British Journal of Mathematical and Statistical Psychology (1983), **36**, 54–80 *Printed in Great Britain*

Non-linear canonical correlation†

Eeke van der Burg and **Jan de Leeuw**

Non-linear canonical correlation analysis is a method for canonical correlation analysis with op scaling features. The method fits many kinds of discrete data. The different parameters are solv an alternating least squares way and the corresponding program is called CANALS. An applica CANALS is discussed and also a study of the stability of the scaling results.

1. Introduction: linear canonical correlation analysis

In classical multivariate analysis canonical correlation is a well-known methoc relating two sets of variables to each other. It was Hotelling who in 1936 intr this type of analysis. Since then a number of scientists have written about car correlation. Amongst others Thomson (1940, 1947), Bartlett (1948), Anderson Hooper (1959), Horst (1961), Meridith (1964), Norman (1965), Mukherjee (196 Stewart & Love (1968), Weiss (1972), Thorndike & Weiss (1973) and Barcikow Stevens (1975). A good overview can be found in Buchanan (1979). In practic questions about the relations between two sets of variables are very common. is a rather frequent use of canonical correlation analysis (CCA) in econometric instance Waugh (1942), Tinter (1946), Finney (1956) and Adelman *et al.* (1969 psychological research CCA is not very common although there are some appl (e.g. Wood & Erskine, 1976). Weiss (1972) gives some references for applicatio research on counselling psychology. Later on in this article we treat an examp political science. We examine the relation between party preferences and opini several issues among members of the Dutch Parliament.

When applying canonical correlation analysis, we consider two sets of varia Let us denote the first set by a $n \times m_1$ matrix $\mathbf{H}_1$ and the second set by a $n \times$ matrix $\mathbf{H}_2$. The columns of $\mathbf{H}_1$ and $\mathbf{H}_2$ represent *variables* and the rows *objects individuals*. A variable or *data vector* consists of *n observations*. The variables o first set and second set both span a *linear subspace*, $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively. Ca correlation analysis looks for *common directions* in the two subspaces. If $\mathbf{L}_1$ an have a subspace in common, we have a case of perfect fit. If $\mathbf{L}_1$ and $\mathbf{L}_2$ have n in common, CCA looks for directions in $\mathbf{L}_1$ and $\mathbf{L}_2$ which are as similar as poss this case the fit is imperfect, and we need a measure for the goodness-of-fit. N the correlation between the corresponding directions is used for this purpose. correlation is called the *canonical correlation* and the directions are called the *canonical variates*. The canonical variates are one-dimensional subspaces of $\mathbf{L}_1$ and will be denoted by $\mathbf{H}_1\mathbf{a}_1$ and $\mathbf{H}_2\mathbf{a}_2$. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are called weig vectors or *canonical weights*. $\mathbf{a}_1$ consists of m_1 elements and $\mathbf{a}_2$ of m_2 elements. finding the best linear combinations in $\mathbf{L}_1$ and $\mathbf{L}_2$, we may look for the second linear combinations, and so on p times. We therefore talk about weight matric

† The present article is an expanded and revised version of a paper presented at the European of the Psychometric Society in Groningen (The Netherlands) in June 1980.

British Journal of Mathematical and Statistical Psychology (1983), **36**, 54–80 *Printed in Great Britain*

Non-linear canonical correlation†

Eeke van der Burg and **Jan de Leeuw**

Non-linear canonical correlation analysis is a method for canonical correlation analysis with optim scaling features. The method fits many kinds of discrete data. The different parameters are solved an alternating least squares way and the corresponding program is called CANALS. An applicatio CANALS is discussed and also a study of the stability of the scaling results.

1. Introduction: linear canonical correlation analysis

In classical multivariate analysis canonical correlation is a well-known method o relating two sets of variables to each other. It was Hotelling who in 1936 introdu this type of analysis. Since then a number of scientists have written about canon correlation. Amongst others Thomson (1940, 1947), Bartlett (1948), Anderson (1 Hooper (1959), Horst (1961), Meridith (1964), Norman (1965), Mukherjee (1966), Stewart & Love (1968), Weiss (1972), Thorndike & Weiss (1973) and Barcikowsk Stevens (1975). A good overview can be found in Buchanan (1979). In practice questions about the relations between two sets of variables are very common. Th is a rather frequent use of canonical correlation analysis (CCA) in econometrics, f instance Waugh (1942), Tinter (1946), Finney (1956) and Adelman *et al.* (1969). I psychological research CCA is not very common although there are some applica (e.g. Wood & Erskine, 1976). Weiss (1972) gives some references for applications research on counselling psychology. Later on in this article we treat an example political science. We examine the relation between party preferences and opinion several issues among members of the Dutch Parliament.

When applying canonical correlation analysis, we consider two sets of variable Let us denote the first set by a $n \times m_1$ matrix $\mathbf{H}_1$ and the second set by a $n \times m_2$ matrix $\mathbf{H}_2$. The columns of $\mathbf{H}_1$ and $\mathbf{H}_2$ represent *variables* and the rows *objects* or *individuals*. A variable or *data vector* consists of n *observations*. The variables of t first set and second set both span a *linear subspace*, $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively. Cano correlation analysis looks for *common directions* in the two subspaces. If $\mathbf{L}_1$ and $\mathbf{L}$ have a subspace in common, we have a case of perfect fit. If $\mathbf{L}_1$ and $\mathbf{L}_2$ have notl in common, CCA looks for directions in $\mathbf{L}_1$ and $\mathbf{L}_2$ which are as similar as possibl this case the fit is imperfect, and we need a measure for the goodness-of-fit. Norn the correlation between the corresponding directions is used for this purpose. Thi correlation is called the *canonical correlation* and the directions are called the *canonical variates*. The canonical variates are one-dimensional subspaces of $\mathbf{L}_1$ an and will be denoted by $\mathbf{H}_1\mathbf{a}_1$ and $\mathbf{H}_2\mathbf{a}_2$. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are called weight vectors or *canonical weights*. $\mathbf{a}_1$ consists of m_1 elements and $\mathbf{a}_2$ of m_2 elements. Af finding the best linear combinations in $\mathbf{L}_1$ and $\mathbf{L}_2$, we may look for the second be linear combinations, and so on p times. We therefore talk about weight matrices

† The present article is an expanded and revised version of a paper presented at the European Mee of the Psychometric Society in Groningen (The Netherlands) in June 1980.

