ORDINAL ECONOMETRICS IN REGIONAL AND URBAN MODELLING

Peter Nijkamp
Piet Rietveld

Researchmemorandum 1981-21

september 1981

ABSTRACT

Geographical research is often suffering from inaccurate and unreliable data. This paper deals with the treatment of soft (ordinal) data in spatial statistics and econometrics. After a brief discussion of soft spatial data, a set of recently developed techniques is introduced aiming at drawing quantitative (cardinal) inferences from a soft data input. These techniques are inter alia multidimensional scaling, rank order statistics, logit analysis, interdependence analysis, discriminant analysis and canonical correlation.

A key stone of this paper is formed by combining Theil's approach to logit analysis with Kendall's rank correlation method based on pairwise comparisons. The paper offers several new perspectives for the treatment of soft data in multivariate techniques (such as multiple regression and clustering analysis) and in decision problems.

Contents

1. Introduction

Regional and urban modelling is playing an increasingly important role in geographical research. Despite much criticism, the use of mathematical and statistical tools has become a prerequisite for an operational analysis of regional and urban phenomena. Mathematical models in geography are based on several assumptions such as:

- the complex relationships in a spatial system can adequately be represented by means of a limited set of mathematical equations describing the successive relevant impacts
- all relevant variables (state variables, goal variables, policy instruments etc.), can be measured in an accurate (cardinal) way
- technical, economic, social and institutional side-conditions prevailing in the spatial system at hand are precisely known and can be specified in an operational way
- the policy aspects of models (e.g. decision-maker(s), goal conflicts, policy measures) can also be included in an operational manner
- the time trajectory and the spatial spillover effects of all variables can be precisely computed
- when the state of the spatial system concerned is characterized by uncertainty (for instance, due to stochastic variables) the probability distribution of the stochastic elements is known.

The above mentioned assumptions imply that most regional and urban models are focussing the attention primarily on precise (sharply defined and cardinally measured) variables, so that all standard numerical operations can be applied to them. Qualitative variables, linguistic attributes, fuzzy characteristics and ordinal aspects are normally left out of consideration. This holds true for both urban and regional models and spatial multivariate methods.

Recently, however, much more attention has been paid to 'soft econometrics' and 'soft data analysis', in which non-cardinal variables are dealt with (especially ordinal variables and fuzzy variables). This paper deals in particular with the treatment of soft data in spatial statistics and econometrics with a special emphasis on ordinal multivariate techniques.

The essential aim of multivariate methods in spatial analysis is to reduce the complexity of phenomena in which many variables or attributes are involved. Given this general feature, it is no surprise that these methods have been applied in various fields of research, such as economics, geography, medicine, biology, etc. (cf. Kendall [1975]).

The starting point of multivariate analysis is normally the data matrix X :

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1I} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{J1} & & x_{JI} \end{bmatrix} \tag{1.1}$$

where I denotes the number of members observed and J the number of variables concerned; x_{ji} denotes the value of the j-th variable for the i-th member (for instance, the i-th region).

In the majority of multivariate methods it is assumed that the variables are measured on a cardinal (interval or ratio) scale. This means that it is meaningful to apply numerical operations to these variables such as summation, subtraction and multiplication (see Rietveld [1980] for a more accurate definition of a cardinal scale of measurement). However, in spatial research often the assumption of cardinal measurability cannot be maintained. For example, when the data are not accurate enough, when variables are involved which can only be measured in a qualitative way (e.g., beauty of landscape), or when latent variables are to be dealt with.

It is important, therefore, to consider the question whether it is possible to develop multivariate methods which are not based on the assumption of cardinal measurability. This has of course, important implications for econometric model-building, since the treatment of soft data has always formed a bottleneck for estimating economic models. Soft econometrics is a recently developed approach to overcoming this problem. This paper presents some techniques for ordinal data and its implications for regional statistical and econometric analyses. A variable j is ordinally measurable when for a series of observations it is possible to indicate the rank order of the observations, but not the differences between the observations. For example, when the rank order of five observations - where the smallest value receives rank 1, the one but smallest receives rank 2, etc. - is: 2, 3, 1, 5, 4, it may be concluded that the first observation is smaller than the fourth one, but not that the difference between the first and the fourth one is larger than the difference between the first and the third one.

It is important to note that several techniques can be applied to perform a multivariate analysis of ordinal data in a spatial context. Some examples of such approaches are given below.

1. The simplest way of achieving a short circuit is to interpret ordinal data as if they were cardinal. Obviously, in this way more information is extracted from the data than is actually contained in them. Kendall [1970], p. 125, indicates that sometimes such an approach may yield satisfactory results. However, since it is based on a questionable assumption, it cannot serve as a general device for dealing with ordinal data.
2. Another way of avoiding the necessity to develop ordinal multivariate methods is the use of order statistics to assign cardinal values to the observations (cf. Rietveld [1980]). A necessary step in this approach is the determination of the probability distribution from which the observations are drawn. This can only be done on *a priori* grounds which makes the results arbitrary. The arbitrariness can be removed to a certain extent by repeating the cardinalization of data for different probability distributions. Obviously, the disadvantage of this approach is that it gives rise to extensive computational work.
3. A third cardinalization method consists of applying multidimensional scaling procedures (cf. Nijkamp [1979, 1980]). These procedures have been devised to transform the ordinal matrix X with dimensions J x I into a cardinal matrix Y with dimensions K x I where K < J. Thus, multidimensional scaling procedures are a means to transform ordinal data about many variables into cardinal data about less variables which reflect as accurately as possible the configuration of the original data. Although multidimensional scaling as such is a sound procedure, its use in the present context may give rise to difficulties. For example, the variables derived may be difficult to interpret, which means that the results of the ensuing multivariate analysis may be less meaningful or require at least a closer examination.
4. Another, recently developed approach concerns a set of ordinal regression models developed by McCullagh [1980]. This approach attempts to develop a general class of regression models that are especially appropriate for ordinal observations on variables. These models are based on various models of stochastic orderings of an ordinality structure. The author proposes two models in particular, viz. the proportional odds model (based on a (linear or non-linear) logit method for the ordered categories of response variables, given the values of covariates) and the proportional hazards model (based on a complementary log-log transform of a hazard function for a response variable that depends on the difference between covariates).
5. An interesting development has also taken place in the field of categorical data analysis. Categorical data are very often collected in survey questionaires, when respondents have to indicate whether or not a certain object is regarded as important. The proportions of affirmative or denying responses

regarding the

objects can then be used as data input for the application of a linear logit model, so that the differences in the ratio of these proportions (across the objects) can next be related to a set of explanatory variables representing the attributes of these objects (see, for instance, Wrigley, 1980).

In light of all these attempts to deal with 'soft' data, it is meaningful to study in more detail ordinal multivariate methods. In the field of (regional) economics this subject has mainly been neglected, whereas in other disciplines, expecially sociology, substantial work has been done on this subject. Given the exploratory character of this paper, we will focus our attention on the main ideas and less on specific statistical aspects or on the feasibility of numerical procedures.

We will deal with the following methods:

Section 2: multiple regression analysis (by means of logit analysis, ordinal rank correlation and constrained regression, respectively).

Section 3: related ordinal regression techniques (such as interdependence and discriminant analysis).

