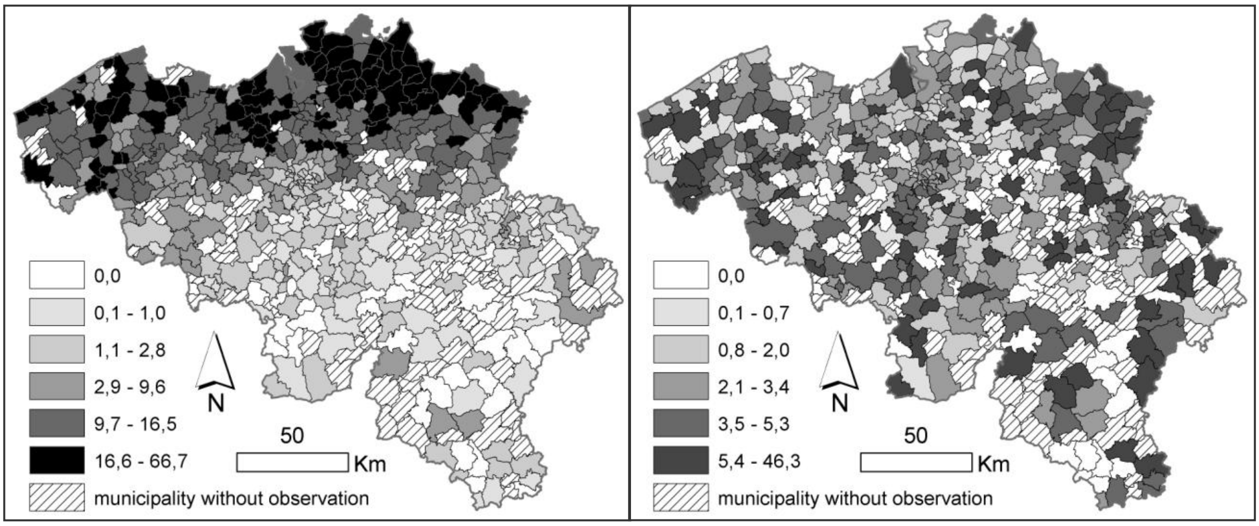
The basic concept in a spatial econometric approach is geographical proximity, and these spatial econometric techniques are increasingly part of the regional scientists' toolbox (Arbia and Fingleton, 2008). The main characteristic of these techniques is the use of a spatial weights matrix to take into account the role of proximity (Ciriaci and Palma, 2008). Such a matrix contains the spatial relations between observations. The relations can be expressed as 1's for neighbours and 0's for non-neighbours, as a function of the distance between observations, or as a function of other factors, like interregional trade (Le Gallo and Dall'erba, 2008). The arbitrary nature of the weights matrix is the main focus of spatial econometrics critics (Arbia and Fingleton, 2008). Nevertheless, the spatial econometric approach proved to be useful to detect spatial (e.g. knowledge) spillover effects, and to counterbalance spatial autorcorrelation. As a consequence, numerous studies use this approach to investigate inequalities and convergence between European regions (Tselios, 2008; Dall'erba and Le Gallo, 2008; Geppert and Stephan, 2008; Le Gallo and Dall'erba, 2008; Olejnik, 2008). Next to spatial autocorrelation, spatial heterogeneity is an important topic in the spatial econometric literature. This heterogeneity means that different processes have a different impact in different places. For instance subsidies may have a different effect in poor and in rich regions. Such groups of regions are also called spatial regimes, and can be considered as a higher level in a spatial multilevel modelling approach.

The ways in which spatial autocorrelation and heterogeneity are treated in spatial econometric models is diverse. Le Gallo and Dall'erba (2008) for instance, compared a distance- and a time-based weights matrix and distinguished only two spatial regimes among European regions (core and periphery) using dummy variables interacting with the other variables. Olejnik (2008) introduced next to a weights matrix with the three nearest neighbours, one dummy variable for new EU member countries. Bosker (2009) also separated Eastern European regions from those of the 'Old Europe' and concluded that in Eastern Europe country-specific factors are more important than in Western Europe. This conclusion is based on an exploratory map and the conditioning of the region's GDP per capita on the GDP per capita of the country the region belongs to. He referred to Quah (1996) who indicated that proximity (neighbour-relative per capita incomes) accounts for more observed regional inequality than hierarchy (state-relative per capita incomes), but that both spatial and national spillovers are important. Tselios (2008) directly incorporated the spatial regimes in the weights matrix by considering only regions which are part of the same welfare regime as potential neighbours. Dall'erba and Le Gallo (2008) stressed the importance of country-effects and mention the use of country dummies, but restricted their analysis to two regimes, core and periphery. Geppert and Stephan (2008) used country dummy variables and concluded that income disparities decreased between countries but not within countries. Finally, the awarded paper of Elhorst and Zeilstra (2007) found evidence of spatial autocorrelation between regions, and of variety in the autocorrelation coefficients across countries. Therefore, they developed a mixed model of random coefficients for regional-level variables and fixed coefficients for national-level variables, the model also contained a spatial weigths matrix based on passenger traffic travel times between regions to counterbalance the spatial autocorrelation. Summarizing, this model combined a spatial multilevel with a spatial econometrics approach since both the institutional framework (country) and the proximity of regions matter.