British Journal of Mathematical and Statistical Psychology (1983), **36**, 54–80 *Printed in Great Britain*

Non-linear canonical correlation†

Eeke van der Burg and **Jan de Leeuw**

Non-linear canonical correlation analysis is a method for canonical correlation analysis with optimal scaling features. The method fits many kinds of discrete data. The different parameters are solved for in an alternating least squares way and the corresponding program is called CANALS. An application of CANALS is discussed and also a study of the stability of the scaling results.

1. Introduction: linear canonical correlation analysis

In classical multivariate analysis canonical correlation is a well-known method of relating two sets of variables to each other. It was Hotelling who in 1936 introduced this type of analysis. Since then a number of scientists have written about canonical correlation. Amongst others Thomson (1940, 1947), Bartlett (1948), Anderson (1958), Hooper (1959), Horst (1961), Meridith (1964), Norman (1965), Mukherjee (1966), Stewart & Love (1968), Weiss (1972), Thorndike & Weiss (1973) and Barcikowski & Stevens (1975). A good overview can be found in Buchanan (1979). In practice questions about the relations between two sets of variables are very common. There is a rather frequent use of canonical correlation analysis (CCA) in econometrics, for instance Waugh (1942), Tinter (1946), Finney (1956) and Adelman *et al.* (1969). In psychological research CCA is not very common although there are some applications (e.g. Wood & Erskine, 1976). Weiss (1972) gives some references for applications in research on counselling psychology. Later on in this article we treat an example from political science. We examine the relation between party preferences and opinions on several issues among members of the Dutch Parliament.

When applying canonical correlation analysis, we consider two sets of variables. Let us denote the first set by a $n \times m_1$ matrix $\mathbf{H}_1$ and the second set by a $n \times m_2$ matrix $\mathbf{H}_2$. The columns of $\mathbf{H}_1$ and $\mathbf{H}_2$ represent *variables* and the rows *objects* or *individuals*. A variable or *data vector* consists of n *observations*. The variables of the first set and second set both span a *linear subspace*, $\mathbf{L}_1$ and $\mathbf{L}_2$ respectively. Canonical correlation analysis looks for *common directions* in the two subspaces. If $\mathbf{L}_1$ and $\mathbf{L}_2$ have a subspace in common, we have a case of perfect fit. If $\mathbf{L}_1$ and $\mathbf{L}_2$ have nothing in common, CCA looks for directions in $\mathbf{L}_1$ and $\mathbf{L}_2$ which are as similar as possible. In this case the fit is imperfect, and we need a measure for the goodness-of-fit. Normally the correlation between the corresponding directions is used for this purpose. This correlation is called the *canonical correlation* and the directions are called the *canonical variates*. The canonical variates are one-dimensional subspaces of $\mathbf{L}_1$ and $\mathbf{L}_2$ and will be denoted by $\mathbf{H}_1\mathbf{a}_1$ and $\mathbf{H}_2\mathbf{a}_2$. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are called weight vectors or *canonical weights*. $\mathbf{a}_1$ consists of m_1 elements and $\mathbf{a}_2$ of m_2 elements. After finding the best linear combinations in $\mathbf{L}_1$ and $\mathbf{L}_2$, we may look for the second best linear combinations, and so on p times. We therefore talk about weight matrices

†The present article is an expanded and revised version of a paper presented at the European Meeting of the Psychometric Society in Groningen (The Netherlands) in June 1980.

0007-1102/83/010054 27 \$02.00/0 © 1983 The British Psychological Society

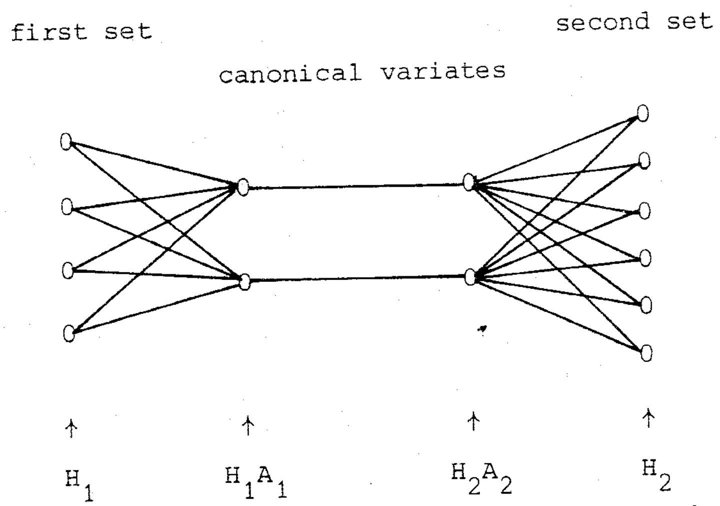

Figure 1. Schematic representation of canonical correlation analysis.

$\mathbf{A}_1(m_1 \times p)$ and $\mathbf{A}_2(m_2 \times p)$. Schematically canonical correlation analysis can be represented as in Fig. 1 (van der Geer, 1971).

Usually only those linear combinations are chosen which form an *orthogonal basis* for $\mathbf{L}_1$ or $\mathbf{L}_2$. Thus the canonical correlation problem is:

> Find weight matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ in such a way that the columns of $\mathbf{H}_1\mathbf{A}_1$ and $\mathbf{H}_2\mathbf{A}_2$ are as similar as possible and so that $\mathbf{H}_1\mathbf{A}_1$ and $\mathbf{H}_2\mathbf{A}_2$ form orthogonal bases.

Starting from *standardized* variables and using a *least squares* formulation, the canonical correlation problem is:

minimize	$(\mathrm{SSQ}(\mathbf{H}_1\mathbf{A}_1 - \mathbf{H}_2\mathbf{A}_2))/np$ over $\mathbf{A}_1$ and $\mathbf{A}_2$
while	$\mathbf{A}_1'\mathbf{H}_1'\mathbf{H}_1\mathbf{A}_1 = n\mathbf{I}$,
	$\mathbf{A}_2'\mathbf{H}_2'\mathbf{H}_2\mathbf{A}_2 = n\mathbf{I}$,
	$\mathbf{h}_j'\mathbf{h}_j = n, \quad j = 1, \ldots, m$
	$\mathbf{h}_j'\mathbf{e} = 0$

Least squares formulation of canonical correlation analysis

The expression $(\mathrm{SSQ}(\mathbf{H}_1\mathbf{A}_1 - \mathbf{H}_2\mathbf{A}_2))/np$ is called the *loss function* or *stress* and is abbreviated as $\sigma(\mathbf{H}, \mathbf{A})$; $\mathrm{SSQ}(\mathbf{X})$ is the sum of squares of the elements of $\mathbf{X}$ and $\mathbf{e}$ is a vector with n ones.