Section 4: clustering and classification methods.

Section 5: principal component analysis (and related subjects such as canonical correlation and partial least squares).

2. Multiple Regression Analysis

2.1. Introduction

Consider the following relationships between y - the variable to be explained - and K explanatory variables $x_1, \ldots, x_K$:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K \tag{2.1}$$

For instance, y might be regional income, while the explanatory variables might be regional investments, infrastructure endowment and labour force.

In this section we consider the question: 'given that I observations are available of the variables y and $x_1, \ldots, x_K$ - measured on an ordinal scale -, is it possible to estimate the values of the β_k (k = 0, 1, ..., K) or to draw conclusions about the extent to which the variables x_k contribute to the explanation of y'.

There are several ways to approach this question. In subsections 2.2 - 2.4 we will critically examine a logit formulation, an approach based on multiple rank correlation coefficients and a multiple regression procedure under constraints, respectively.

2.2. Logit Analysis[1)]

In this subsection we will show how a logit analysis based on data about pairwise comparisons of observations, can be used to determine the relative importance of the explanatory variables. Consider all pairs of observations $(i, i' ; i \neq i')$ which will be numbered as $n = 1, \ldots, N$, where $N = I\ (I - 1)$. We introduce new variables w and z_k $(k = 1, \ldots, K)$ which are related to y and x_k $(k = 1, \ldots, K)$ by means of the following dominance relationships:

$$\begin{aligned}
&\text{if for the pair } (i,i') : y_i > y_{i'} \text{, then } w_n = 1\\
&\text{if for the pair } (i,i') : y_i < y_{i'} \text{, then } w_n = 0\\
&\text{if for the pair } (i,i') : x_{ki} > x_{ki'} \text{, then } z_{kn} = 1\\
&\text{if for the pair } (i,i') : x_{ki} < x_{ki'} \text{, then } z_{kn} = 0
\end{aligned} \tag{2.2}$$

These variables can be summarized in a column vector $\underline{w}$ with N elements and an N x K matrix Z:

$$\underline{w} = \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_N \end{bmatrix} \quad , \; Z = \begin{bmatrix} z_{11}, & \cdot & \cdot & \cdot & , z_{1K} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ z_{N1}, & \cdot & \cdot & \cdot & , z_{NK} \end{bmatrix} \tag{2.3}$$

Every row of Z consists of a series of zeros and ones. A certain row combination of zeros and ones will be called a <u>regime</u> l $(l = 1, \ldots, L)$. There are in principle $L = 2^K$ different regimes. Let F_l denote the number of rows in Z with a certain regime l. Let F_{0l} and F_{1l} denote the number of rows of regime l in Z such that the corresponding value of w_n is equal to 0 and 1, respectively. Then we have by definition: $F_l = F_{0l} + F_{1l}$ and $\sum_l F_l = N$.

A numerical example may clarify the meaning of the symbols defined above. Assume that $I = 4$ and $K = 2$ and that the observations of the y and x_k are:

1) The authors thank Professor Franz Palm for his valuable comments on this section.

$$\underline{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 2 \end{bmatrix} \quad , \quad \underline{x}_1 = \begin{bmatrix} 4 \\ 1 \\ 3 \\ 2 \end{bmatrix} \quad , \quad \underline{x}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 4 \end{bmatrix} \tag{2.4}$$

When the pairs of observations are considered in the following order: (1,2), (1,3), (1,4), (2,1), (2,3), ... , (4,3), we arrive at the following results of $\underline{w}$ and Z:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{2.5}$$

The four regimes appearing in (2.5) are (0,0), (1,0), (0,1) and (1,1). The corresponding frequencies are summarized in Table 1:

	regime l			
	$l = 1$	$l = 2$	$l = 3$	$l = 4$
	(0,0)	(1,0)	(0,1)	(1,1)
F_{1l}	3	1	1	1
F_{0l}	1	1	1	3
F_l	4	2	2	4

Table 1. Frequencies for various regimes of explanatory variables.

Let y, x_1 and x_2 denote net migration, the vacancy rate and the quality of infrastructure, respectively. A regime (1,1) means in this case that when two regions A and B are considered, both the vacancy rate and the quality of infrastructure in region A are higher than those in region B. Table 1 shows that there are 4 pairs of regions with this regime: in 3 cases the net migration of region A is smaller than that of region B; in one case the reverse holds true.

The information contained in Table 1 can be used for a standard logit analysis in the following way (cf. Theil [1971a,b] and Upton [1978]). Let p_l denote the probability that w is assigned a value 1 when regime l holds. A regime l is described by a (K x 1) vector $\underline{m}_l$ consisting of 1 and 0 elements m_{lk} (k=1, ..., K). Then the usual assumption in this type of problem is that p_l depends on the structure of regime l in the following way:

$$\ln \left(\frac{p_l}{1-p_l}\right) = \gamma_0 + \gamma_1 m_{l1} + \gamma_2 m_{l2} + \ldots + \gamma_K m_{lK} \qquad (2.6)$$

The right hand side of (2.6) shows an additive structure with dummy variables. If desired, interaction effects between the variables k and k' can be included by adding parameters $\delta_{kk'}$ $(k \neq k')$ when both m_{lk} and $m_{lk'}$ are equal to 1 (cf. Theil [1971a]). The expression $\ln (p_l / (1 - p_l))$ at the left hand side of (2.6) is termed the logit of p_l. Its main feature is that it transforms in a monotone increasing way p_l - falling in the [0,1] interval - to a variable ranging from $-\infty$ to ∞ . For a further discussion of the specification of (2.6) and its relationships with the entropy concept we refer to Theil [1971a].

Equation (2.6) does not contain an error term. The reason is that in the left hand side no observed variable is included. When we want to estimate the parameters γ_k , we have to replace the probabilities p_l by the observed relative frequencies $f_l = F_{l1} / F_l$. In that case there is a clear reason to include an error term, since the relative frequencies f_l may differ from the probabilities p_l. Thus the relationship to be estimated is:

$$\ln \left(\frac{f_l}{1 - f_l}\right) = \gamma_0 + \gamma_1 m_{l1} + \gamma_2 m_{l2} + \ldots + \gamma_K m_{lK} + \varepsilon_l \qquad (2.7)$$

where ε_l is the error term.

Theil [1971a] shows that a weighted least squares method is appropriate to estimate the parameters when it may be assumed that the relative frequencies f_l are based on independent random samples of size F_l from binomial distributions with probability p_l of success. In that case it can be shown that the large sample expectation and variance of ε_l are 0 and $1 / (F_l (f_l) (1-f_l))$, respectively. Consequently, weighted least squares (a special case of generalized least squares) can be applied, the weights being proportional to $\sqrt{F_l (f_l)(1-f_l)}$ (which serves to take account of the sample size).

This means that regimes for which f_l = 1 or 0 do not play a role in the estimation of the γ. We also see that the larger F_l (the number of observations in a regime), the heavier the weight of that regime in the determination of the parameters.

An important difficulty inherent in the estimation of (2.7) in the context of ordinal data analysis is that Theil's assumption that the f_l 's are based on independent random samples of size F_l is not valid. The frequencies F_{1l} refer to pairs of observations which are derived from the original set of observations in a systematic way. For our numerical example is this clearly displayed by Table 1, where we find that $f_1 + f_4 = 1$ and $f_2 + f_3 = 1$.