2 Measuring proximity and hierarchy

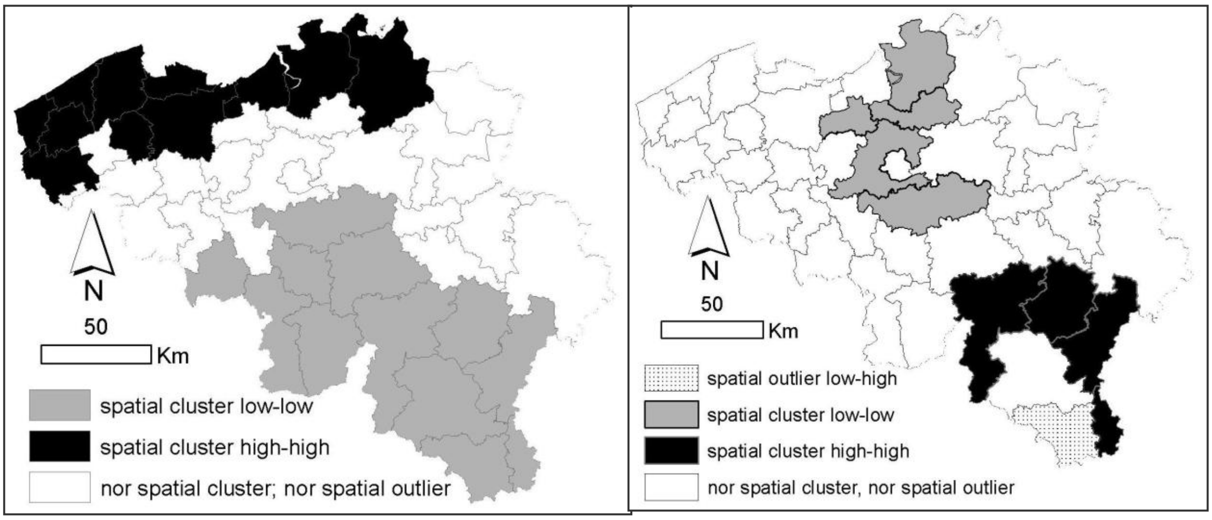
2.1 Measuring proximity

Proximity, but also hierarchy, can be seen as spatial autocorrelation, being the coincidence of value similarity with locational similarity. Spatial correlation is mostly measured using the Moran's I. This and other related indicators use the values of neighbouring spatial units to investigate if these have similar values. In this way, the proximity effect is measured. The Moran's I is a kind of Pearson correlation coefficient, but in stead of looking at the correlation between two different variables, it measures the correlation between the value of a variable and the value of the same variable for neighbouring observations (Legendre, 1993; Griffith and Layne, 1999).



Figures 1 and 2: average % cycling (left) and carpooling (right) employees in Belgian municipalities

Figures 1 and 2 show the share of cycling and carpooling in Belgian municipalities. A clear pattern appeared for bicycle use, municipalities where cycling is popular are located in the proximity of municipalities where cycling is also popular. As a consequence, the corresponding Moran's I is high (0.72). The spatial pattern for carpooling is less clear, logically, the Moran's I (0.34) is lower than the one for cycling.



Figures 3 and 4: LISA maps of cycling (left) and carpooling (right) in Belgian districts (arrondissements); weights matrix based on the 5 nearest districts

Next to overall measures for spatial autocorrelation, also Local Indicators of Spatial Association exist (LISA; (Anselin, 1995). Such measures have a value for each observation and indicate both spatial clusters and spatial outliers. These indicators detect spatial patterns and are often more meaningful than a single overall measure like the Moran's I. Two distinct spatial patterns can be distinguished in the LISA maps in Figures 3 and 4. For cycling, a cluster of districts with high values is detected in the north (high-high), and a cluster with low values in the south (low-low) of Belgium. Measuring the share of carpooling, a kind of core-periphery pattern appears, with a cluster of districts where carpooling is less popular in the centre of Belgium (around Brussels), and a carpool-cluster in the south and the east. A LISA map of carpooling using the municipality level (Figure 5) in stead of the arrondissement level (Figure 4), shows that the spatial pattern differs at different scales.

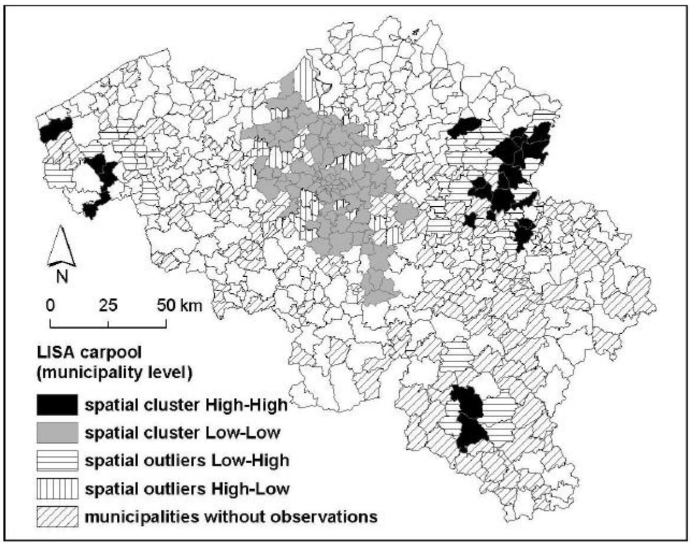


Figure 5: LISA map of carpooling at the municipality level (weights matrix based on all municipalities within 30 km)

2.2 Measuring hierarchy

Hierarchy means that observations within the same group have similar values, or by rephrasing Tobler's first law of geography: 'everything is related to everything else, but things in the same region are more related than things in different regions'. Multilevel indicators in the first place measure the within-area and the between-area correlation. The Variance Partition Coefficient (VPC) is regularly used to check which amount of the total variance can be attributed to the different levels (Rasbash et al., 2005; Uthman, 2008). The VPC is the relative importance of the different levels in the random part of a multilevel model (see 3.2). For cycling, 49% of the total variance can be attributed to the district level, 5% to the municipality level and 46% to the worksite level. The latter includes also the variance of the individual level (Tranmer and Steel, 2001), for which no data were available. Carpooling is less explained by a spatial hierarchy, since 95% of the variance can be attributed to the lowest level.

