Chapter 16
Qualireg, A Qualitative Regression Method

Circumstances can produce themselves under which no cardinal data are available (see Ancot and Paelinck, 1990). To allow drawing inferences about at least the direction of influence of certain potentially explanatory variables, only available as ordinal data ("rankings"), methods should be developed to treat that problem. The method described here—*QUALIREG*—resulted from work on a qualitative multi-criteria method—*QUALIFLEX*, originated by Paelinck (1976)—which is detailed first, after which the logic of *QUALIREG* will be introduced.

A first application to test the method is then presented, followed by a typical spatial econometric one, to wit estimating first- and second-order contiguity effects.

16.1 Qualiflex

Suppose three objects, O, to be ranked according to three criteria, C, along which they can initially only be separately ranked. Table 16.1 presents such a case.

The relative importance of the criteria is known only in an ordinal manner, i.e., again their ranking (vector **w**).

The optimal ranking of O_1 through O_3 is to be derived out of the $3! = 6$ possible rankings. Those rankings can be classified according to elementary permutations (i.e., permutations of neighboring objects). Table 16.2 presents those rankings, starting from [+++, ++, +], with the rankings being denoted by R_i.

A possible measure of the agreement of two rankings is a so-called rank correlation coefficient, denoted τ, and having values in [–1, 1], much like an ordinary simple correlation coefficient. The easiest choice for attributing values is to divide the *interval* in equal parts, in this case *0.66*, and hence obtain the values shown in the last column of Table 16.2. The values of τ are computed with respect to R_1.

The observed rankings can be laid out in matrix form as Table 16.3 shows for the three rankings of Table 16.1; a descending order is rated *+1*, an ascending one *–1*.

These tables should obviously be skew-symmetric, i.e., the absolute values of symmetric terms are equal, but their signs are opposite. τ-values based on these tables can be computed with respect to R_1 (e.g., the upper-triangular sum of the first ranking of Table 16.3 is *1*, which divided by the sum of all positive scores gives *0.33*,

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_16,

Table 16.1 A qualitative multicriteria table

w (rankings)	C\O	O_1	O_2	O_3
+++	C_1	++	+++	+
++	C_2	++	+	+++
+	C_3	+++	+	++

Table 16.2 Elementary permutations for three elements

R_1	+++	++	+	1
R_2	++	+++	+	0.33
R_3	+++	+	++	0.33
R_4	++	+	+++	–0.33
R_5	+	+++	++	–0.33
R_6	+	++	+++	–1

as in Table 16.2). That part of Table 16.3 also reveals that the maximum correlation coefficient may be observed permuting O_1 and O_2 appears, as the ranking then is identically R_1. This that is the very clue that led to the method that follows.

The idea is to find a ranking that has maximum (possibly weighted) correlation with—or, alternatively, minimum (possibly weighted) so-called Kendall distance (see Paelinck, 1985, pp. 80–98; it is a linear transform of Kendall's rank correlation coefficient; see Sect. 16.2) to—the observed individual rankings This implies constructing a new table—or matrix—from the observed ones by summing them with the appropriate weights, and then—by permuting rows and columns—obtaining a maximum upper-triangular sum. Consider the three parts of Table 16.3, and suppose all weights to be equal. Then the resulting sum table (here the tables may be simply added up) is Table 16.4.

The ranking presented in Table 16.1 is the optimal one. But what if the criteria are only qualitatively ranked? A first possibility is to inspect the weight triangle (or, for higher dimensions, the hyper-triangle). Its endpoints are *(1, 0, 0), (0.5, 0.5, 0)*

Table 16.3 Rankings fot three criteria

Ranking		O_1	O_2	O_3
First	O_1	0	–1	1
	O_2	1	0	1
	O_3	–1	–1	0
Second	O_1	0	1	–1
	O_2	–1	0	1
	O_3	1	–1	0
Third	O_1	0	1	1
	O_2	–1	0	–1
	O_3	–1	1	0

Table 16.4 Sum table from Table 16.3

O_i	O_1	O_2	O_3
O_1	0	1	1
O_2	–1	0	1
O_3	–1	–1	0

and *(0.33, 0.33, 0.33):* see Fig. 16.1. This figure is a two-dimensional cut through the three-dimensional space generated by the weight axes.

For point *(1, 0, 0)* the optimal ranking is (O_2, O_1, O_3), (see Table 16.1). For point *(0.5, 0.5, 0)* the optimal ranking can be calculated (by adding part 1 and 2 of Table 16.3) to be (O_1, O_2, O_3), and this ranking is optimal again in point *(0.33, 0.33, 0.33)* as previously mentioned. The conclusion is that for a relatively high weight attributed to C_1, the optimal ranking would be (O_2, O_1, O_3), with O_2 the "best" (first in rank) object; for lower C_1-weights, (O_1, O_2, O_3) would be optimal with C_1 the "best" object.