We conclude that the ε_l in (2.7) cannot be assumed to be distributed independently. Therefore, a generalized least squares estimation of (2.7) is adequate. The obvious difficulty is that the covariance matrix V is not known and that it seems to be impossible to describe V by means of a small number of parameters, as is sometimes done in time series regressions. How can one proceed in this situation? Three directions can be chosen.

1. The simplest way is to ignore the problem and to apply ordinary least squares. In that case the estimated parameters are unbiased, but the variances will be higher compared to the results of generalized least squares (cf. Theil [1971b]).
2. Another way is the use of iterative procedures. For example: start with an estimation by means of ordinary least squares. Use the resulting estimated errors to construct an estimated covariance matrix $\hat{V}$ and apply generalized least squares based on $\hat{V}$, and so forth.
3. A third approach aims at directly approximating the covariance structure of the ε_l 's as follows:

   Consider the set of I original observations. This set can be used to generate I sets of I - 1 observations, each set containing the I original observations but one. For each set the values of f_l and $\ln (f_l / (1-f_l))$ can be determined. The series of I values for the logits can be used to calculate the covariance matrix of the logits. This matrix can be used as an approximation of V so that generalized least squares can be employed.

It is clarifying to pay some attention to the number of observations and parameters in specification (2.7). The number of parameters in (2.7) is equal to K + 1. The maximum number of observations (regimes) is equal to 2^K. This means that when the actual number of observations is equal to the maximum possible number, the number of degrees of freedom increases rapidly with increasing K. However, there are several reasons why the actual number of regimes in the estimation is smaller than 2^K. Especially when I is not so large, for some regimes F_{0l} or F_{1l} (or both) may be equal to zero and - as shown above - such a regime cannot be used to estimate the parameters.

Another reduction in the number of observations is due to the interdependencies by means of the concept of 'complementary regimes'. A regime l' is a complement of a regime l when the sum of $\underline{m}_l$ and $\underline{m}_{l'}$ is a vector $\underline{\iota}$, exclusively consisting of unit elements:

$$\underline{m}_l + \underline{m}_{l'} = \underline{\iota} \tag{2.8}$$

In our numerical example the complement of regime 1 is 4 and of regime 2 is 3. It follows from the definition of F_{01} and F_{11} that for complementary regimes we have the following redundancy conditions:

$$F_{01} = F_{11'} \text{ , } F_{11} = F_{01'} \text{ , and } F_l = F_{l'} \text{ , and therefore: } f_l = 1-f_{l'} \text{ .}$$

Consequently for each pair of complementary regimes holds the following condition:

$$\ln \frac{f_l}{1-f_l} + \ln \frac{f_{l'}}{1-f_{l'}} = 0 \tag{2.9}$$

Combining (2.9) with (2.7) yields for all complementary pairs:

$$\ln \frac{f_l}{1-f_l} + \ln \frac{f_{l'}}{1-f_{l'}} = 2\,\gamma_0 + \sum_k \gamma_k + \varepsilon_l + \varepsilon_{l'} = 0 \tag{2.10}$$

We may conclude, therefore, that for all complementary pairs (l,l') and (s,s') we have:

$$\varepsilon_{s'} = \varepsilon_l + \varepsilon_{l'} - \varepsilon_s \tag{2.11}$$

Consequently, when in (2.7) the relative frequency f is given for the regime l, l' and s, the value of $f_{s'}$ does not add any useful information for the determination of the parameters γ_k . This means that when there are L regimes (L even), only the frequencies of ½L + 1 regimes contain useful information on the parameters (the set of ½L + 1 regimes contains only one pair of complementary regimes). In our numerical example we have K = 2 and hence the number of parameters is equal to 3.

In such a case, in general, the parameters can be determined, while the estimated errors are zero. Indeed, we can derive that $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 0$ and $\gamma_0 = -\gamma_1 = -\gamma_2 = \ln 3$.

At the end of the presentation of this approach, we may conclude that, despite some difficult estimation problems, an ordinal analogon has been developed for multiple regression which does justice to the ordinal character of the data.

What is the essentially new idea of this approach? A close examination shows that the approach consists of two building stones: 1) a method to transform the ordinal data matrix X and the vector $\underline{y}$ in a vector of relative frequencies $f_1, \ldots, f_L$ by means of pairwise comparisons; and 2) an estimation procedure based on specification (2.6). The main elements of the building stones have been developed by Kendall [1970] and Theil [1971a], respectively. The novelty of the method developed here is thus the combination of the two building stones.

This logit analysis enables us to draw inferences about the probability that a certain regime (based on a dominance via pairwise comparisons of individual ordinal explanatory variables) will lead to a dominance of the left hand side variable (also based on a pairwise comparison of a set of observations). Especially in case of soft spatial data, this approach may be a promising one for the derivation of cardinal conclusions from ordinal data. It should be noted that the strength of this approach is based on the fact that both the left hand side and the right hand side variables may be measured in ordinal units.

2.3. Ordinal Analogue of Multiple Regression and Correlation Analysis

Beside the previous linear model, one may also use a correlation analysis. The approach discussed in this subsection is based on structural similarities between product-moment correlation coefficients and rank correlation coefficients. We will first show the nature of these similarities.

The ordinary product-moment correlation coefficient for cardinal data on a set of variables u_i and v_i reads as:

$$r = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})^2 \sum_i (v_i - \bar{v})^2}} \tag{2.12}$$

where $\bar{u}$ and $\bar{v}$ are the mean values of the u_i and v_i respectively.

We also present the regression coefficient b_{vu} following from the estimation of the relationship between v and u:

$$v_i = a + b_{vu} u_i$$

The best linear unbiased estimator of b reads as:

$$b_{vu} = \frac{\sum_i (u_i - \bar{u})(v_i - \bar{v})}{\sum_i (u_i - \bar{u})^2} \tag{2.13}$$

We turn now to some correlation coefficients proposed for ordinal data. First, we will discuss correlations among pairs of variables.

Kendall [1970] proposed to use the method of paired comparisons in the following way. Consider all $\frac{1}{2}I(I-1)$ pairs (i, i') of ordinal observations of two variables x and y. Let S^+ be the number of pairs for which x and y are <u>concordant</u>, i.e., the number of pairs for which $\{x_i > x_{i'}$ and $y_i > y_{i'}\}$ or $\{x_i < x_{i'}$ and $y_i < y_{i'}\}$. Let S^- be the number of pairs for which $\{x_i > x_{i'}$ and $y_i < y_{i'}\}$ or $\{x_i < x_{i'}$ and $y_i > y_{i'}\}$. Let T_x and T_y be the number of ties in x and y, respectively. When no ties appear, Kendall's coefficient of rank correlation is defined as the number of concordant pairs minus the number of discordant pairs divided by the total number of pairs:

$$\tau = \frac{S^+ - S^-}{S^+ + S^-} \tag{2.14}$$

When ties are present, the following correction is applied:

$$\tau_b = \frac{S^+ - S^-}{\sqrt{S^+ + S^- + T_x}\,\sqrt{S^+ + S^- + T_y}} \tag{2.15}$$

For the latter case, Somers (1962) provided an alternative measure which will later also appear to be meaningful:

$$d_{yx} = \frac{S^+ - S^-}{S^+ + S^- + T_y} \tag{2.16}$$

For these three measures it can be proved that the extreme values are -1 and +1, respectively.