2. CANALS: non-linear canonical correlation analysis

The canonical correlation analysis discussed so far can be viewed as *linear* CCA. Linear refers to the fact that the CCA results are invariant under linear transformations of the data. We would now like to introduce *non-linear canonical correlation analysis*, a method realized in the program CANALS (de Leeuw, 1973; van der Burg & de Leeuw, 1978). Non-linear CCA is canonical correlation analysis with results invariant under certain *non-linear transformations* of the data. More specifically we define non-linear CCA as a method that looks for weight matrices $\mathbf{A}_1$

and $\mathbf{A}_2$, exactly like classical CCA, but at the same time looks for an *optimal scaling* of the data. The data consist of variables that can take only a finite number of values, the *category scores*. In non-linear CCA the category scores of each variable are rescaled in such a way that the rescaled variables belong to a space of 'permitted transformations' (the *optimal scaling space*). The rescaled values of the category scores, the *category quantifications*, are optimal in the sense that they optimize the canonical correlations, and we can therefore speak of *optimally scaled* variables.

The type of non-linear transformations permitted is determined by the assumed *measurement level* of each variable, which can be *discrete nominal*, *ordinal* or *numerical*. Discrete means that we consider the original category scores as coming from a domain of discrete values, instead of a continuous interval. Thus we can speak as well of *category numbers* instead of category scores.

A *discrete* measurement level implies that observations from one category remain in the same category under rescaling, or putting it in another way, observations with the same category number get the same category quantification. If $\mathbf{h}_j$ is an original data vector consisting of n observations, $\mathbf{q}_j$ is a rescaled data vector consisting of n observations and $\sim$ is the relation 'in the same category', the discrete character of a variable is formulated in terms of the following restrictions:

$$h_{ij} \sim h_{kj} \quad \rightarrow \quad q_{ij} = q_{kj}, \quad (i, k = 1, \ldots, n; j = 1, \ldots, m).$$

The *nominal* measurement level imposed on a variable does not imply any further restrictions for that variable.

The *ordinal* measurement level implies also that the order of the original category numbers should be maintained. If $<$ is the empirical order relation of the data vector $\mathbf{h}_j$, the following restriction holds:

$$h_{ij} < h_{kj} \quad \rightarrow \quad q_{ij} \leqslant q_{kj}, \quad (i, k = 1, \ldots, n; j = 1, \ldots, m).$$

The *numerical* measurement level implies that only linear transformations of the data vectors are permitted:

$$q_{ij} = c_j h_{ij} + k_j, \quad (i = 1, \ldots, n; j = 1, \ldots, m; \ c_j, k_j \text{ constants}).$$

For a more detailed treatment of the different measurement levels see Young *et al.* (1976) and Gifi (1980, 1981).

We denote the optimal scaling space, defined by the original data vector $\mathbf{h}_j$, for each variable by $C(\mathbf{h}_j)$ $(j = 1, \ldots, m)$. In the following least squares formulation of the non-linear canonical correlation problem, the vectors $\mathbf{q}_1, \ldots, \mathbf{q}_m$ are parameters, in contrast with the formulation of the classical CCA, where $\mathbf{h}_1, \ldots, \mathbf{h}_m$ were the data. We use the names (original) data vectors or variables for the $\mathbf{h}_j$, and the names rescaled data vectors or optimally scaled variables for the $\mathbf{q}_j$. The term variable is used in a double meaning, in concrete sense as a data vector $\mathbf{h}_j$, and in abstract sense as a characteristic on which objects can have a score (category number). Thus for each variable $j = 1, \ldots, m$ the CANALS program rescales the variables or data vectors $\mathbf{h}_j$ into optimally scaled variables $\mathbf{q}_j$.

Collecting the first m_1 vectors $\mathbf{q}_1, \ldots, \mathbf{q}_{m_1}$ in the columns of $\mathbf{Q}_1$ and the last m_2 vectors $\mathbf{q}_{m_1+1}, \ldots, \mathbf{q}_m$ in the columns of $\mathbf{Q}_2$, and using the same notation as in section 1 for the loss or stress, namely:

$$\sigma(\mathbf{Q}, \mathbf{A}) = (\mathrm{SSQ}(\mathbf{Q}_1 \mathbf{A}_1 - \mathbf{Q}_2 \mathbf{A}_2))/np.$$

the least squares formulation of non-linear CCA is then:

minimize	$\sigma(\mathbf{Q}, \mathbf{A})$ over $\mathbf{A}_1, \mathbf{A}_2$ *and* $\mathbf{q}_1, \ldots, \mathbf{q}_m$
while	$\left.\begin{array}{l}\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I},\\ \mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I},\end{array}\right\}$ orthogonality restrictions
	$\mathbf{q}_j'\mathbf{q}_j = n,$
	$\mathbf{q}_j'\mathbf{e} = 0, \quad j = 1, \ldots, m$
	$\mathbf{q}_j \in C(\mathbf{h}_j).$

Least squares formulation of non-linear canonical correlation analysis

3. The ALS-algorithm and its subproblems

The loss function $\sigma(\mathbf{Q}, \mathbf{A})$, as defined in the preceding section, has two major types of parameters: the weight matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, called *model parameters*, emanating from the canonical correlation problem, and the matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, called *scaling parameters*, because they represent the scaling part of the non-linear CCA. The non-linear canonical correlation problem can be solved in many ways. We prefer an *alternating least squares* (ALS) method. Characteristic of ALS-methods is that they solve least squares problems *iteratively* by splitting up the complete set of parameters into 'nice' subsets and *solving subproblems* for subsets of parameters at a time (holding the other parameters fixed at their present value).

This treatment of non-linear CCA agrees with the treatment of additive structure analysis (ADDALS) of de Leeuw *et al.* (1976), regression analysis (MORALS/CORALS) of Young *et al.* (1976), principal components analysis (HOMALS and PRINCALS) of de Leeuw & van Rijckevorsel (1980) and three way PCA (TUCKALS) of Kroonenberg & de Leeuw (1980). In this series canonical correlation analysis with m sets (OVERALS), canonical discriminant analysis (CRIMINALS), partial canonical correlation analysis (PARTALS) and path analysis (PATHALS) will follow (Gifi, 1980, 1981).

