Chapter 12
Selecting Spatial Regimes by Threshold Analysis

The existence of differential spatial regimes has been revealed on different occasions (see for instance Arbia and Paelinck, 2003a, b; also see Chap. 14). Hence the necessity exists for developing workable specifications to compute possible frontiers or thresholds between those regimes.

The next section describes one possible method. Sects. 12.2 and 12.3 then apply it to two spatial models using Dutch data: an income generating model, and an activity complex model.

12.1 Method

Assume the following model

$$y = ax, \tag{12.1}$$

$$a = a_1 \mid x \leq x^*, \tag{12.2}$$

$$a = a_2 \mid x \geq x^*. \tag{12.3}$$

The model can be respecified as follows, for each observation *i*

$$y_i = a_1 u_i x_i + a_2 (1 - u_i) x_i + \varepsilon_i, \tag{12.4}$$

$$(u_i - \eta)(x_i - x^*) \leq 0, 0 < \eta < 1, \tag{12.5}$$

$$u_i = u_i^2, \tag{12.6}$$

the estimation being, e.g., performed by minimizing $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$. This model was first tested on the following data, appearing in Table 12.1.

The values for a_1 and a_2 are respectively *2* and *1*, the first four observations being governed by a_1. The value of x^*, which is not unique, should lie between *8* and *11*. Estimation reproduced the values of a_1 and a_2, $x^* = 8{,}1333$, with the u_is correctly

D.A. Griffith, J.H.P. Paelinck, *Non-standard Spatial Statistics and Spatial Econometrics*, Advances in Geographic Information Science 1,
DOI 10.1007/978-3-642-16043-1_12,

Table 12.1 Test data for the model defined by Eqs. (12.4)–(12.6)

Variable	Values							
y_i	4	10	12	16	11	13	14	17
x_i	2	5	6	8	11	13	14	17

partitioning the observations. The objective function adopted the value zero, and all restrictions and optimality conditions were satisfied.

If one wants to replace Eq. (12.2) or (12.3) by a strict inequality, the following specification could be used

$$(u_i - \eta)(x_i - x^* - \theta) \geq 0 \tag{12.7}$$

where η is defined as in Eq. (12.5), and θ is an appropriately chosen small positive number. If $x_i - x^* > \theta$, $u_i = 1$; if $x_i - x^* = \theta$, u_i could be equal to either *0* or *1*, whichever value would give the best estimation result; if $x_i - x^* < \theta$, $u_i = 0$. In both Eqs. (12.5) and (12.6) the specification (parameter η) prevents u_i from being zero if the second factor of Eq. (12.7) is non-zero (positive or negative, depending on the case).

12.2 Spatial Income Generating Model

This model was initially developed in Paelinck and Klaassen (1979, pp. 21–23). Its aim is to measure the spatial interdependence between regional incomes or products.

Let **y** be the column vector of regional incomes; then the model in its simplest form is specified as

$$\mathbf{y} = \mathbf{Ay} + \mathbf{b}, \tag{12.8}$$

where matrix **A** integrates some spatial interaction operator. In the present case, a first-order contiguity matrix, $\mathbf{C_1}$, has been selected for matrix **A** to compute total incomes over neighboring regions.

As can be seen from Fig. 12.1, in the Netherlands all regions, except for the provinces of Flevoland and Utrecht, are peripheral, with three purely maritime (Friesland, Noord- and Zuid-Holland), so five "pseudo-border" correcting (additive) parameters have been introduced into the model (Paelinck, 1996b, pp. 4–8). Moreover, the "reaction" parameters *a* and *b* have been split, according to Eq. (12.4) together with constraints (see Appendix).

Table 12.2 presents the data for regional (provincial) products (1987, Dfl 10^6; source: van Gastel and Paelinck, 1995, p. 152). Table 12.3 presents the spatial contiguity structure (degrees of contiguity) for the Netherlands (same source).