At first sight there is not much similarity between these ordinal measures and the above mentioned measures for cardinal data: the ordinal measures are based on counting frequencies of discordant and concordant pairs, while the cardinal measures are based on measuring distances with respect to the mean. It will be shown, however, that the structure underlying the ordinal and the cardinal measures (cf. Hawkes [1971] and Ploch [1974]) is the same.

The first step to prove the similarity is to rewrite (2.12) and 2.13) such that the mean values $\bar{u}$ and $\bar{v}$ disappear. It is not difficult to show that for cordinal data:

$$\sum_i (u_i - \bar{u})^2 = \frac{1}{2I} \sum_i \sum_j (u_i - u_j)^2$$

$$\sum_i (u_i - \bar{u})(v_i - \bar{v}) = \frac{1}{2I} \sum_i \sum_j (u_i - u_j)(v_i - v_j)$$

Accordingly, when we set $u_{ij} = u_i - u_j$ and $v_{ij} = v_i - v_j$, correlation coefficient (2.12) and regression coefficient (2.13) can be rewritten as:

$$r = \frac{\Sigma u_{ij} v_{ij}}{\sqrt{\Sigma u^2_{ij}} \sqrt{\Sigma v^2_{ij}}} \tag{2.12'}$$

$$b_{vu} = \frac{\Sigma u_{ij} v_{ij}}{\Sigma u^2_{ij}} \tag{2.13'}$$

In (2.12') and (2.13') the summation relates to all possible pairs.

In the second step we introduce the following operation for the ordinal data. For all pairs (i, j), we may set:

$$\begin{aligned} x_{ij} &= 1 \quad && \text{if} \quad x_i > x_j \\ x_{ij} &= 0 && \text{if} \quad x_i = x_j \\ x_{ij} &= -1 && \text{if} \quad x_i < x_j \end{aligned}$$

The variable y_{ij} can be defined in the same way. Thus we arrive at two vectors consisting of N^2 elements being equal to 1, 0, or -1. The term $(S^+ - S^-)$ can be expressed in terms of x_{ij} and y_{ij} in a straightforward way, so that $S^+ - S^- = \frac{1}{2} \Sigma x_{ij} y_{ij}$. Given this result it is not difficult to see that:

$$\tau_b = \frac{\Sigma x_{ij} y_{ij}}{\sqrt{\Sigma x^2_{ij} \Sigma y^2_{ij}}} \tag{2.15'}$$

and

$$d_{yx} = \frac{\Sigma x_{ij} y_{ij}}{\Sigma x^2_{ij}} \tag{2.16'}$$

When we compare (2.12') and 2.13') with (2.15') and 2.16') we conclude that, although the correlation coefficients are based on different concepts, they give rise to completely identical analytical expressions.

Thus, our conclusion is that there is a strict correspondence between cardinal and ordinal correlation coefficients (and between cardinal and ordinal regression coefficients as well) in case of a pairwise treatment of data.

Hawkes [1971] and Ploch [1974] argue that these similarities provide a sufficient base for developing partial and multiple correlation coefficients for multivariate ordinal data along the same lines as for cardinal data. Because this would be a very convenient result, this approach deserves a closer examination.

There is at least one argument supporting this. Kendall [1970] has shown that, in developing a partial correlation coefficient for ordinal data based on τ (assuming that no ties occur), one may arrive at a formulation which is completely similar to the formulation of the partial product moment correlation coefficient:

$$\Theta_{mk.l} = \frac{\tau_{mk} - \tau_{ml}\,\tau_{kl}}{\sqrt{1 - \tau^2_{ml}}\sqrt{1 - \tau^2_{kl}}} \tag{2.17}$$

where $\Theta_{mk.l}$ denotes the partial correlation between m and k, given l. This expression is obviously an indication that in some cases ordinal partial correlation coefficients may be dealt with in the same way as their cardinal counterparts. It has to be added, however, that Quade [1974] has indicated several ways to conceptualize a partial correlation coefficient for ordinal data and that not all of them lead to relationships like (2.17).

Further correspondences between ordinal and cardinal measures in the multivariate case have not been found, however. Consequently, the approach of deriving regression coefficients by means of ordinal correlation coefficients is only partly justified. For some empirical applications we refer to Ploch [1974], while Namboodiri et al. [1975] and Blalock [1976] give a more thorough discussion of the above mentioned approach. Thus our conclusion regarding multiple correlation is that ordinal regression and correlation analyses may lead to dissimilar analytical expressions compared to cardinal analyses, except for some specific cases.

2.4. Multiple Regression under Constraints

In this subsection we will deal with ordinal data on y and $x_1, \ldots, x_K$ from (2.1) again in an alternative way, viz. by imposing side-conditions associated with the ordinal values of these variables.

Let $cy, cx_1, \ldots, cx_K$ be the unknown cardinal values corresponding to the ordinal variables. Thus, when $y_3 \geq y_4$, then $cy_3 \geq cy_4$, etc. Accordingly, we arrive at a series of I-1 inequalities (for instance, spatial differences) for the cy_i:

$$cy_{i_1} \geq cy_{i_2} \geq \ldots \geq cy_{i_I} \tag{2.18}$$

where i_1 is the index of the largest observation, i_2 indicates the one but largest observation, etc. The same set of inequalities can be developed for the explanatory variables.

Next one may try to use (2.18) as logical constraints for deriving cardinal units. The information that there is a linear relationship between y and the x_k may be used to determine the cardinal values corresponding to y and the x_k. As a first step in the analysis we consider a constrained regression procedure expressed as the following mathematical programming problem:

$$\left\{\begin{array}{ll} \min\limits_{cy_i, cx_{ki}, \beta_k}! & \sum\limits_i (cy_i - \beta_0 - \beta_1 cx_{1i} - \ldots - \beta_K cx_{Ki})^2 \\ \text{subject to} & cy_{i_1} \geq cy_{i_2} \geq \ldots \geq cy_{i_I} = 1 \\ & cx_{1j_1} \geq cx_{1j_2} \geq \ldots \geq cx_{1j_I} = 1 \\ & \vdots \\ & cx_{Kl_1} \geq cx_{Kl_2} \geq \ldots \geq cx_{Kl_I} = 1 \end{array}\right. \qquad (2.19)$$

Obviously, in (2.19) the cardinal values of y and x_k and the values of the parameters β_k are determined simultaneously. It is a programming problem with (K + 1) (I + 1) variables and (K + 1) I constraints. The variables cy and cx_k have been standardized by imposing that the smallest value is equal to one.

It is not difficult to see that (2.19) as it stands here attains its minimum when all cardinal values are equal to 1 and when $\Sigma\ \beta_k = 1$. This result, when all variables form one big tie, is less meaningful; it is an indication that (2.19) has been designed to serve too many ends on the basis of too little information. When more restrictions can be imposed on the problem, better outcomes may be expected, however.