Detecting spatial clusters in a hierarchical approach is less meaningful, since the higher level units are pre-given clusters. However, observations can be classified as high-high, low-low, high-low and low-high to indicate the 'performance' of units in a multilevel context (Subramanian et al., 2001). In other words, such a classification indicates if an observation has a high or low value relative to the other observations in the same area, which has on its turn a low or high value compared to the other areas.

3 Modelling proximity and hierarchy

3.1 Modelling proximity: spatial econometrics

Spatial econometric regressions model ‘proximity’ by incorporating a spatial weights matrix W. The most common ways are the spatial lag and the spatial error model given in equations (1) and (2) (Anselin, 1988; Anselin, 2002).

Spatial lag model: $y = \rho Wy + X\beta + \varepsilon$ (1)

Spatial error model: $y = X\beta + \varepsilon$ with $\varepsilon = (1 - \rho W)^{-1}u$ (2)

Common causes of spatial dependence in regional data are measurement problems caused by the arbitrary delimitation and/or aggregation of spatial units of observations (MAUP) and, most importantly, the presence of spatial externalities and spillover effects. Next to the aforementioned Moran’s I statistic, a set of maximum likelihood tests can be performed to detect this spatial dependence. A first Lagrange Multiplier (LM) test is the the LMerr test, with as null hypothesis the absence of spatial dependence, and as alternative hypothesis the spatial error model. The second LM test, the LMlag test, has the same null hypothesis and as alternative hypothesis the spatial lag model (for more details see Anselin (2001)). Bera and Yoon (1993) and Anselin et al. (1996) also provide a robust version of these tests, respectively denoted as the RLMerr and RLMlag tests. These tests are commonly used to check if data are spatially autocorrelated, and if so, which type of spatial econometric model is the most appropriate to counterbalance this spatial dependence.

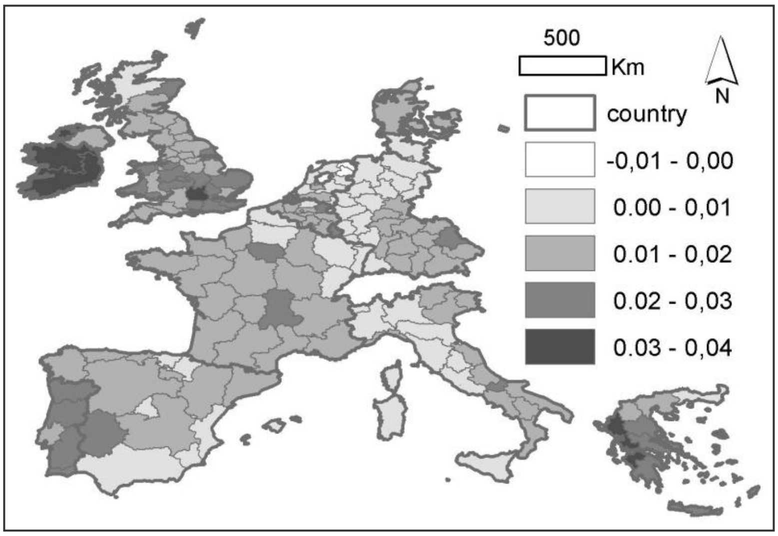


Figure 6: regional labour productivity growth in EU 12 (1980-2002)

We applied this methodology (Anselin, 2005) on labour productivity growth data of 173 European regions (Figure 6;see Fiaschi et al. (2009) for an explanation of data and model). As the diagnostics of the first model in Table 2 show, a spatial error model forces itself on the data. However, such a spatial error model (the second model in Table 2; Figure 7) does not explain all residual variance of the first model. Since European regions are nested in countries, the VPC can detect if the country level is a source of unexplained variance. Indeed, the second model does not take into account the country level, and does not contain variables whose dimension is typically national. Table 3 confirms this country hypothesis as no less than 64% of the variance in regional labour productivity growth can be attributed to the country level. The spatial error model does not account for this spatial hierarchal autocorrelation since 71% of the variance in the residuals of the second model in Table 3 can be attributed to the country level. A third model, a simple OLS regression with a dummy variable for each country, seems to be superior to the spatial error model. Indeed, the LM tests could not detect spatial autocorrelation anymore, and the adjusted R^2 raised from 0.50 for the first model to 0.74 for the third model. Logically, no variance in the residuals of the last model can be attributed to the country level. To conclude, the detected spatial error correlation is informative about country level omitted regressors. Indeed, among others Anselin, Florax and Rey (2004) already reported that spatial error correlation can be caused by the omission of relevant regressors. This implies that when modelling regional phenomena, we may not ignore the hierarchical structure of the data, i.e. the nesting of regions in countries.