In practice, one can randomly scan the weight triangle and find out the zones where certain rankings are optimal. Anyway, in each of those points a matrix permutation is necessary, but it can be shown that this is equivalent to a quadratic assignment problem.

For applications, one can consult Ancot and Paelinck (1982, 1985 and 1986).

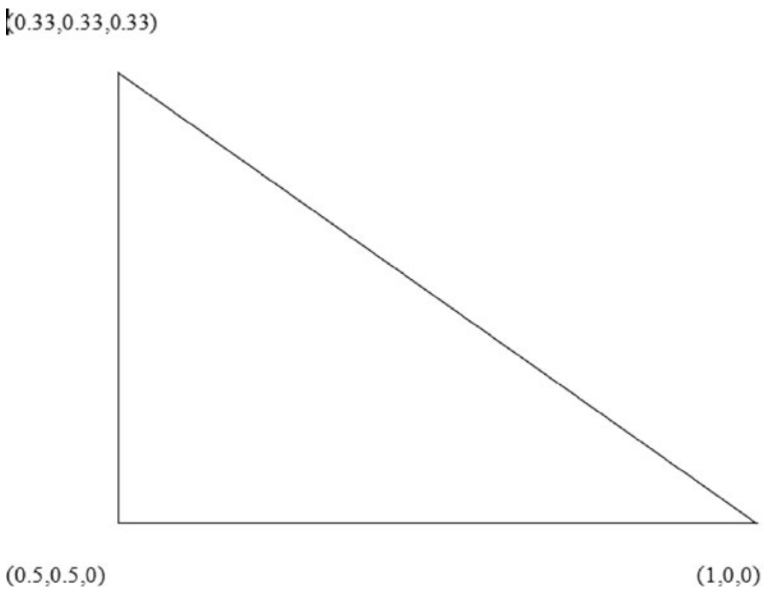

Fig. 16.1 Weight triangle for three criteria

16.2 Qualireg

The problem studied in Sect. 16.1 was to derive an optimal ranking for given rankings and criteria weights. One could ask whether the inverse problem—derive the weights given the final ranking and the initial rankings—has a meaning.

The solution to this optimal ranking problem is equivalent to qualitative regression (for first results, again see Ancot and Paelinck, 1986). Consider the equation

$$\mathrm{y_i = a\,x_i + b\,z_i + c}, \tag{16.1}$$

in which parameters a and b have to be estimated. In this case, y_i, x_i and z_i would be elements of three rankings of the respective variables.

Three rank correlation coefficients, $\tau(y, x)$, $\tau(yz)$ and $\tau(xz)$ can be calculated from these rankings (see Kendall, 1955). These correlation coefficients can be used in the classical regression parameter estimation equation, yielding

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & \tau_{xz} \\ \tau_{xz} & 1 \end{bmatrix} \begin{bmatrix} \tau_{yx} \\ \tau_{yz} \end{bmatrix} \tag{16.2}$$

The method has been applied to the data reported in Table 16.5 (source: Plante, 2005).

The data of Table 16.5 have been reduced to their rankings, and the above method applied, rendering the results reported in Table 16.6, which are compared with the original OLS estimates). Their analysis only shows the workings of QUALIREG, with no spatial effects, save distances, having been introduced.

Here the signs of OLS and QUALIREG correspond, but not the relative magnitudes. This result is due, on the one hand to the "data reduction"—"impoverishment"—previously mentioned, and on the other hand to standardization

Table 16.5 Data for a QUALIREG application

Observations	y	x	z
1	2533	53	19
2	962	18	28
3	426	33	35
4	7226	60	3
5	94	20	46
6	411	17	42
7	101	21	61
8	102	27	70
9	27	19	68
10	158	23	69
11	76	24	63
12	269	16	62

Note: y stands for population densities in Northern Virginia counties, x for the share of non-agricultural activity in total activity, and z for distance from Washington

Table 16.6 QUALIREG and OLS estimation results compared

Parameter region	a	b	R^2	t_a	t_b
OLS	0.66	–0.43	0.79	2.50	–2.23
QUALIREG	0.23	–0.66	0.54	0.34	–0.96

of the QUALIREG estimators (i.e., division by the appropriate standard errors, as correlation coefficients have been used). Moreover, for QUALIREG the parameters R^2, t_a and t_b are "pseudo" test criteria, because they can be computed only by analogy. These estimators were obtained as follows.

From general OLS analysis

$$R^2 = \mathbf{y'X'(X'X)^{-1}Xy/y'y}. \tag{16.3}$$

For normalized variables **y** and two τ-variables, this equation translates into

$$\mathbf{R^2} = \begin{bmatrix} \tau_{y,x_1}; \tau_{y,x_2} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \tau_{x_1,x_2} \\ \tau_{x_1,x_2} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \tau_{y,x_1}; \tau_{y,x_2} \end{bmatrix} \tag{16.4}$$

and

$$\sigma = (1 - R^2)^{.5}. \tag{16.5}$$

Applying this result to the numbers obtained renders

$R^2 = 0.5388$
$\sigma = 0.6791$
$t_a = 0.3393$
$t_b = -0.9631$

As noted previously, these are all "pseudo-values", because they are simple "analogs" to the theoretical ones. In this case, though R^2 might be significant, a and b are not; this finding is to be expected, as the original data have been "impoverished".