A first way to improving the result arises when for some of the variables in (2.1) the cardinal values are known in advance. For example, when all explanatory variables are cardinally measured (i.e. the cx_{ki} are known beforehand), (2.19) can be transformed into the following quadratic programming problem:

$$\left\{\begin{array}{ll} \min\limits_{cy_i, \beta_k}! & \sum\limits_i (cy_i - \beta_0 - \beta_1 cx_{1i} - \ldots - \beta_K cx_{Ki})^2 \\ \text{subject to} & cy_{i_1} \geq cy_{i_2} \geq \ldots \geq cy_{i_I} = 1 \end{array}\right. \qquad (2.20)$$

This reduced ordinal data problem may give rise to more meaningful results. Nievergelt [1971] arrives at essentially the same formulation when he tries to estimate the weights β_k of a utility function, where the x_k's are the arguments of the utility function and where a series of I alternatives has been placed in order of attractivity.

A second source of additional information for improving the results of (2.19) can be obtained when the explanatory variables can be distinguished in various classes. For example, Nijkamp [1980] classifies the explanatory variables of regional income in an economic profile and a socio-geographical profile. Multidimensional scaling methods are then used for each profile to derive cardinal values for one or several variables representing the profiles. This cardinal information can then be used for an ordinary multiple regression procedure when y is measured on a cardinal scale. When y is measured on an ordinal scale, however, formulation (2.20) can be used as a meaningful tool.

This approach is obviously a two-step procedure: first the number of ordinal variables is replaced by a smaller number of cardinal variables, while next the derived cardinal values are used to estimate the weights β_k. Evidently, it is worthwhile to consider the possibility of integrating the two steps. This would mean that the derivation of cardinal variables by means of multidimensional scaling uses also information concerning the position of these variables in a larger causal structure. This integration can be carried out in the following way. Let the K variables be divided in T profiles or classes (T < K). Assume that per profile t only one cardinal variable will be determined. This variable will be denoted by z_t, so that the information on each category t for all members i (e.g., regions) is measured in cardinal units. A short-hand description of a multidimensional scaling procedure is the following:

$$\begin{cases} \min\limits_{z_{ti}}! \quad \text{stress}\,(z_{t1}, \ldots, z_{tI}) \\ \text{subject to} \quad (z_{t1}, \ldots, z_{tI}) \in A_t \end{cases} \tag{2.21}$$

Here, stress $(z_{t1}, \ldots, z_{tI})$ is a measure of the discordance between the ordinal data and the z_{ti}'s, while $(z_{t1}, \ldots, z_{tI}) \in A_t$ denotes the transformation relationships from ordinal to cardinal data used in multidimensional scaling. For the ease of notation, the multidimensional scaling variables referring to the variable y will be denoted by an index 0. Then the integration we are aiming at can be reached by solving:

$$\begin{cases} \min! \quad \mu \sum\limits_i (z_{0i} - \gamma_0 - \gamma_1 z_{1i} - \cdots - \gamma_T z_{Ti})^2 \\ \qquad\quad + \sum\limits_{t=0}^{T} \lambda_t \,\text{stress}\,(z_{t1}, \ldots, z_{tI}) \\ \text{subject to} \quad (z_{t1}, \ldots, z_{tI}) \in A_t \qquad t = 0, 1, \ldots, T \end{cases} \tag{2.22}$$

The outcome of (2.22) depends on the weights μ and $\lambda_0, \lambda_1, \ldots, \lambda_T$ attached to the various terms of the objective function. It is not difficult to see that the two-step procedure mentioned above is a special case of (2.22), viz. by first solving (2.22) with $\mu = 0$, and subsequently with the values for the z_{ti}'s obtained in the first step (with $\lambda_0 = \lambda_1 = \ldots = \lambda_T = 0$). It may be concluded that this method has more arbitrary elements than the previous ones.

3. Related Ordinal Multiple Regression Techniques

The ordinal multiple regression techniques discussed in section 2 can also be applied to related problems where regression analysis plays an implicit role, for instance, in subselection and classification problems. Two methods that are fairly common in geographical research, viz. interdependence analysis and discriminant analysis, will briefly be discussed here.

3.1. Interdependence Analysis

Interdependence analysis is a method aiming at selecting a set of variables from a larger data set such that the selected variables represent the original data set as good as possible (see Kendall [1975] and Blommestein et al [1980]). This method is based on multiple regression, since the criterion in the selection procedure is the multiple correlation coefficient between each of the discarded variables (the variables to be explained) and the selected variables (the explanatory variables). We conclude, therefore, that the approaches to ordinal multiple regression dealt with in subsections 2.2 - 2.4 are equally useful for ordinal interdependence analysis.

3.2. Discriminant Analysis

The aim of discriminant analysis is the determination of a decision rule which assigns individuals (in a broad sense) to certain predetermined classes on the basis of their characteristics, such that the probability of misclassification is as small as possible. Let the characteristics of the individuals be denoted by $x_1, \ldots, x_K$ and assume that there are only two classes. Then a frequently used form of the decision rule is (assuming that the variables are measured on a cardinal scale):

$$\text{if} \quad \sum_k \beta_k x_{ki} > c \text{ , assign x to class A}$$
$$(3\text{-}1)$$
$$\text{if} \quad \sum_k \beta_k x_{ki} \leq c \text{ , assign x to class B}$$

From this formulation of a decision rule, it is clear that there is a high similarity between regression analysis and discriminant analysis. For example, Kendall [1975, p.94] considers linear regression with a nominally measured regressand y as identical to discriminant analysis.

When the scale of measurement of the x_k's is ordinal, a decision rule can be conceptualized in the following way. It can no longer refer to one individual and therefore we propose referring it to a pair of individuals (i,i'). It indicates to which class (A or B) individual i has to be assigned, given the characteristics for which individual i is larger than i', given the characteristics for which i is smaller than i', and given the class to which alternative i' belongs.

This formulation of a decision rule enables one to employ the logit function (2.7) for ordinal discriminant analysis. The only necessary adjustment concerns the right hand side of (2.7), where a dummy variable has to be added indicating whether or not alternative i' belongs to class A. It is not difficult to determine the reference value c as introduced in (3.1). This value is in ordinal discriminant analysis equal to zero, since $\ln p/(1-p) = 0$ implies: $p = .50$

It is finally interesting to note, that also in discriminant analyses with discrete explanatory variables similar specifications of the decision rule are used (cf. Goldstein and Dillon [1978]).

Our final conclusion is that related ordinal multivariate problems can be dealt with in a meaningful way by means of the tools described in section 2.

4. Cluster Analysis

Consider the above mentioned data matrix (1.1.) and suppose it is measured in ordinal units. Then one may raise the question whether it is possible to develop a cluster technique for these ordinal data, especially because cluster techniques have gained much popularity in geographical research.

Clustering aims at deriving sets of "similar" individuals or variables. Some authors use the term clustering only in connection with individuals and employ the term classification in connection with variables. We will normally use the term clustering; only when misunderstandings might arise we will indicate whether we mean clustering of individuals or of variables. It is interesting to note that clustering implies the transformation of numerical data to data measured on a nominal scale.

There are many types of clustering methods (see Hartigan [1975]). Clustering methods can be distinguished among others according to:

- the similarity criterion
- the objective function (e.g., the objective may be: maximize the similarity within clusters, minimize the similarity between clusters, or employ some mixed objective)
- the way in which clusters are combined (hierarchical versus non-hierarchical).