3.1. *Defining the sequence of subproblems*

As a result of the large number of parameters in the non-linear canonical correlation problem, there is much freedom in the *choice of the subsets and the order of solving the subproblems.* At a global level, a natural way of partitioning the parameters is into subsets $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$. But there are several ways of alternating:

→1 $\mathbf{A}_1, \mathbf{A}_2$	→1 $\mathbf{A}_1, \mathbf{A}_2$	→1 $\mathbf{A}_1$
└2 $\mathbf{Q}_1, \mathbf{Q}_2$	2 $\mathbf{Q}_1$	2 $\mathbf{Q}_1$
	3 $\mathbf{A}_1, \mathbf{A}_2$	3 $\mathbf{A}_2$
	└4 $\mathbf{Q}_2$	└4 $\mathbf{Q}_2$

Young *et al.* (1976) use the first and second method. They do not realize that the orthogonality constraints on one set are violated as soon as one of the columns of $\mathbf{Q}_1$ or $\mathbf{Q}_2$ is updated. Because the $\mathbf{q}_j$ has to lie in the optimal scaling space $C(\mathbf{h}_j)$, the orthogonality restriction on the set to which the $\mathbf{q}_j$ belongs cannot be kept during iteration. Young *et al.* solve this by reorthogonalizing the corresponding canonical variates, each time $\mathbf{Q}_1$ or $\mathbf{Q}_2$ is updated. But this is against the rules of alternating

$A_2 = S_{xx}^{-} S_{xy} A_1$

least squares methods. So their MORALS/CORALS algorithm is incorrect and consequently divergence may occur. Their algorithm is a proper ALS-method only in the case of one variable in one of the sets.

We choose the third method for the CANALS program. As we want to use the same solutions for the scaling parameters as Young *et al.* use, we meet the same difficulty of violating the orthogonality constraints. To overcome this difficulty we modify the algorithm in such a way that we still have a non-linear generalization of CCA, but not the one suggested in Section 2. The modification we make is minimizing the loss function under *only one orthogonality restriction.* The advantage of this modified algorithm is that the solutions for the parameters are easy to compute. The CANALS formulation of non-linear CCA is the following:

minimize	$\sigma(\mathbf{Q}, \mathbf{A})$ over $\mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{q}_1, \ldots, \mathbf{q}_m$
while	$\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I}$ *or* $\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I}$,
	$\mathbf{q}_j'\mathbf{q}_j = n$,
	$\mathbf{q}_j'\mathbf{e} = 0, \quad j = 1, \ldots, m$
	$\mathbf{q}_j \in C(\mathbf{h}_j)$.

CANALS formulation of non-linear CCA

We now have to prove that the CANALS problem formulated above is a real generalization of the linear CCA problem formulated in Section 1. Furthermore we have to show that it does not matter which orthogonality restriction is kept for the solutions of the canonical weights. We will state this in a theorem in the next section.

3.2. *Generalizing linear CCA*

To prove that the CANALS problem is a real generalization of linear CCA as formulated in Section 1 means that we have to show that we get equivalent results for the weight matrices whether we solve the linear CCA problem or the CANALS problem. Define two matrices as *equivalent* if they differ only by a diagonal or orthogonal transformation. The difference between the CANALS problem(s) and the linear CCA problem is in the number of orthogonality constraints, so the following theorem proves the generalization.

Theorem

Minimization of $\sigma(\mathbf{Q}, \mathbf{A})$ over $\mathbf{A}_1$ and $\mathbf{A}_2$ with

$$\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I} \quad \text{and} \quad \mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I}, \tag{1}$$

minimization of $\sigma(\mathbf{Q}, \mathbf{A})$ over $\mathbf{A}_1$ and $\mathbf{A}_2$ with $\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I}$. (2)

minimization of $\sigma(\mathbf{Q}, \mathbf{A})$ over $\mathbf{A}_1$ and $\mathbf{A}_2$ with $\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I}$. (3)

all give equivalent results for $\mathbf{A}_1$ and $\mathbf{A}_2$.

Proof

We suppose for convenience that the columns of $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are linearly independent. The proof can easily be generalized to the case in which we do not make this assumption. Solving problem (1) for $\mathbf{A}_1$ and $\mathbf{A}_2$ is a standard canonical correlation problem (see, for instance, van de Geer, 1971). The parameters $\mathbf{A}_1$ and $\mathbf{A}_2$ can be expressed in terms of the eigenvectors of the matrix:

$$\mathbf{T} = (\mathbf{Q}_1'\mathbf{Q}_1)^{-\frac{1}{2}}\mathbf{Q}_1'\mathbf{Q}_2(\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}.$$

Suppose the truncated singular value decomposition of this matrix is (Stewart, 1973):

$$\mathbf{T} = \mathbf{Z}\mathbf{\Lambda}\mathbf{W}'.$$

with $\mathbf{Z}$ a $m_1 \times p$ matrix and $\mathbf{W}$ a $m_2 \times p$ matrix, both with orthonormal columns, and with $\mathbf{\Lambda}$ a $p \times p$ diagonal matrix containing the p largest eigenvalues. Then the loss function is minimal for the following values of $\mathbf{A}_1$ and $\mathbf{A}_2$:

$$\mathbf{A}_1{}^1 = \sqrt{n}(\mathbf{Q}_1'\mathbf{Q}_1)^{-\frac{1}{2}}\mathbf{Z}.$$

$$\mathbf{A}_2{}^1 = \sqrt{n}(\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}\mathbf{W}.$$

We have added a superscript to $\mathbf{A}_1$ and $\mathbf{A}_2$ to show that we have solved problem (1). The solution of problem (2) is seen by reformulating the sum of squares:

$$\begin{aligned}&\mathrm{SSQ}(\mathbf{Q}_1\mathbf{A}_1 - \mathbf{Q}_2\mathbf{A}_2)\\ &\quad = \min \mathrm{tr}\,(n\mathbf{I} - 2\mathbf{A}_1'(\mathbf{Q}_1'\mathbf{Q}_1)^{\frac{1}{2}}\mathbf{T}(\mathbf{Q}_2'\mathbf{Q}_2)^{\frac{1}{2}}\mathbf{A}_2 + \mathbf{A}_2'(\mathbf{Q}_2'\mathbf{Q}_2)^{\frac{1}{2}}(\mathbf{Q}_2'\mathbf{Q}_2)^{\frac{1}{2}}\mathbf{A}_2).\end{aligned}$$

This expression is minimized over $\mathbf{A}_2$, unrestricted, for

$$\mathbf{A}_2{}^2 = (\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}\mathbf{T}'(\mathbf{Q}_1'\mathbf{Q}_1)^{\frac{1}{2}}\mathbf{A}_1.$$