Fig. 12.1 Map of the Dutch provinces

Table 12.2 Provincial products in the Netherlands, 1987

Province	Abbreviation	Number	Product
Groningen	Gr	1	22,675
Friesland	Fr	2	13,229
Drente	Dr	3	11,187
Overijssel	Ov	4	25,448
Flevoland	Fp	5	3703
Gelderland	Gl	6	43,861
Utrecht	Ut	7	27,801
Noord-Holland	NH	8	77,997
Zuid-Holland	ZH	9	102,864
Zeeland	Zl	10	10,908
Noord-Brabant	NB	11	59,242
Limburg	Lb	12	29,036

Table 12.3 Contiguity structure of the Netherlands

Province	Gr	Fr	Dr	Ov	Fp	Gl	Ut	NH	ZH	Zl	NB	Lb
Gr	0	1	1	2	2	3	3	3	4	5	4	4
Fr	1	0	1	1	1	2	2	2	3	4	3	3
Dr	1	1	0	1	2	2	3	3	3	4	3	3
Ov	2	1	1	0	1	1	2	2	2	3	2	2
Fp	2	1	2	1	0	1	1	1	2	3	2	2
Gl	3	2	2	1	1	0	1	2	1	2	1	1
Ut	3	2	3	2	1	1	0	1	1	2	2	2
NH	3	2	3	2	1	2	1	0	1	2	2	3
ZN	4	3	3	2	2	1	1	1	0	1	1	2
Zl	5	4	4	3	3	2	2	2	1	0	1	2
NB	4	3	3	2	2	1	2	2	1	1	0	1
Lb	4	3	3	2	2	1	2	3	2	2	1	0

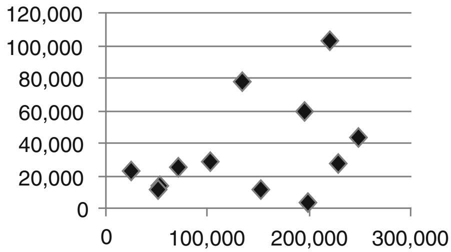

Fig. 12.2 Specifying condition (12.5): the $x_r - y_r$ relation

A typical equation for a region *r* now may be written as follows

$$y_r = a_1 u_r x_r + a_2(1 - u_r)x_r + b_1 z_r + b_2(1 - z_r) + c_r + \varepsilon_r, \qquad (12.9)$$

where y_r denotes a regional product and x_r the sum of products in neighboring regions (here divided by *10*, for reasons of magnitude similarity), u_r and z_r are binary variables, and c_r denotes the “pseudo-border” coefficients previously mentioned.

Figure 12.2 hereafter has served to specify the condition referred to in Eq. (12.5). It suggests replacing x_i of Eq. (12.5) by the ratio x_r / y_r.

Table 12.4 presents the results of the estimation procedure (derived by means of a constrained gradient method; see Fylstrom et al., 1998). Model (12.8) including interdependent endogenous variables, an SDLS estimation procedure (see Chap. 11, Sect. 11.1.3) was used, whose optimization principle, as said there, is the minimization of the sum of squared differences between the observed and the endogenously computed y_r variables.

Rather than indicating two regimes, all four combinations of the reaction parameters are present: a_1–b_1 (four times), a_2–b_1 (five times), a_1–b_2 (twice) and a_2–b_2 (once, the case of Flevoland, a recent small new province). The value of the threshold is *0.3486*. Corrective constants c_r are positive for Noord- and Zuid-Holland (two

Table 12.4 Results of the estimation procedure of model (12.8)

Province	$a_1 = 3.8934$	$a_2 = 0.8328$	$b_1 = 10378$	$b_2 = -8276$	c_r	y_r (est.)
Gr	x		x			24,145
Fr		x	x		220	10,774
Dr		x	x			16,598
Ov	x		x			22,945
Fp		x		x	−2473	7052
Gl		x	x			41,250
Ut		x	x		−46	36,258
NH	x		x		3074	77,518
ZH	x		x		1552	102,061
Zl		x	x			10,208
NB	x			x		57,831
Lb				x		32,260

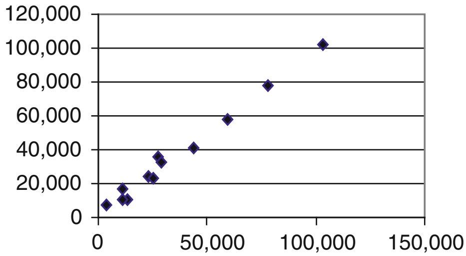

Fig. 12.3 Observed and computed values of y_r

heavily exporting provinces, with important harbors), and negative for Flevoland (see the remark above), the other corrections being negligible.

Figure 12.3 compares the observed and computed values of the endogenous variables; Theil's *U* (Theil, 1961, p. 32) has value *0.0380*, showing the close connection between observed and computed values.

12.3 A Spatial Activity Complex Model

To subject the model to a more terse test, a so-called "attraction model" (first developed by Klaassen; see Paelinck and Klaassen, 1979, pp. 23–30) was evaluated.