The method finally allows use of even very poor data in order to assess at least the signs (directions) of the partial relations hypothesized.

16.3 Spatial Setting

Next consider spatial interaction effects.

In order to address these effects, a classical spatial specification has been selected, to wit the estimation of first- and second-order contiguity effects. An interdependent linear specification has been selected, which has first been estimated by

Simultaneous Dynamic (here Spatial) Least Squares with endogenously estimated computed values (see Sect. 11.1.3).

The model is specified as follows

$$\mathbf{y} = a\mathbf{C_1 y} + b\mathbf{C_2 y} + c\mathbf{i}, \tag{16.6}$$

where $\mathbf{y}$ is a vector of regional products, $\mathbf{C_1}$ and $\mathbf{C_2}$ are the first- and second-order contiguity matrices, $\mathbf{i}$ is the unit vector, and *a, b* and *c* are parameters to be estimated.

The application concerns the regional products of 11 Belgian regional units, presented in Table 16.7; with their products (in millions of Euros) for 2002, and also with the average contiguity products of order 1 and 2. Only 11 units appear, as the extra-territorial units have been included in the "Brussels Capital" region. The data quantities and map are those of Chap. 14 and, for the contiguity degrees, they are taken from Kaashoek et al. (2004; see Table 16.8).

Table 16.7 Belgian regional units and products, 2002

Number	Unit	Product	C_1	C_2
1	Antwerp	41,483.5	21,307	21,590.1
2	Walloon Brab.	7,639	19,324.2	25,289.1
3	Flem. Brab.	23,232.5	23,908	11,118
4	East Fland.	26,070.5	26,395.8	17,690.7
5	West Fland.	22,766	22,085.9	19,776.8
6	Limburg	14,617.9	27,118.3	17,534
7	Hainaut	18,101.2	17,292.1	23,876.3
8	Namur	6,752.3	11,553.7	21,671.7
9	Luxemburg	3,835.8	11,695.6	15,897.7
10	Liège	16,638.8	11,215.5	32,115.2
11	Bruss. Cap.	42,805.4	23,232.5	23,382.4

Table 16.8 First and second order contiguity degrees for Belgian spatial units

Units	1	2	3	4	5	6	7	8	9	10	11
1		2	1	1	2	1	2			1	2
2	2		1	2	2	2	1	1	2	1	2
3	1	1		1	2	1	1	2	2	1	1
4	1	2	1		1	2	1	2		2	2
5	2	2	2	1			1	2			
6	1	2	1	2			2	2	2	1	2
7	2	1	1	1	1	2		1	2	2	2
8		1	2	2	2	2	1		1	1	
9		2	2			2	2	1		1	
10	2	1	1	2		1	2	1	1		2
11	2	2	1	2		2	2			2	

Table 16.9 Estimation results from Tables 16.7 and 16.8 using model (16.6)

Parameters	a	t_a	B	t_b	R^2
SSLS	0.7963	1.1558	0.3279	0.4464	0.5237
QUALI1	0.9167	1.0501	0.5031	0.5763	0.5128
QUALI2	0.4459	0.4328	0.2209	0.2114	0.1581

Table 16.9 presents three results: the first is the SSLS estimators obtained from the original data (the constant has been omitted, because it cannot appear in the following results); and, the other two are QUALIREG estimators. The first of these latter estimates relates to the average ranks of the contiguity products, whereas the second relates to the ranking of the C_1–C_2 numbers in Table 16.7. The sum of absolute values of residuals has been minimized, rendering very robust estimators.

In this case, not only do the signs correspond, but also the ranking of the values obtained. QUALI1 furnishes the closest fit. Both of these outcomes are remarkable.

Of note is that the SSLS R^2 is not particularly high, despite the low number of df (the four lowest out of the eleven endogenous regional products have been fixed, as they had a tendency to turn negative). This finding corresponds to outcomes reported in Chap. 14, in which each spatial unit (they are the same as in this study) had its own reaction coefficients. Factors are at least twofold: spatial bias (see Paelinck, 2000b; Paelinck et al., 2005, pp. 25–26) and spatial asymmetry proper (due to the differences in economic structure of each of the Belgian spatial units; see Chap. 14).

Nevertheless, the present exercise is concerned with studying only the efficiency of the QUALIREG method.

16.4 Conclusion

It now appears possible to perform spatial econometric exercises with very poor (read: qualitative, ordinal) data, rankings rather than cardinal data. This is of utmost importance when working on multiregional problems in developing countries where the latter sort of information often is not available (for examples, see Ancot and Paelinck, 1990)

Further experience is certainly required with the method proposed, but the improvement with the initial Ancot-Paelinck approach (see Ancot and Paelinck, 1986, referred to above) is very clear.