In this paper we will only deal with the first mentioned feature: the similarity criterion.

When the aim is a clustering of ordinally measured variables, it is not difficult to find a similarity criterion. Kendall's rank correlation coefficient defined in (2.5) is a good indicator for the interdependence between two variables, that is closely related to the notion of similarity between two variables. When a cluster C consists of more than two variables, an adequate similarity index (based on the rank correlation coefficient) is:

$$s(C) = \min_{j,j' \in C} \tau_{j,j'} \qquad (4.1)$$

Thus, s(C) indicates the minimum correlation between all pairs of variables in cluster C.

Serious difficulties arise when the aim is a clustering of individuals on the basis of ordinal data. This is clearly exemplified by the following data matrix, describing the outcomes of two variables for four individuals:

$$X = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{bmatrix} \qquad (4.2)$$

It is tempting to state that in (4.2) the first and second individual are better candidates to form a cluster than the 2-nd and 3-rd individual, since

$$\sum_j (x_{j1} - x_{j2})^2 \leq \sum_j (x_{j2} - x_{j3})^2 \qquad (4.3)$$

This numerical operation with ordinal data, however, may lead to false conclusions. For instance, suppose that the underlying cardinal values were:

$$X = \begin{bmatrix} 10 & 60 & 65 & 100 \\ 95 & 100 & 90 & 10 \end{bmatrix} \qquad (4.4)$$

a cluster between the 2-nd and 3-rd individual should be preferred. Obviously the root of this problem is a mis-interpretation of the ordinal data matrix X.

In the present section, we will show that it is yet possible to draw certain conclusions about clusterings based on ordinal data, although in most cases the conclusions will not be strong, as the distance metric (4.3) has only a limited relevance. A discussion of distance properties of multivariate techniques (inter alia in the case of qualitative variates) can be found in Gower [1966].

Consider a pair of individuals (i,m). Let $s(i,m)$ denote the similarity between i and m. Then the following statement is in accordance with an ordinal matrix $X = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_I)$:

$$\text{if } \underline{x}_i \leq \underline{x}_m \leq \underline{x}_n \text{ , then : } \begin{array}{l} s(i,m) \geq s(i,n) \quad \text{and} \\ s(m,n) \geq s(i,n) \end{array} \tag{4.5}$$

Thus we arrive at an ordinal similarity measure.
It is not difficult to prove that this measure has the following properties:

reflexivity : $s(i,m) \geq s(i,m) \quad \forall\ (i,m)$
transitivity : if $s(i,m) \geq s(l,n)$ and if $s(l,n) \geq s(k,r)$,
then $s(i,m) \geq s(k,r) \quad \forall\ (i,m), (l,n), (k,r)$.

It cannot be proved that this measure is complete, however. Completeness would mean that for all combinations of pairs (i,m), (l,n) either $s(i,m) \geq s(l,m)$ or $s(i,m) \leq s(l,n)$; in other words, it would imply that it is possible to indicate for all combinations of pairs which of either pair is most similar.

We will illustrate the similarity measures $s(i,m)$ by means of the matrix X in (4.2). An incomparable combination of pairs will be denoted by u. In Table 2 we represent the results of a combination-wise comparison of the similarity index for all pairs of individuals.

(i,m) \ (l,n)	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
(1,2)	=	u	u	u	u	u
(1,3)	u	=	$\geq$	u	u	u
(1,4)	u	$\leq$	=	u	u	$\leq$
(2,3)	u	u	u	=	$\geq$	u
(2,4)	u	u	u	$\leq$	=	$\leq$
(3,4)	u	u	$\geq$	u	$\geq$	=

Table 2. Results of a combination-wise comparison of the similarity index $s(.,.)$ for the pair of alternatives (i,m) and the pair (l,n)

The table clearly shows that most of the combinations are incomparable. We illustrate its meaning for the clustering problem by means of the second row. This row implies that a necessary condition for a common membership by individuals 1 and 4 of the same cluster is, that also individual 3 is a member of that cluster. Thus, the information contained in this row implies that a clustering such as $C_1 = \{2,3\}$ and $C_2 = \{1,4\}$ is not consistent with the ordinal data matrix X.

It appears, however, that in general several clusterings exist which are in accordance with the information of the type of Table 2. For example, when we consider the ways in which two clusters can be formed that are consistent with Table 2, we arrive at:

$$\begin{array}{lll} 1. & C_1 = \{1,2\} & C_2 = \{3,4\} \\ 2. & C_1 = \{1\} & C_2 = \{2,3,4\} \\ 3. & C_1 = \{2\} & C_2 = \{1,3,4\} \\ 4. & C_1 = \{4\} & C_2 = \{1,2,3\} \end{array} \qquad (4.6)$$

We conclude that we need an additional criterion to reduce the number of feasible clusterings. One positive way is to use the median $\underline{x}^m$ as a <u>reference point</u>. For example, when J = 2, we arrive at 4 possible clusters:

$$
\begin{aligned}
C_1 &= \{\underline{x} \mid \underline{x} \geq \underline{x}^m\} \\
C_2 &= \{\underline{x} \mid \underline{x} \leq \underline{x}^m\} \\
C_3 &= \{\underline{x} \mid x_1 \geq x_1^m ,\ x_2 < x_2^m\} \\
C_4 &= \{\underline{x} \mid x_1 < x_1^m ,\ x_2 \geq x_2^m\}
\end{aligned}
\qquad (4.7)
$$

When (4.7) is applied to (3.2), we find:

$C_1 = C_2 = \phi$, $C_3 = \{1,2\}$, $C_4 = \{3,4\}$,

which is one of the feasible clusters in (4.6).

It is not difficult to show that a clustering along these lines is always in accordance with the information contained in Table 2 (and hence with (4.5), irrespective of the number of variables J, the number of individuals I, or the number of reference points used (e.g., in addition to the median one may also use the quartile positions). The proof reads as follows:

A general way to describe an arbitrary cluster in this situation is:

$$
C^{rs} = \{\underline{x} \mid \underline{x}^r \leq \underline{x} \leq \underline{x}^s\} \qquad (4.8)
$$

where $\underline{x}^r$ and $\underline{x}^s$ are vectors with reference values.

Condition (4.5) states that, when $\underline{x}_i \leq \underline{x}_m \leq \underline{x}_n$ and when individuals i and n are in the same cluster, also individual m should be in that cluster. This condition is satisfied by (4.8), since when $\underline{x}_i \in C^{rs}$ and $\underline{x}_n \in C^{rs}$, (4.8) implies that also $\underline{x}_m \in C^{rs}$.

We may conclude that given an ordinal data matrix X, clustering of variables is not essentially different from a situation with cardinal data. The clustering of individuals is more difficult with an ordinal X, however. We proved that a consistent clustering can be achieved by using reference points (such as the median). Of course, the cluster results depend on the reference points used.

5. Principal Component Analysis

5.1. Introduction

The aim of principal components analysis is the representation of J variables by a smaller number of variables (called components) with a high degree of accuracy. When the data matrix X is cardinal, this can be achieved as follows. We describe X as a series of J row vectors : $\underline{x}_1', \ldots, \underline{x}_J'$.