For $\mathbf{A}_1$ the same results hold as for (1). Thus:

$$\mathbf{A}_1{}^2 = \sqrt{n}(\mathbf{Q}_1'\mathbf{Q}_1)^{-\frac{1}{2}}\mathbf{Z}.$$

$$\mathbf{A}_2{}^2 = \sqrt{n}(\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}\mathbf{W}\mathbf{\Lambda}.$$

The solution of problem (3) is analogous to that of problem (2):

$$\mathbf{A}_1{}^3 = \sqrt{n}(\mathbf{Q}_1'\mathbf{Q}_1)^{-\frac{1}{2}}\mathbf{Z}\mathbf{\Lambda}.$$

$$\mathbf{A}_2{}^3 = \sqrt{n}(\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}\mathbf{W}.$$

We see that $\mathbf{A}_1{}^1$ and $\mathbf{A}_2{}^1$ are equivalent to $\mathbf{A}_1{}^2$ and $\mathbf{A}_2{}^2$ and also to $\mathbf{A}_1{}^3$ and $\mathbf{A}_2{}^3$. Note that we find different results for the minimum stress. By substituting the results for $\mathbf{A}_1$ and $\mathbf{A}_2$ we get:

$$\min \text{stress} = 2 - \frac{2}{p}\sum_{s=1}^{p}\lambda_s \quad (1),$$

$$\min \text{stress} = 1 - \frac{1}{p}\sum_{s=1}^{p}\lambda_s^2 \quad (2 \text{ and } 3).$$

3.3. *The difference between the various generalizations*

Let us denote the three generalizations of linear CCA by:

(1) the original non-linear CCA problem (Section 2),
(2) the CANALS formulation with the first restriction,
(3) the CANALS formulation with the second restriction.

In the preceding section we minimized the loss function over the canonical weights; substitution of the various solutions gives a stress (as a function of the scaling parameters $\mathbf{Q}$) for every problem. By substituting the canonical weights the imposed orthogonality restrictions are automatically satisfied. This means that the problem left after eliminating the weights is minimizing the stress $\sigma(\mathbf{Q})$ over $\mathbf{Q}$, while the columns of $\mathbf{Q}$ are standardized and belong to the optimal scaling space, conditions which are the same for all three problems. As we saw in the preceding section, the stress values (minimized over weights) are:

$$\sigma_1(\mathbf{Q}) = 2 - \frac{2}{p}\sum_{s=1}^{p} \lambda_s,$$

$$\sigma_2(\mathbf{Q}) = 1 - \frac{1}{p}\sum_{s=1}^{p} \lambda_s^2,$$

$$\sigma_3(\mathbf{Q}) = 1 - \frac{1}{p}\sum_{s=1}^{p} \lambda_s^2.$$

We see that $\sigma_2 = \sigma_3$, which means that both the CANALS problems are the same after eliminating the canonical weights, and thus give the same results for $\mathbf{Q}$. We also see that $\sigma_1 \neq \sigma_2$, which means in general that σ_1 and σ_2 have different minima and also that these minima occur for different values of $\mathbf{Q}$. An exception is the case that the number of dimensions equals one ($p = 1$). Thus if $p > 1$ the solutions for the scaling parameters of (1) are different from those of (2) and (3). In the previous section we found that the solutions for the canonical weights of the three problems are equivalent, but as the solutions for the scaling parameters of the original non-linear CCA problem and the CANALS problem(s) are different, the solutions of (1) and (2) are not comparable any more. However, the two CANALS solutions are comparable: the solutions for the scaling parameters are identical and the solutions for the canonical weights are equivalent.

3.4. *Switching from one condition to the other*

As stated in Section 3.1, we use the third sequence of parameter subsets for the ALS-method of the CANALS program. We could choose one orthogonality restriction and minimize the loss function over the parameter subsets consecutively. But it is difficult to solve for parameter subsets which are also restricted by an orthogonality constraint. So the trick in the CANALS program is that we minimize the loss function over the parameters of one set, while keeping the orthogonality restriction on the other set. The following scheme gives the order of solving the parameters and the corresponding restriction. Remember that, when we solve for say $\mathbf{A}_1$, the other parameters $\mathbf{A}_2$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are fixed.

→1 $\mathbf{A}_1$ with $\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I}$,
2 $\mathbf{Q}_1$ with $\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{I}$,
3 $\mathbf{A}_2$ with $\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I}$,
└4 $\mathbf{Q}_2$ with $\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I}$.

We switch from one orthogonality restriction to the other during the iteration process. We do this by transforming the canonical weights in such a way that the other condition holds without changing the stress value. This can be done using the transformation matrix $\mathbf{S}$ which satisfies

$$\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{S}\mathbf{S}'$$

if the first restriction holds, or using the matrix $\mathbf{R}$ which satisfies

$$\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{R}\mathbf{R}'$$

if the second condition holds. Here all the parameters $\mathbf{A}$ and $\mathbf{Q}$ are fixed. Every symmetric matrix can be split into a product of two regular matrices. In the CANALS program we use a Choleski decomposition (Stewart, 1973). If

$$\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{A}_1 = n\mathbf{I} \quad \text{and} \quad \mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2 = n\mathbf{S}\mathbf{S}'$$

the matrices

$$\mathbf{A}_1^+ = \mathbf{A}_1\mathbf{S} \quad \text{and} \quad \mathbf{A}_2^+ = \mathbf{A}_2(\mathbf{S}')^{-1}$$

are the transformed weights which give the same stress and the other condition, because:

$$\mathrm{SSQ}(\mathbf{Q}_1\mathbf{A}_1 - \mathbf{Q}_2\mathbf{A}_2) = \mathrm{tr}\,(n\mathbf{S}\mathbf{S}' + n\mathbf{I} - 2\mathbf{A}_1'\mathbf{Q}_1'\mathbf{Q}_2\mathbf{A}_2) = \mathrm{SSQ}(\mathbf{Q}_1\mathbf{A}_1^+ - \mathbf{Q}_2\mathbf{A}_2^+)$$

and

$$\mathbf{A}_2^{+\prime}\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2^+ = (\mathbf{S}^{-1})\mathbf{A}_2'\mathbf{Q}_2'\mathbf{Q}_2\mathbf{A}_2(\mathbf{S}^{-1})' = n\mathbf{I}.$$