Let y_{ir} represent the production level (or value added) of activity sector *i* in region *r*; the model here is specified as

$$y_{ir} = \sum_j a_{ijl}\, y^*_{jr} + b_{il}, \qquad 12.10$$

where the y^*_{jr} represent spatially discounted (from Table 12.3) aggregations (potentials) of activity sector production levels (index *j*). The index *l* denotes the relevant regime to which the parameters belong.

Table 12.5 Endogenous activity sectors

1. Industry, public utilities and minerals.
2. Building and construction activities.
3. Trade, catering and repairs.
4. Transport and communication activities.
5. Banking and insurance.
6. Realty and business services.
7. Health care and veterinary services.
8. Cultural, sports and recreational activities.
9. Other services.

Nine endogenous sectors were employed (Table 12.5) for the Netherlands case study.

Appendix 12.5 reports their production (value added) levels. The following sectors have been considered as exogenous: agriculture, forestry and fishery (10), crude oil and gas plants (11), public sector (12).

Because only 12 observations per activity sector are available, aggregated explanatory variables were constructed, whereby contiguous regions have been exogenously discounted at 50%, the aggregation being the following ones: 10, 11; 1, 2; 3 through 9, 12.

Figures 12.4 through 12.7 portray the $y_{ir} / y^*{}_{jr}$ ratios (inverses of the ratios used in Sect. 12.2) and the $y_{ir} / \sum_r y_{ir}$ one for sector 1; they suggest (at least) three regimes.

Forty-eight binary variables are to be used for only 2 regimes, generating a heavy *0–1* mathematical program; no solution could be reached within 40 hours, so a different strategy was developed.

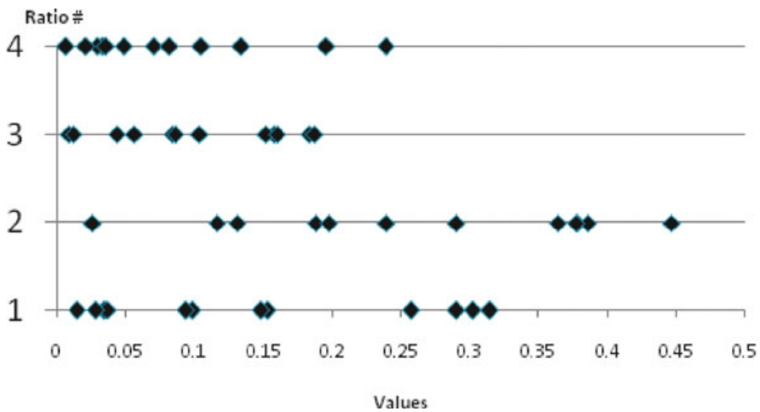

Fig. 12.4 Distribution of xt/yt ratios fot the four Dutch regions

For n observations and k relevant parameters, combining the relative frontiers between the values of Figs. 12.4, 12.5, 12.6, and 12.7, generates $[n(n-1)/2]^k$ simple quadratic programs. In the present exercise the cut-off points have been selected visually, so sub-optimal outcomes might be expected.

Coming back to parameters $a_{ijl,}$ and b_{il} of Eq. (12.10), some outliers might produce reverse results compared to a general trend as revealed, e.g., in Fig. 12.2; therefore an extra condition was introduced, to wit the sign equality of parameters a_{ijl},, $\forall l$. A way to impose that constraint would be

$$\mathrm{abs}\left(\sum_{l} \mathrm{a_{ijl}}\right) = \sum_{l} \mathrm{abs}\ (\mathrm{a_{ijl}}), \tag{12.11}$$

or, alternatively, binary conditions such as

$$\mathrm{a_{ijl}}{}^{*}\mathrm{a_{ijm}} \geq 0, \tag{12.12}$$

which would lead to a quadratic program with non-linear constraints. In practice, and in order to obtain classical results, necessary conditions [like those exposed through Eqs. (12.11) or (12.12)] have been replaced by inspecting all possible sign combinations (i.e., $2^3=8$, as the level parameters, b_{il}, were left free).

Finally, to allow for comparability of results, logarithms are used, yielding non-dimensional elasticities as parameters. Table 12.6 presents the estimation results.

One finding is that only two sectors (1 and 9) satisfy without constraint the equal sign condition; they also show great similarity inside each parameter group, and, as should be expected, the lowest U-values. Sectors with one active constraint number four, sectors with two number two, and only one sector needed all three constraints. The sectors needing two constraints also show the highest U-values (the only ones exceeding 0.05), even higher than the sector needing all three constraints.