Then the first component $\underline{p}'$ has to be determined such that the difference between each $\underline{x}_j'$ and $a_j\ \underline{p}'$ is as small as possible. The factor a_j is a scaling factor to allow for the fact that the J variables can be measured in different dimensions. Thus the first component $\underline{p}$ can be found by solving:

$$\min_{a_j,\, p_i} ! \quad \sum_i \sum_j (x_{ij} - a_j p_i)^2 \tag{5.1}$$

This means that the matrix X consisting of I J parameters is approximated by the matrix $\underline{a}\ \underline{p}'$, based on I + J parameters.

The second component can be found by repeating this procedure for the data matrix consisting of the errors remaining after the first step. In general, component n is based on the errors remaining after step n - 1.

Principal component analysis is also a widely used tool in quantitative geographical research, so that it is extremely interesting to explore the possibilities of ordinal principal component techniques.

5.2. Ordinal Approaches to Principal Component Analysis

Is it possible to extract components when X is ordinal? We will discuss several proposals all dealing with the extraction of one component.

1. Kendall [1970] proposes to base the components on the I sums of the elements in the columns of X. Thus, first one calculates:

$$s_i = \sum_j x_{ji} \qquad i = 1, \ldots, I \tag{5.2}$$

Subsequently the I individuals are ranked according to the s_i .
For example, when $s_{i'}$ is set equal to I, etc., Kendall proves that this procedure yields the maximum average correlation (of the Spearman type) between the rankings in X and the component. Thus by means of this procedure we maximize:

$$\frac{1}{J} \sum_j \rho_j \tag{5.3}$$

where ρ_j is the Spearman type correlation between the component and variable j. This component is very easy to compute, and Kendall [1975] shows an application of it for an analysis of crop productivity in various countries. He reports that there was a striking agreement between the first principal component based on cardinal data and the component based on (5.2) for ordinal values being in accordance with the cardinal ones.

Yet, there is a weak point in this approach. It can be illustrated by means of the following data matrix:

$$X = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 4 & 3 & 1 & 2 \end{bmatrix} \qquad (5.4.)$$

In this case the column sums are all equal, which means that the component consists of equal outcomes for all individuals. This is a strange result when we realize that in (5.1) the scaling factor a_j may be positive as well as negative. Following the lines of (5.1) we should conclude that in (5.4), (1 2 4 3) or(4 3 1 2) would be perfect components, since they do not give rise to any remaining errors.

In more general terms this objection against (5.2) can be formulated as follows:
criterion (5.3) is not meaningful, as it ignores the possibility of negative correlations. Better criteria would therefore be:

$$\max ! \quad \frac{1}{J} \sum_j |\rho_j| \qquad (5.5.a)$$

or:

$$\max ! \quad \frac{1}{J} \sum_j \rho_j^2 \qquad (5.5.b)$$

It is not difficult to see that these criteria - when applied to (5.4) - yield the desired outcomes. It is important to note, however, that there is no straightforward way to determine the solution of (5.5.a) or (5.5.b), as was the case with (5.2).

2. Another approach discussed by Kendall [1970] is based on a special hierarchical frequency of ordinal outcomes. This will be illustrated by the following X matrix:

$$X = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \\ 1 & 3 & 2 & 4 \end{bmatrix} \qquad (5.6)$$

This method uses the number of variables with outcome I, I - 1, sequentially etc., obtained by each individual. For example, individual 4 will receive rank 4 since it includes two outcomes 4 in its column. Individual 1 will receive rank 3 since it has the other outcome 4. Further, the second individual gets rank 2 since it has two values equal to 3 and rank 1 is for individual 3.

Kendall dismisses this approach, however, since it is not self-consistent. This can be seen when the same procedure is followed, but now starting with the value 1. It is easy to see that a rank order is achieved which is different from the order when we start with value 4. This is obviously an unattractive property.

3. Another approach, suggested by Ehrenberg [1952], is to base a component on the number of variables in regard to which individual i is ranked higher than i'. It is interesting to note the similarity between this idea and the principle of majority voting between pairs of alternatives. Indeed, the problem of deriving a common component from a series of rankings is very similar to the problem of finding a social welfare function based on a series of preference relationships. Arrow [1951] has shown that such an aggregation of preferences is only possible under rather restrictive assumptions.

   A well-known illustration of the difficulties in this respect is based on the following ranking of three alternatives i by three persons j:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 3 & 1 \end{bmatrix} \qquad (5.7)$$

When majority voting is used to select an alternative from the pair (1,2), alternative 2 will be chosen. Voting between alternatives 2 and 3 leads to the selection of alternative 3. Voting between alternatives 1 and 3 leads to the selection of alternative 1. The aggregated preference relation obtained in this way is intransitive (cf. Section 4) which is obviously unsatisfactory.

We conclude that this third approach will give rise to the same problems as in social welfare theory. Up to now there has not been much progress in solving these problems. Therefore, this approach is not very promising, unless less restrictive assumptions are being made.

4. The last approach is related to the proposals in (4.5), but instead of Spearman's correlation coefficient it uses the Kendall approach. Thus the component has to be determined such that a maximum is attained for one of the two following criteria:

$$\frac{1}{J} \sum_j |\tau_j| \tag{5.8.a}$$

or

$$\frac{1}{J} \sum_j \tau_j^2 \tag{5.8.b}$$

For the moment, our conclusion is that especially the (adjusted) first method and the fourth method may provide a meaningful approach to ordinal principal component analysis.

5.3. Related Techniques

At the end of this section, we will also pay some attention to canonical correlation and partial least squares, since these methods are closely related to principal component analysis (cf. Kendall [1975] and Wold [1979]).

The aim of canonical correlation analysis is the determination of components from two data sets X_1 and X_2 such that the correlations between the components are as high as possible. Partial least squares can be conceived of as a generalization of canonical correlation analysis since it deals with the analysis of correlations between components derived from more than two data sets.

We will illustrate for canonical correlation analysis how it can be carried out when X_1 and X_2 are ordinal. Let the number of variables in X_1 and X_2 be J_1 and J_2 respectively. Then the components $\underline{p}_1$ and $\underline{p}_2$ are the solution of:

$$\max ! \quad \frac{1}{J_1} \sum_{j_1} |\tau_{j_1}| + \frac{1}{J_2} \sum_{j_2} |\tau_{j_2}| + |\tau_{1,2}| \tag{5.7.a}$$

or

$$\max ! \quad \frac{1}{J} \sum_{j_1} \tau_{j_1}^2 + \frac{1}{J_2} \sum_{j_2} \tau_{j_2}^2 + \tau_{1,2}^2 \tag{5.7.b}$$

In these formulations, $\tau_{1,2}$ denotes Kendall's correlation coefficient between components 1 and 2.
Further, τ_{j_1} and τ_{j_2} denote Kendall's correlation coefficient between component 1 and variable j_1 for 1 = 1,2 , respectively. It should be noted, that in (4.9) an equal weight is given to the correlations internal to an X_1 and the external correlations between the X_1's.

In conclusion, ordinal principal component techniques may find fruitful applications in geographical research.

6. Conclusion

We conclude that it is in principle possible to develop multivariate methods for ordinal data that are related to corresponding methods for cardinal data without making mis-interpretations concerning the character of ordinal data. The methods developed in this paper show the power of soft econometric methods in regional and urban modelling. Clearly, this field has to be explored much further. For example, in further elaborations more attention has to be paid to:

- statistical properties of the ordinal data methods (see also McCullagh, 1980)
- statistical tests related to the methods
- computational aspects
- the occurrence of ties
- the possibility that part of the variables are ordinal and others are cardinal.