3.5. *Satisfying both conditions*

Since a solution, asymmetric in the weights, is somewhat difficult to interpret it is natural to transform the solution for the canonical weights to values which satisfy both conditions. We saw in Section 3.2 that the CANALS problem under the second restriction has solutions for $\mathbf{A}$ (and fixed $\mathbf{Q}$) which equal:

$$\mathbf{A}_1^+ = \sqrt{n}(\mathbf{Q}_1'\mathbf{Q}_1)^{-\frac{1}{2}}\mathbf{Z},$$
$$\mathbf{A}_2^+ = \sqrt{n}(\mathbf{Q}_2'\mathbf{Q}_2)^{-\frac{1}{2}}\mathbf{W}\mathbf{\Lambda}.$$

The diagonal transformation $\mathbf{\Lambda}^{-1}$ of the second weight matrix ensures that both conditions hold. By this transformation the stress value will change to a value

$$\text{stress} = 2 - \frac{2}{p}\sum_{s=1}^{p}\lambda_s,$$

which is the stress of the original non-linear CCA problem. This does not mean that the solutions we found are the solutions of the original problem. This is true for the canonical weights only when the scaling parameters are fixed. We showed in Section 3.3 that the CANALS solutions are not comparable with the solutions of the original non-linear CCA problem and the above transformation does not really change the solutions, but merely changes the scale of the canonical weights of the second set.

4. The solution of the subproblems

As a result of the alternating least squares method, the parameter subsets are not uniquely notated by $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$. We need an extra index to mark the update number. We use a superscript for this purpose. If we use a zero index for the initial values, the order in computing the CANALS parameters is:

$\mathbf{A}_1{}^0, \mathbf{A}_2{}^0, \mathbf{Q}_1{}^0, \mathbf{Q}_2{}^0$
$k = 0$
1 $\mathbf{A}_1{}^{3k+1}$
2 $\mathbf{Q}_1{}^{k+1}$
3 $\mathbf{A}_1{}^{3k+2}, \mathbf{A}_2{}^{3k+1}$
no 4 $\mathbf{A}_2{}^{3k+2}$
5 $\mathbf{Q}_2{}^{k+1}$
6 $\mathbf{A}_1{}^{3k+3}, \mathbf{A}_2{}^{3k+3}$
7 $k = k+1$
yes 8 stop ?
9 $\mathbf{A}_1{}^{3k+1}, \mathbf{A}_2{}^{3k+1}$

In the following sections we discuss the solution for the different parameters, while holding the other parameters constant. We leave all the update numbers and mark the newly computed value of the parameters with a + sign to show that we are dealing with a new update.

4.1. *Model parameters or canonical weights*

The standard solution for the model parameters involves the computation of the pseudo-inverse of the matrix $\mathbf{Q}_1$ or $\mathbf{Q}_2$. Again we prefer an alternating *least squares approximation*. For the weights of the first set this is:

$$\mathbf{A}_1{}^+ = \mathbf{A}_1 + \theta \mathbf{E}_{js},$$

with

$$\mathbf{E}_{js} = \begin{cases} 1 \text{ for element } (j, s), \\ 0 \text{ for all other elements.} \end{cases} \quad (j = 1, \ldots, m_1;\ s = 1, \ldots, p).$$

The parameter θ is found by substituting $\mathbf{A}_1{}^+$ in the loss function and by differentiating the loss over θ:

$$\theta = \mathbf{q}_j'(\mathbf{Q}_1\mathbf{A}_1 - \mathbf{Q}_2\mathbf{A}_2)_s/n. \quad (j = 1, \ldots, m_1;\ s = 1, \ldots, p).$$

$(\mathbf{Q}_1\mathbf{A}_1 - \mathbf{Q}_2\mathbf{A}_2)_s$ is the sth column of matrix $(\mathbf{Q}_1\mathbf{A}_1 - \mathbf{Q}_2\mathbf{A}_2)$. The matrix $\mathbf{A}_1$ can be approximated several times before the next parameter subset is calculated. For $\mathbf{A}_2{}^+$ an analogous procedure can be used. In this case we have:

$$\theta = \mathbf{q}_j'(\mathbf{Q}_2\mathbf{A}_2 - \mathbf{Q}_1\mathbf{A}_1)_s/n. \quad (j = m_1 + 1, \ldots, m;\ s = 1, \ldots, p).$$

4.2. *Scaling parameters*

Till now we denoted the subsets for the scaling parameters as $\mathbf{Q}_1$ and $\mathbf{Q}_2$, but this is a short notation for the *m subsets* $\mathbf{q}_1, \ldots, \mathbf{q}_{m_1}$ and $\mathbf{q}_{m_1+1}, \ldots, \mathbf{q}_m$. In the CANALS program we solve for the optimal values of the parameters $\mathbf{q}_j$, while holding $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}, \mathbf{q}_{j+1}, \ldots, \mathbf{q}_m$ at a fixed value. De Leeuw (1977) proved that optimizing the loss function over normalized $\mathbf{q}_j \in C(\mathbf{h}_j)$ can be solved by optimizing the loss function over unnormalized $\mathbf{q}_j \in C(\mathbf{h}_j)$, and standardizing afterwards. It is possible to divide the solution for scaling parameters into even more steps. To show this the stress must be rewritten. Denote a row of matrix $\mathbf{A}_1$ or $\mathbf{A}_2$ by $\mathbf{a}_j'$.

$$\mathrm{SSQ}(\mathbf{Q}_1\mathbf{A}_1-\mathbf{Q}_2\mathbf{A}_2)=\mathrm{SSQ}\left(\mathbf{q}_j\mathbf{a}_j'-\mathbf{Q}_2\mathbf{A}_2+\sum_{\substack{k=1\\k\neq j}}^{m_1}\mathbf{q}_k\mathbf{a}_k'\right)$$

$$=\mathrm{SSQ}\left(\mathbf{q}_j\mathbf{a}_j'-\frac{1}{\mathbf{a}_j'\mathbf{a}_j}\left(\mathbf{Q}_2\mathbf{A}_2-\sum_{\substack{k=1\\k\neq j}}^{m_1}\mathbf{q}_k\mathbf{a}_k'\right)\mathbf{a}_j\mathbf{a}_j'\right)+\text{constant}$$

$$=\mathrm{SSQ}(\mathbf{q}_j\mathbf{a}_j'-\bar{\mathbf{q}}_j\mathbf{a}_j')+\text{constant}$$

with

$$\bar{\mathbf{q}}_j=\frac{1}{\mathbf{a}_j'\mathbf{a}_j}\left(\mathbf{Q}_2\mathbf{A}_2-\sum_{\substack{k=1\\k\neq j}}^{m_1}\mathbf{q}_k\mathbf{a}_k'\right)\mathbf{a}_j.$$