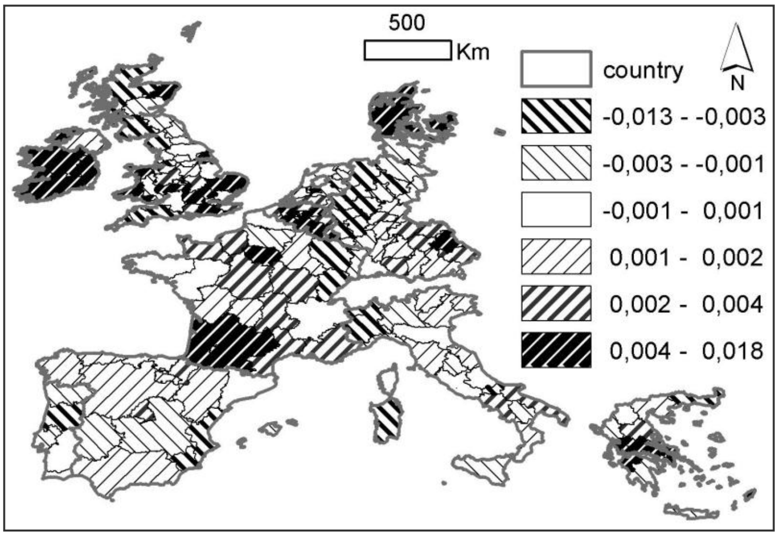


Figure 7: residuals of the spatial error model

Variable	OLS Est.	OLS S.D.	spatial error model Est.	spatial error model S.D.	OLS with country dummies Est.	OLS with country dummies S.D.
Intercept	-0.0092	0.0072	-0.030	0.0092	0.0038	0.0066
PROD.REL.80	-0.0091	0.0015	-0.011	0.0017	-0.016	0.0024
INV.RATE			0.012	0.010		
EMP.GR	-0.0057	0.0017	-0.0041	0.0016	-0.0047	0.0015
ECO.DEN	0.00041	0.00041				
SCF.on.GVA	0.13	0.068	0.28	0.062	0.19	0.057
SCF.on.GVA2	-0.78	0.44	-1.3	0.39	-0.87	0.35
D.SHARE.AGRI	0.015	0.012	0.043	0.012	0.024	0.012
CON.80	0.0057	0.017	0.018	0.014	-0.031	0.013
MIN.80	-0.023	0.0096	-0.023	0.0080	-0.026	0.0076
NON.MRKT.SER.80	-0.0029	0.0071	-0.018	0.0062	-0.027	0.0078
FIN.80	0.0038	0.023	0.055	0.020	0.035	0.018
TRANS.80	-0.00080	0.019			0.017	0.015
OTHER.SERV.80	0.040	0.012	0.042	0.011	0.029	0.011
BE					0.0053	0.0014
DK					0.0042	0.0035
ES					-0.0091	0.0019
FR					0.00056	0.0010
GR					-0.012	0.0025
LU					0.012	0.0039
IE					0.014	0.0031
IT					-0.0069	0.0014
NL					0.00034	0.0016
PT					-0.013	0.0034
UK					-0.0035	0.0018
Adjusted R^2	0.50				0.74	
LM error (df = 1)	8.69	p = 0.0032			2.31	p = 0.13
LM lag (df = 1)	0.027	p = 0.87			1.60	p = 0.21
robust LM error (df = 1)	11.00	p = 0.00091			1.07	p = 0.30
robust LM lag (df = 1)	2.33	p = 0.13			0.36	p = 0.55
Lambda			0.98			
LR test value			30.22	p: 3.86e-6		
Asymptotic standard err.			0.012	z: 79.38		
Log likelihood (error model)			7024.524			
AIC			-1376.9			

Table 2: results of the three models on European regional labour productivity growth

EU12 n=173	y (Growth Rate) estimate (s.e.)	VPC	Spatial error model estimate (s.e.)	VPC
country level	4.4 e-5 (2.0e-5)	64%	3.4 e-5 (1.5e-5)	71%
region level	2.5 e-5 (2.8e-6)	36%	1.4 e-5 (1.6e-6)	29%
constant	0.019 (0.002)		0.0022 (0.0018)	
-2 loglikelihood	-1308.932		-1405.201	

Table 3: Variance Partition Coefficients of the dependent variable and of the spatial error model residuals

3.2 *Modelling hierarchy: spatial multilevel modelling*

A multilevel regression model (Goldstein, 1995) has, next to a residual at the lowest level (e_{ij}), also a residual at a higher level (u_{oj}). More formally, this can be written as:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij} \tag{3}$$

$$\text{and } \beta_{0j} = \beta_0 + u_{0j} \tag{4}$$

with i = lowest level (e.g. region) and j = second level (e.g. country)

This model allows different level 2 units to have different intercepts and is therefore called random intercept model. The u_{0j}–terms are the level 2 random effects or the level 2 residuals. Multilevel modelling not only has the advantage of getting a better understanding and more clear interpretation of the effects of higher levels, but ignoring the fact that data are grouped often causes underestimated standard errors of regression coefficients (Maas and Hox, 2004; Rasbash et al., 2005).

Table 4 gives the results of two multilevel models with their corresponding empty models (models without variables, but with a hierarchical structure). The data are a selection out of the Belgian database Home-to-Work-Travel (Vanoutrive et al., 2009), and contain 2690 worksites where both carpooling and cycling occur. Since the models have three levels, the aforementioned model in equations (3) and (4) is extended with one extra level, but the principles remain the same. The standard approach to evaluate a multilevel model is looking at the residuals. The so-called caterpillar plots (Figure 8) can be derived for every level. A peculiar feature of spatial multilevel models is that these residuals can be mapped, like in Figures 9 and 10. The plots in Figure 8 show that the bicycle model largely explains the variance between districts, only three of the 43 districts significantly differ from zero.