At this stage, no SDLS-computations were performed, despite the interdependent specification of the model. Seven out of nine U-values are sufficiently low that no further corrections were deemed necessary (the two high U-values are probably related, to regional accountancy for sector 5, and the requirement of another specification, in terms of explanatory variables, for sector 8).

No significance magnitudes have been shown. However, they could be computed deleting the series of variables corresponding to the zero values.

Moreover no detailed study was made of the different combinations of a_{ijl} and b_{jl} parameters as presented in Table 12.6 (nine tables would be necessary for such a comparison). From that table, however, a remarkable finding can be derived, namely that parameters $a_{2\,l}$ are all non-negative in the solution. As for the other a_{il} parameters, five out of nine are non-positive, but in this exploratory study no detailed analysis has been made of the various cases.

Table 12.6 Estimation results of model (12.10)

S\P	a_{11}	a_{12}	a_{13}	a_{21}	a_{22}	a_{23}	a_{31}	a_{32}	a_{33}	b_1	b_2	b_3	U
1	−2.1819	−2.9368	−3.3316	3.7595	4.4289	4.8247	0.3968	0.4027	0.3721	−17.14	−17.01	−17.12	0.0038
2	1.2198	1.7931	0	3.2543	3.3275	3.4147	−2.6764	−2.7218	2.7692	−7.1934	−10.757	3.8513	0.0432
3	−2.4663	−3.0223	−2.9632	1.1667	3.2278	3.2312	1.4375	0	0.2275	−0.8157	0.1089	−2.3046	0.0248
4	0.2763	0.4163	0.6399	1.2082	1.2566	0	−1.7744	−1,7890	0.0685	9.6793	10.868	10.0858	0.0404
5	−2.7904	−3.4291	0	3.3726	3.3467	0.6926	−6.2102	−0.0158	0	55.907	−1.3444	−0.0898	0.0861
6	−0.0149	0	−0.2519	1.0556	0.1847	0	0	0.9497	1.1796	−3.4719	−3.8712	−1.9225	0.0202
7	1.5117	0	.2852	2.9476	1.1018	1.1558	−2.1952	−0.4717	−0.4853	−11.57	1.3841	−0.4828	0.0073
8	0	1.1709	1.1709	0.0560	4.1141	4.2962	−.0645	0	−3.6705	4.2579	−3.6466	−4.1888	0.1039
9	−0.7343	−0.8851	−0.8655	1.3096	1.4254	1.4840	1.2425	1.4705	1.6052	−14.53	−16.26	−17.75	0.0002

12.4 Conclusion

A workable method to flexibly select parameter regimes has been presented. It is a member of a class of non-standard estimators, many of which will be used in further spatial econometric work. They are indispensable companions of non-standard specifications that will also be required in the field of spatial econometrics.

Recent experience has indeed shown that systems of regions often reveal two regimes. In Coutrot et al. (2009), the introduction of a second regime lifted the R^2 from 0.5156 to 0.9990. Moreover, the regions were behaviorally very different, which separated the main regional activity poles from the minor centers. Another example will be provided in Chap. 14.

In the light of this, still limited, experience, it seems that an appropriate specification-cum-estimation strategy is to systematically test for the presence of multiple regimes.

12.5 Appendix

Activity production levels

R\S	1	2	3	4	5	6	7	8	9	10	11	12
1	1700	408	951	649	18	1307	605	104	451	294	3691	1067
2	1507	481	1051	418	96	1320	458	159	373	598	515	879
3	1085	316	814	245	33	1031	396	95	293	335	625	664
4	3644	981	2273	818	39	2505	889	187	657	599	99	1663
5	345	129	570	93	2	609	138	45	212	352	0	294
6	5399	1519	4316	1310	291	5034	1631	397	1451	922	29	3318
7	2491	890	3473	1127	473	4138	1202	281	1211	274	4	2032
8	6928	1864	7454	4423	502	8594	2346	957	2154	834	84	4483
9	12,323	2846	8782	5058	881	10,947	2930	830	2914	2121	182	6547
10	1841	301	802	408	9	770	264	64	239	244	0	594
11	10,093	2040	5866	1814	178	6842	1815	496	1481	1119	46	3411
12	4225	786	2308	960	177	2729	1100	237	726	580	5	1632

Source: CBS, StatLine; numbers relate to 1993 and are expressed in 10^6 EUROS