For $\mathbf{q}_j\in C(\mathbf{h}_j)$ the stress can further split into:

$$\mathrm{SSQ}(\mathbf{q}_j\mathbf{a}_j'-\bar{\mathbf{q}}_j\mathbf{a}_j')=\mathbf{a}_j'\mathbf{a}_j(\mathrm{SSQ}(\mathbf{q}_j-\hat{\mathbf{q}}_j)+\mathrm{SSQ}(\hat{\mathbf{q}}_j-\bar{\mathbf{q}}_j)),$$

with $\hat{\mathbf{q}}_j$ a vector of n elements, consisting of the mean quantifications of $\bar{\mathbf{q}}_j$ over categories, i.e. element i of $\hat{\mathbf{q}}_j$ equals the mean of all those elements of $\bar{\mathbf{q}}_j$ which correspond to the same category number as the ith element. We can now see the steps in solving for $\mathbf{q}_j$ ($j=1,\ldots,m_1$). The first step consists of computing $\bar{\mathbf{q}}_j$, which corresponds to *minimization of the stress over* $\mathbf{q}_j$ *unrestricted*. The second step is the *scaling step* and consists of computing:

(*a*) $\hat{\mathbf{q}}_j$ the vector of category means of the vector $\bar{\mathbf{q}}_j$. This corresponds to the *discrete* restrictions on the variable.
(*b*) $\tilde{\mathbf{q}}_j$ the *regression* of vector $\hat{\mathbf{q}}_j$ on $\mathbf{h}_j$, the original variable. This corresponds to the measurement level restrictions of each variable. The type of regression depends on the assumed variable type.

The *nominal* regression is unrestricted, thus $\tilde{\mathbf{q}}_j=\hat{\mathbf{q}}_j$ (Fig. 2 left).
The *ordinal* regression is monotone because of the order restrictions imposed on the variable. Thus $\tilde{\mathbf{q}}_j$ is a *monotone transformation* of $\mathbf{h}_j$ determined by $\hat{\mathbf{q}}_j$ (Fig. 2 centre).
The *numerical* regression is linear: $\tilde{\mathbf{q}}_j$ is a *linear transformation* of $\mathbf{h}_j$ determined by $\hat{\mathbf{q}}_j$ (Fig. 2 right).

As the vector $\hat{\mathbf{q}}_j$, which should be regressed, can take only k_j different values (k_j = number of categories of variable j), the regression of $\hat{\mathbf{q}}_j$ on $\mathbf{h}_j$ can be a weighted regression of a vector consisting of the k_j category means of $\hat{\mathbf{q}}_j$ on the vector of the k_j different category numbers of $\mathbf{h}_j$. The weights in the regression correspond to the frequencies of each category. Figure 2 shows the three types of scaling. The figures are made as if all the frequencies were one for all categories.

The third step in computing the scaling parameters consists of *standardizing* vector $\tilde{\mathbf{q}}_j$ to unit variance and mean zero. This gives the new update $\mathbf{q}_j^{+}$, the *optimally scaled variable*.

For vector $\mathbf{q}_j$ in the second set the steps are analogous. The roles of $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are interchanged, as for $\mathbf{A}_1$ and $\mathbf{A}_2$.

Theoretically there are many ways of treating *missing observations* (Gifi, 1981). In the CANALS program we choose *multiple categories* for missing observations so that every missing score gets its own quantification. The first and last steps in solving for scaling parameters include missing scores. Consequently the quantification of a missing category is optimal in the sense that it minimizes the stress value. The

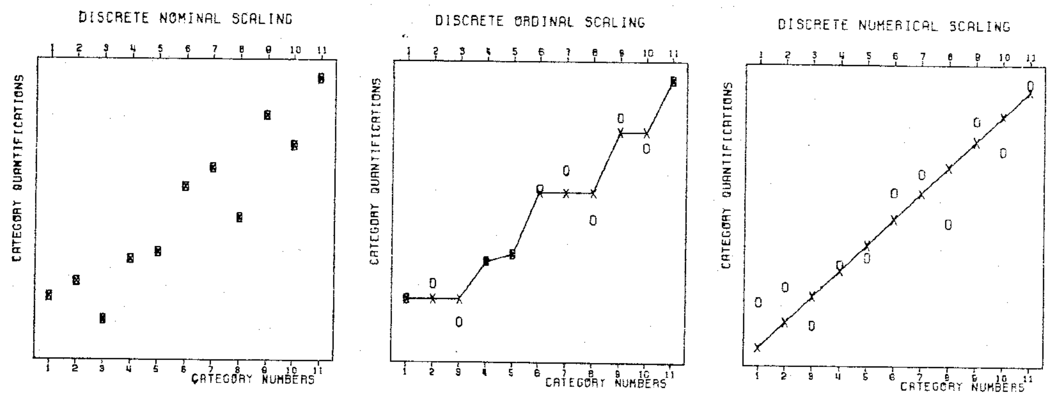

Figure 2. Different types of scaling. (o) category means ($\hat{\mathbf{q}}_j$), × regressed category means ($\tilde{\mathbf{q}}_j$).

second step excludes missing scores. Another way of treating missing observations is to use single categories for missing scores for each variable, or to delete missing scores from the stress. For technical reasons the last two approaches are not implemented in the CANALS program. The single approach can be simulated by making a single category for missing observations for each variable and by treating this variable as nominal with one extra category. The multiple approach has the advantage that it is a very comprehensible and interpretable way of treating missing observations. However the multiple approach also has a disadvantage, as there may arise many *unique patterns* in the data. The program will make as perfect a fit as possible by placing a unique object far away and by placing all the other objects close together. One should always be cautious when a perfect fit occurs as it may be a degenerate solution as exemplified in the following section.

This completes our introduction to non-linear canonical correlation analysis. We have seen that there are many ways of choosing an alternating least squares method. We discussed the choice realized in the CANALS program, as well as estimation methods for the different parameters. In the next section we deal with an application of CANALS and compare non-linear CCA with linear CCA. Finally the stability of the scaling results of this application is studied.

5. Application of CANALS: party preferences and issues in the Dutch Parliament

We use political data from a Parliament Survey to illustrate the CANALS program. Before describing the data we introduce the Dutch political system by quoting the following passages from Daalder & Rusk (1972, pp. 143–144).