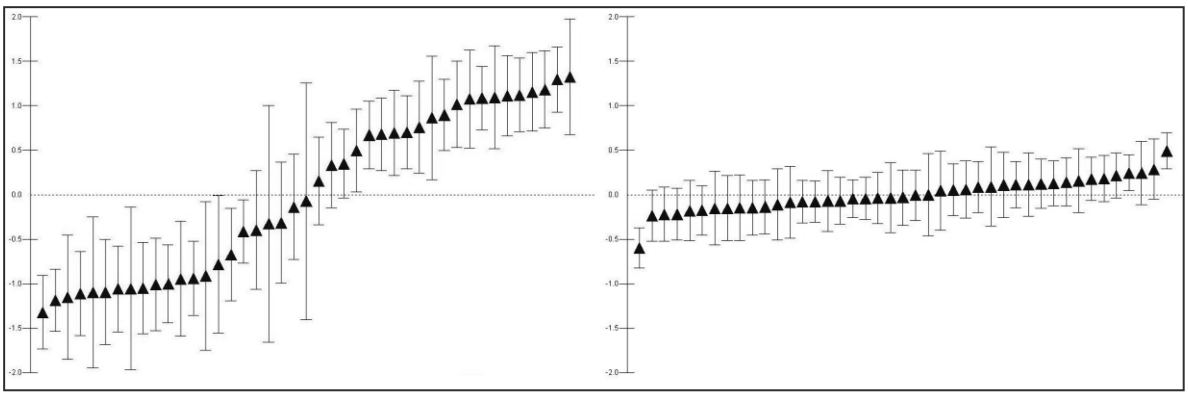
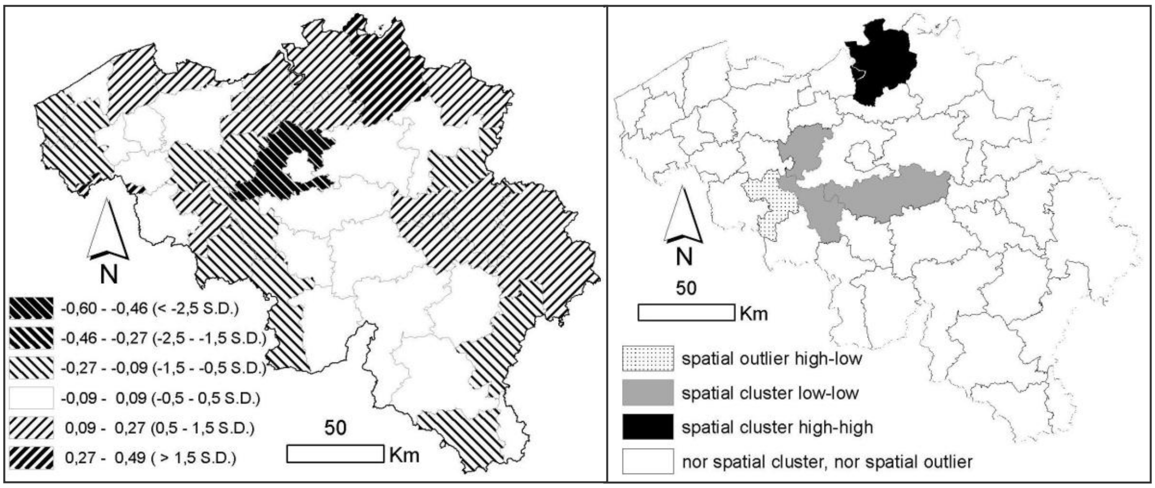


Figure 8: caterpillar plot: level 3 residuals of the empty (left) and full (right) bicycle models together with their 95% confidence intervals

	empty bicycle model		full bicycle model		empty carpool model		full carpool model	
random part	est.	s.d.	est.	s.d.	est.	s.d.	est.	s.d.
Level 3 (arrondissement) n= 43	0.90	0.21	0.06	0.02	0.005	0.002	0.00	0.00
Level 2 (municipality) n = 375	0.09	0.02	0.05	0.01	0.005	0.002	0.003	0.002
Level 1 (worksite) n = 2690	0.83	0.02	0.64	0.02	0.20	0.01	0.17	0.00
fixed part								
Constant	-3.00	0.15	-3.55	0.26	0.63	0.02	0.54	0.04
# Employees (log)			-0.31	0.04			-0.14	0.02
Car accessibility (log)			-1.45	0.45			-0.31	0.13
- Rail accessibility (log)			0.20	0.17			-0.22	0.07
Job Density (log)			-0.13	0.08			-0.03	0.03
Regular work schedules (log)			0.02	0.01			0.03	0.01
Slope (log)			-1.24	0.20			0.01	0.06
Flanders region			1.35	0.27			0.03	0.04
Walloon region			-0.21	0.28			0.06	0.04
economic sector dummy variables								
Manufacturing			-0.45	0.05			0.19	0.02
Electricity; gas; water			-0.56	0.13			0.06	0.07
Construction			-1.19	0.16			0.31	0.08
Retail			-0.55	0.07			0.04	0.04
Transport; communications			-1.10	0.08			0.16	0.04
Finance			-0.90	0.10			-0.01	0.05
Producer services			-0.73	0.09			0.13	0.04
Health			-0.12	0.06			-0.22	0.03
Higher Education			0.21	0.09			-0.25	0.05
Public Transport companies			-0.28	0.11			-0.21	0.05
Police			0.48	0.13			-0.21	0.07
Local Governments			0.51	0.06			-0.05	0.03
-2 Loglikelihood	7424.609		6603.185		3337.275		2948.272	