An important question concerning the newly developed methods is whether they give rise to outcomes that differ much from the outcomes of methods based on cardinal data. In order to test the sensitivity of the results, it is meaningful to perform both a cardinal analysis and an ordinal analysis to the same data set, in which the data in the ordinal analysis are obtained from an ordinal transformation of the original cardinal data base (see Figure 1.) .

a. cardinal data → cardinal multivariate methods → statements about the structure of the data

b. cardinal data → ordinal data → ordinal multivariate methods → statements about the structure of the data

Fig. 1. Input-output schemes for multivariate methods.

In these cases one may analyze the ordinal data matrix of which the corresponding cardinal values are known. Then we will be able to compare the specificity of the outcomes of ordinal and cardinal methods.

This comparison is also important for several fields of research or decision-making where ordinal data are already used as a source of information (e.g. certain multiobjective decision methods). We expect that ordinal data techniques may have important side-effects on various numerical methods in spatial analysis.

References

Arrow, K.J., Social Choice and Individual Values, Wiley, New York, 1951.

Blalock, H.M., Can We Find a Genuine Ordinal Slope Analogue? in D.R. Heise (ed.), Sociological Methodology, Jossey-Bass Publishers, San Fransisco, 1976, pp. 195-229.

Blommestein, H.J., P. Nijkamp and P. Rietveld, A Multivariate Analysis of Spatial Inequalities, in W. Buhr and P. Friedrich (eds.), Regional Development under Stagnation, Nomos, Baden-Baden, 1980.

Ehrenberg, A.S.C., On Sampling from a Population of Rankers, Biometrika, Vol. 39, 1952, pp. 82-87.

Goldstein, M. and W.R. Dillon, Discrete Discriminant Analysis, Wiley, New York, 1978.

Gower, J.C., Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis, Biometrika, Vol. 53, No. 3, 1966, pp. 325-338.

Hartigan, J.A., Clustering Algorithms, Wiley, New York, 1975.

Hawkes, R.K., The Multivariate Analysis of Ordinal Measures, American Journal of Sociology, Vol. 76, pp. 908-926.

Kendall, M.G., Rank Correlation Methods, Griffin, London 1970.

Kendall, M.G., Multivariate Analysis, Griffin. London, 1975.

McCullagh, P., Regression Models for Ordinal Data, Journal of the Royal Statistical Society, B, Vol. 42, 1980, pp. 109-142.

Namboodiri, N.K., L.F. Carter, and H.M. Blalock, Applied Multivariate Analysis and Experimental Designs, McGraw-Hill, New York, 1975.

Nievergelt, E., Ein Beitrag zur Lösung von Entscheidungsproblemen mit Mehrfacher Zielsetzung, Die Unternehmung, Vol. 25, 1971, pp. 101-126.

Nijkamp, P., Multidimensional Spatial Data and Decision Analysis. John Wiley, New York/London, 1979.

Nijkamp, P., Soft Econometric Models, Research Memorandum 1980-5, Department of Economics, Free University, Amsterdam. 1980.

Ploch, D.R., Ordinal Measures of Association and the General Linear Model, in H.M. Blalock (ed.), Measurement in the Social Sciences, MacMillan, London, 1974, pp. 343-368.

Quade, D., New Parametric Partial Correlation, in H.M. Blalock (ed.), Measurement in the Social Sciences, MacMillan, London, 1974, pp. 369-398.

Rietveld, P., Multiple Objective Decision Methods and Regional Planning, North Holland Publishing Company, Amsterdam, 1980.

Somers, R.H., A New Asymmetric Measure of Association for Ordinal Variables, American Sociological Review, Vol. 27, pp. 799-811.

Theil, H., On the estimation of Relationships Involving Qualitative Variables, American Journal of Sociology, Vol. 76, 1971a, pp. 103-154.

Theil, H., Principles of Econometrics, Wiley, New York, 1971b.

Upton, G.J.G., The Analysis of Cross-Tabulated Data, Wiley, New York, 1978.

Wold, H., Model Construction and Evaluation when Theoretical Knowledge is Scarce; An Example of the Use of Partial Least Squares, Cahier 79-06 du Departement d'Econometrie, Faculté des Sciences Economiques et Sociales, Université de Genève, 1979.

Wrigley, N., Categorical Data, Repeated-Measurement Research Designs, and Regional Industrial Surveys, Regional Studies, vol.14, 1980, pp. 455-471.

Serie Research Memoranda:

1979-1	A.A. Schreuder-Sunderman, Frans Blommaert en Hein Schreuder,	Werknemers en Soc. Jaarverslag.
1979-2	H. Schreuder,	De maatschappelijke verantwoordelijkheid van Ondernemingen.
1979-3	P. Nijkamp en P. Rietveld,	Multilevel Multi-objective Models in a Multiregional System.
1979-4	J. Arntzen, G. Bornmalm-Jardelöw and P. Nijkamp,	Duality, Segmentation and Dynamics on a Regional Labour Market.
1979-5	P. Nijkamp, A. Soffer	Soft Multicriteria Decision Models for Urban Renewal Plans.
1979-6	drs. A.J. Mathot	A Model of choosing a car with or without a credit.
1979-7	H. Blommestein, P. Nijkamp en W. van Veenendaal,	Shopping Perceptions and preferences: A multidimensional Attractivenerss Analysis of Consumer and Entrepreneurial Attitudes.
1979-8	H.J.Blommestein and F.C. Palm	The Aggregate Demand for Money in the Netherlands- a new look at a study of the Bank of the Netherlands.
1980-1 Jan.	P. Nijkamp and H. Voogd	New Multicriteria Methods for Physical Planning by Means of Multidimensional Scaling Techniques.
1980-2	Hidde P. Smit	Medium- and Long-Term Models for the Escap-Region - A review of existing models and a proposal for a new model system -
1980-3 april	P.v. Dijck en H. Verbruggen	Productive Employment in Developing Countries' Exporting Industries
1980-4	P. Nijkamp en L. Hordijk	Integrated Approaches to Regional Development Models; A survey of some Western European Models
1980-5	P. Nijkamp	Soft Econometric Models. An analysis of Regional Income Determinants
1980-6	P. Nijkamp en F. van Dijk	Analysis of Conflicts in Dynamical Environmental Systems via Catastrophe Theory.
1980-7 juni	E. Vogelvang	A short term econometric model for the consumer demand of roasted coffee in the Netherlands
1980-8	N. van Hulst	De effectiviteit van de geleide loonpolitiek in Nederland.
1980-9 Oct. 1980	P. Nijkamp	A survey of Dutch integrated Energy-Environmental-Economic Policy Models
1980-10 Oct. 1980	P. Nijkamp	Perspectives for Urban analyses and policies
1980-11 Oct. 1980	P. Nijkamp	New developments in multidimensional geographical data and policy analysis
1980-12	F.C. Palm, E. Vogelvang en D.A. Kodde	Efficient Estimation of the Geometric Distributed Lag Model; Some Monte Carlo Results on Small Sample Properties