Table 4: results of the bicycle and carpool models



Figures 9 and 10: level 3 residuals of the full bicycle model (left) and the corresponding LISA map (right)

Both Chaix et al. (2005b) and Groenwegen et al. (1999) calculated the Moran's I for the residuals of a multilevel model to measure the unaccounted spatial autocorrelation. As Figures 9 and 10 show, the bicycle model does not eliminate all spatial autocorrelation at the district level (Moran's I = 0.11). However, the low level 3 variance of the full bicycle model reduces the importance of this 'residual spatial pattern'. The use of a third level does largely remove the spatial autocorrelation at the municipality level, since the estimates of this second level are made relative to the third level. Figure 11 shows the level 2 residuals of the carpool model and due to the model design, the average level 2 residual within an arrondissement approximates zero. As a consequence, detecting spatial patterns is not useful, but the map indicates the level 2 observations with a high or low value relative to the other observations in the same level 3 area. This can be used as a kind of performance index.

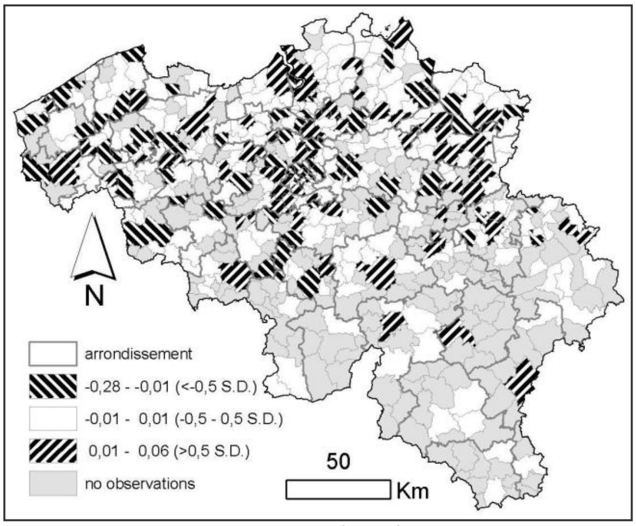


Figure 11: level 2 residuals of the full carpool model

It is important to notice that the residuals of these multilevel models are shrunken residuals. These residuals are not just the average of the lower level units, but also take into account the other observations of the entire dataset. This is an advantage for areas with a small number of observations, since they in general have a large standard deviation. In other words, the observed area mean is shrunken towards the centre of the whole population using a shrinkage factor (Rasbash et al., 2005). This shrinkage is relevant when discussing the use of country dummy variables. Langford et al. (1998) shows that fixed parameter estimates obtained by using dummies for areal units, can differ from the shrunken parameter estimates in a multilevel model. The desirability of this shrunk effect '*in any particular model is a matter of debate, but a general rule is the categorical variables included in the model as a level will produce shrunken, or conditional estimates of contrasts whilst including a categorical variable as a variable, for example, as a set of dummy vectors, will produce unconditional estimates'* (Langford et al., 1998). Including dummy variables for broad classes of spatial location can still be useful, but this not allow other coefficients in the model to vary spatially (Brunsdon et al., 1998). Moreover, a hierarchical spatial regression analysis can be enriched by creating a random slope model. The multilevel model in equations (3) and (4) can be extended towards such a random slope model, as given in equations (5) and (6). The level 2 regression lines have now an own slope. This kind of model is clearly superior to one with a myriad of dummy variables interacting with different other variables.

$$y_{ij} = \beta_0 + \beta_{1j}x_{ij} + u_{0j} + e_{ij} \tag{5}$$

$$\text{and } \beta_{1j} = \beta_1 + u_{1j} \tag{6}$$

4 A comparison between spatial econometrics and spatial multilevel modelling

By means of a comparison between the spatial econometrics and the spatial multilevel approach, we will now briefly discuss some relevant topics about regressions using spatial data. A multilevel structure can easily reflect administrative structures with different government levels, e.g. municipalities which are part of regions which are part of countries. An advantage of multilevel modelling is the possibility to incorporate more than two levels (spatial scales) in a straightforward way. However, a regression model can contain more than one, or a complex spatial weights matrix. But spatial multilevel modelling remains superior in modelling different scales simultaneously and in overcoming both the atomistic fallacy of individual-based studies, and the ecological fallacy of aggregated research.

Spatial econometrics on the other hand uses a unique set of neighbours for every observation, which preserves a relation between neighbouring regions separated by a national boundary. Indeed, economic processes often neglect administrative or other boundaries and many processes are influenced more by the distance between objects than by their hierarchical setting.

Spatial econometric models seldom account for missing neighbouring countries, like the absence of Switzerland in models of the European Union. As a result, some regions ‘jump’ over missing regions to find neighbours, or find neighbours in only one direction. A multilevel approach on the other hand, suffers less from this ‘missing data’ problem, but neglects all cross-boundary effects.

Both approaches deal with the 'Modifiable Areal Unit Problem' (MAUP), but while the spatial multilevel approach delivers a better understanding of the scale issue, the spatial econometrics approach better handles the zonation issue and the related spurious spatial autocorrelation. The difference between compositional and contextual data on the other hand, is merely a topic of spatial multilevel modelling.

Space remains a challenging thing to incorporate in statistical models since several distinct processes operate at several different spatial scales, scales which vary over space as well. We started this paper with a list of geographic concepts and while focussing on proximity and hierarchy, concepts like scale, distance and location were mentioned. It seems that regional scientists may not stick to a single spatial concept. We now will finish with some comments on the combination of the proximity and the hierarchy approaches.

5 Combining the two approaches

Both spatial econometrics and multilevel modelling are established regression techniques which are extended to other econometric approaches, like for instance multilevel SEM (Structural Equation Modelling; (Muthén and Muthén, 2006)) and spatial econometric SEM (Oud and Folmer, 2008). Also hybrid approaches exist., like Case (1991) who imposed a block structure in the weights matrix and formed as a result a hierarchical spatial model, where all units that share a common higher order level are considered to be neighbours. Anselin (2002) warns for side effects of this type of model and it seems preferable to use a standard multilevel model in stead of a modified spatial econometrics approach. A real combination of both approaches is Elhorst and Zeilstra (2007), and also Langford et al. (1999) incorporates spatial effects in a multilevel model. But the latter concludes: '*models can easily become very complex, and this is why we emphasize the need for hypotheses to be properly specified before modelling begins… …Complex models can easily be built, but less easily interpreted,…'*

Model parsimony remains an important evaluation criterion for models. As a consequence, complex models which incorporate both hierarchy and proximity must be treated with care. In many cases, accessibility variables can cover distance effects and as a result, remove the necessity to impose a hierarchical or spatial dependence structure on a model. In the carpool model, car and rail accessibility variables account for an important part of the spatial variation, as do the economic sector dummy variables, since different economic sectors have different spatial patterns. The use of country dummy variables is another example of an approach that can reduce model complexity. We showed in a model of labour productivity growth that a set of country dummy variables removed the need for a spatial error model. It seems that more effort should go not towards increasingly complex models, but towards the spatial assessment of models. Only few attempts are made to assess spatial multilevel models for spatial autocorrelation, or to look if a spatial econometric model ignores a hierarchical structure. This paper illustrated how the proximity and hierarchy approach can learn from each other.

6 Acknowledgements

We are very grateful to everybody involved in the 2008 Prepare ERSA summer school for creating a platform where we could discuss the topics dealt with in this paper.

Part of this research is conducted within the ADICCT-project (Assessing and Developing Initiatives of Companies to control and reduce Commuter Traffic) which is financed by Belgian Science Policy in the Science for a Sustainable Development research programme.

7 References

Anselin, L., Bera, A.K., Florax, R., and Yoon, M.J., 1996, Simple diagnostic tests for spatial dependence: Regional Science and Urban Economics, v. 26, p. 77-104.

Anselin, L., 2002, Under the hood - Issues in the specification and interpretation of spatial regression models: Agricultural Economics, v. 27, p. 247-267.

Anselin, L., 1988, Spatial econometrics: Methods and models: Dordrecht, Kluwer Academic Publishers, 308 pp.

Anselin, L., 2001, Spatial econometrics, *in* Baltagi, B., ed., A Companion to Theoretical Econometrics: Oxford, Blackwell, p. 310-330.

Anselin, L., 1995, Local Indicators of Spatial Association-LISA: Geographical Analysis, v. 27, p. 93-115.

Anselin, L., 2005, Exploring Spatial Data with GeoDa^{TM} : A Workbook: Urbana, Center for Spatially Integrated Social Science, 226 pp.

Anselin, L., Florax, R.J.G.M., and Rey, S.J., 2004, Advances in spatial econometrics: Methodology, tools and applications: Berlin, Springer-Verlag, 513 pp.

Arbia, G. and Fingleton, B., 2008, New spatial econometric techniques and applications in regional science: Papers in Regional Science, v. 87, p. 311-317.

Bera, A. and Yoon, M.J., 1993, Specification testing with misspecified alternatives: Econometric Theory, v. 9, p. 649-658.

Bosker, M., 2009, The spatial evolution of regional GDP disparities in the 'old' and the 'new' Europe: Papers in Regional Science, v. 88, p. 3-27.

Brunsdon, C., Fotheringham, S., and Charlton, M., 1998, Geographically Weighted Regression-Modelling Spatial Non-Stationarity: The Statistician, v. 47, p. 431-443.

Case, A.C., 1991, Spatial Patterns in Household Demand: Econometrica, v. 59, p. 953-965.

Chaix, B., Merlo, J., Subramanian, S.V., Lynch, J., and Chauvin, P., 2005a, Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmo, Sweden, 2001: American Journal of Epidemiology, v. 162, p. 171-182.

Chaix, B., Merlo, J., and Chauvin, P., 2005b, Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in France: Journal of Epidemiology and Community Health, v. 59, p. 517-526.

Ciriaci, D. and Palma, D., 2008, The role of knowledge-based supply specialisation for competitiveness: A spatial econometric approach: Papers in Regional Science, v. 87, p. 453-476.

Dall'erba, S. and Le Gallo, J., 2008, Regional convergence and the impact of European structural funds over 1989-1999: A spatial econometric analysis: Papers in Regional Science, v. 87, p. 219-244.

Duncan, C., Jones, K., and Moon, G., 1998, Context, composition and heterogeneity: Using multilevel models in health research: Social Science & Medicine, v. 46, p. 97-117.

Elhorst, J.P. and Zeilstra, A.S., 2007, Labour force participation rates at the regional and national levels of the European Union: An integrated analysis: Papers in Regional Science, v. 86, p. 525-549.

Fellman, J.D., Getis, A., and Getis, J., 1999, Human Geography: Boston, WCB/McGraw-Hill.

Fiaschi, D., Lavezzi, A.M., and Parenti, A., 2009, Productivity Convergence across European Regions: the Impact of Structural and Cohesion Funds: Pisa, Discussion Papers del Dipartimento di Scienze Economiche - Università di Pisa, n. 84, 36 pp.

French, K.M. and Jones, K., 2006, Impact of definition on the study of avoidable mortality: Geographical trends in British deaths 1981-1998 using Charlton and Holland's definitions: Social Science & Medicine, v. 62, p. 1443-1456.

Geppert, K. and Stephan, A., 2008, Regional disparities in the European Union: Convergence and agglomeration: Papers in Regional Science, v. 87, p. 193-217.

Goldstein, H., 1995, Multilevel Statistical Models, London, Arnold.

Griffith, D.A. and Layne, L.J., 1999, A casebook for spatial statistical data analysis: US, Oxford University Press, 524 pp.

Groenewegen, P.P., Leufkens, H.G., Spreeuwenberg, P., and Worm, W., 1999, Neighbourhood characteristics and use of benzodiazepines in The Netherlands: Social Science & Medicine, v. 48, p. 1701-1711.

Johnston, R., Jones, K., Propper, C., and Burgess, S., 2007, Region, Local Context, and Voting at the 1997 General Election in England: American Journal of Political Science, v. 51, p. 640-654.

Kwan, M.P. and Weber, J., 2008, Scale and accessibility: Implications for the analysis of land use-travel interaction: Applied Geography, v. 28, p. 110-123.

Langford, I.H., Bentham, G., and McDonald, A.-L., 1998, Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European Community: Statistics in Medicine, v. 17, p. 41-57.

Langford, I.H., Leyland, A.H., Jon, R., and Goldstein, H., 1999, Multilevel Modelling of the Geographical Distributions of Diseases: Applied Statistics, v. 48, p. 253-268.

Le Gallo, J. and Dall'erba, S., 2008, Spatial and sectoral productivity convergence between European regions, 1975-2000: Papers in Regional Science, v. 87, p. 505-526.

Legendre, P., 1993, Spatial Autocorrelation: Trouble or New Paradigm?: Ecology, v. 74, p. 1659-1673.

Luke, D.A., 2004, Multilevel Modeling, Thousand Oaks, Sage, 79 pp.

Maas, C.J.M. and Hox, J.J., 2004, Robustness issues in multilevel regression analysis: Statistica Neerlandica, v. 58, p. 127-137.

Manley, D., Flowerdew, R., and Steel, D., 2006, Scales, levels and processes: Studying spatial patterns of British census variables: Computers, Environment and Urban Systems, v. 30, p. 143-160.

Miller, H.J., 2004, Tobler's First Law and Spatial Analysis: Annals of the Association of American Geographers, v. 94, p. 284-289.

Mohan, J., Twigg, L., Barnard, S., and Jones, K., 2005, Social capital, geography and health: a small-area analysis for England: Social Science & Medicine, v. 60, p. 1267-1283.

Muthén, L.K. and Muthén, B.O., 2006, Mplus User's Guide: Los Angeles.

Olejnik, A., 2008, Using the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU25: Papers in Regional Science, v. 87, p. 371-385.

Orford, S., 2000, Modelling Spatial Structures in Local Housing Market Dynamics: A Multilevel Perspective: Urban Studies, v. 37, p. 1643-1671.

Oud, J.H.L. and Folmer, H., 2008, A structural equation approach to models with spatial dependence: Geographical Analysis, v. 40, p. 152-166.

Pumain, D. and Saint-Julien, T., 2001, Les interactions spatiales: Paris, Armand Colin.

Quah, D.T., 1996, Regional convergence clusters across Europe: European Economic Review, v. 40, p. 951-958.

Rasbash, J., Steele, F., Browne, W., and Prosser, B., 2005, A User's Guide to MLwiN Version 2.0: Bristol, Centre for Multilevel Modelling University of Bristol. 256 pp.

Schwanen, T., Dieleman, F.M., and Dijst, M., 2004, The Impact of Metropolitan Structure on Commute Behavior in the Netherlands: A Multilevel Approach: Growth and Change, v. 35, p. 304-333.

Subramanian, S.V., Duncan, C., and Jones, K., 2001, Multilevel perspectives on modeling census data: Environment and Planning A, v. 33, p. 399-417.

Tranmer, M. and Steel, D.G., 2001, Ignoring a level in a multilevel model: evidence from UK census data: Environment and Planning A, v. 33, p. 941-948.

Tselios, V., 2008, Income and educational inequalities in the regions of the European Union: Geographical spillovers under welfare state restrictions: Papers in Regional Science, v. 87, p. 403-431.

Uthman, O.A., 2008, Geographical variations and contextual effects on age of initiation of sexual intercourse among women in Nigeria: a multilevel and spatial analysis: International Journal of Health Geographics, v. 7.

Vanoutrive, T., Van Malderen, L., Jourquin, B., Thomas, I., Verhetsel, A., and Witlox, F., 2009, Carpooling and employers: a multilevel modelling approach, *in* Macharis, C. and Turcksin, L., eds., proceedings of the BIVEC-GIBET Transport Research Day. Brussels, VUBPRESS, p. 335-349.