> The Netherlands, like many other European states has a parliamentary system. The cabinet as the chief political executive organ has no independent electoral mandate, but must be formed on the basis of the strength of the parliamentary parties. As no single party in the Netherlands has come even near the majority for over half a century, this has meant coalition building among numerous groups. Secondly, an extreme form of national proportional representation does away with any direct electoral link between individual members of parliament and individual constituencies. Voters choose from a large number of rival parties each of which carries on separate national campaigns. Seats are apportioned to parties on the basis of their national percentage of the valid votes. Within parties, candidates are elected largely according to their predetermined rank-order on the party list.
>
> Both the need to sustain cabinets and the absence of direct ties with distinct groups of voters make a party loom large in the perception and behaviour of individual legislators.
>
> The Dutch case seems to offer an example of a political system in which parties have potentially a very high control over the behaviour and perceptions of individual legislators. The dependence

of nomination and renomination on internal party processes could give central party organs a strong weapon with which to discipline deviant behaviour. The electorate gives a mandate to a party, not to individual members of the parliament: defiant members cannot bring their case to individual constituents but must satisfy themselves with such hearing as they can get within the party or leave.

5.1. *Description of the data*

In 1972 the members of the Dutch Parliament (MPs) were interviewed. Among other things, the MPs gave their opinions on a number of issues and their preference votes for the political parties. The issues concerned development aid, abortion, law and order, income differences, worker participation, taxation and defence. The opinions were measured on a nine-point scale of which the lowest and the highest category were described (Table 1). The party preferences were recorded in a table of rank

Table 1. The issues and the meanings of the lowest and the highest category

	1 DEVELOPMENT AID	
the government should spend more money on aid to developing countries	(1) . . . . (9)	the government should spend less money on aid to developing countries
	2 ABORTION	
the government should prohibit abortion completely	(1) . . . . (9)	a woman has the right to decide for herself about abortion
	3 LAW AND ORDER	
the government takes too strong action against public disturbances	(1) . . . . (9)	the government should take stronger action against public disturbances
	4 INCOME DIFFERENCES	
income differences should remain as they are	(1) . . . . (9)	income differences should become much less
	5 PARTICIPATION	
only management should decide important matters in industry	(1) . . . . (9)	workers must also have participation in decisions important for industry
	6 TAXATION	
taxes should be increased for general welfare	(1) . . . . (9)	taxes should be decreased so that people can decide for themselves how to spend their money
	7 DEFENCE	
the government should insist on shrinking the Western armies	(1) . . . . (9)	the government should insist on maintaining strong Western armies

orders. The scores in this table tell us the rank order each member of the parliament gave to the different parties (2 = highest preference, 15 = lowest preference). The lowest score (2) was always used for the MP's own party. For our illustration we only consider the preferences for the four largest parties, which are:

PvdA—labour party (socialists) (39)
ARP —Anti Revolutionary Party (christian democrats) (13)
KVP —catholic party (christian democrats) (35)
VVD —'liberal' party, referred to as conservatives (16)

The figure in parentheses is the number of MPs. The other parties in 1972 were:

CHU —Christian Historical Union (christian democrats) (10)
D'66 —democrats 66 (liberals) (11)

DS70 —democrats 77 (conservative social democrats) (7)
PSP —pacifistic socialists (2)
PPR —radical party (defiant MPs of the KVP) (2)
GPV —conservative calvinistic party (2)
SGP —conservative calvinistic party (2)
BP —farmers party (0)
DJ —conservative 'one man' party (1)
NMP —merchants party (1)
CPN —communist party (0)

In total 141 members of parliament, out of 150, were interviewed. The communists and farmers refused. As far as we know, the party preferences and the opinions have always been analysed separately (see, for instance, van de Geer & de Man, 1974; Daalder & van de Geer, 1977). We wonder whether the opinions of members of parliament, who have, for example, a sympathy for the PvdA, really differ from those of MPs with a sympathy for the VVD, and furthermore which issues can be associated with any such difference. So we use the opinions on the issues and the preferences for the four largest parties for a non-linear canonical correlation analysis. Furthermore we also want to know what advantage we will get from a non-linear analysis rather than a linear one. We therefore analysed the data with the discrete ordinal and numerical scaling options.

5.2. *Canonical correlations and canonical loadings*

Before analysing the data three individuals were removed because they had too many missing scores, leaving 138 members. A two dimensional CANALS analysis with ordinal discrete restrictions shows a canonical correlation of 1·00 in the first dimension. In Section 4.3 we noted that a perfect fit can mean that we are dealing with a unique pattern in the data. This analysis shows a correlation of −1·00 between the (rescaled) variable 'law and order' and the first canonical variate of both sets, and also between the (rescaled) preference for the PvdA and the first canonical variates, whilst the other variables correlate badly. We therefore expect one or more MPs to have a unique score pattern on 'law and order' and preference for the PvdA. There is one person with a missing observation for 'law and order' and a very high and unique preference for the PvdA. When we remove this MP from the data the relation disappears. The following analyses concern the remaining 137 MPs.

The canonical correlations of a two dimensional CANALS analysis with discrete ordinal restrictions and one with discrete numerical restrictions are given in Table 2.

Table 2. Canonical correlations

	Dim. 1	Dim. 2
Ordinal	0·921	0·916
Numerical	0·853	0·726

We see that for both the ordinal and the numerical analysis, the data fit rather well.

As we prefer to look at plots instead of numbers, we plotted the optimally scaled variables in the *canonical spaces* spanned by the canonical variates of each set. The correlations between the optimally scaled variables and the canonical variates, the *canonical loadings*. are standard output of the CANALS program. As the optimally scaled variables and the canonical variates are standardized, the canonical loadings correspond to the *projections* of the optimally scaled variables on the canonical variates. We do not give the figures for the canonical spaces here. Because the canonical correlations of both the ordinal and numerical solution are rather high the canonical loadings of the first set are very similar to the canonical loadings of the second set. Therefore we averaged the canonical variates over the sets so as to get one figure for the two sets together and computed the correlations between the optimally scaled variables and the mean canonical variates. These correlations, also called *component loadings* analogous to principal components analysis, can be computed from the canonical loadings: when multiplied by $\sqrt{[(1+\lambda)/2]}$, the canonical loadings of the optimally scaled variables and canonical variates of the same set give the component loadings of this set. For the first dimension λ is λ_1 and for the second dimension λ is λ_2 (see Table 2). As we also standardized the mean canonical variates the component loadings correspond to the projections of the optimally scaled variables in the mean canonical space (Fig. 3). The horizontal axis is the first mean canonical variate and the vertical axis is the second mean canonical variate. We see in Fig. 3 that 'income differences' and 'participation' are the most important issues after which 'taxation', 'defence' and 'law and order' follow. 'Abortion' and

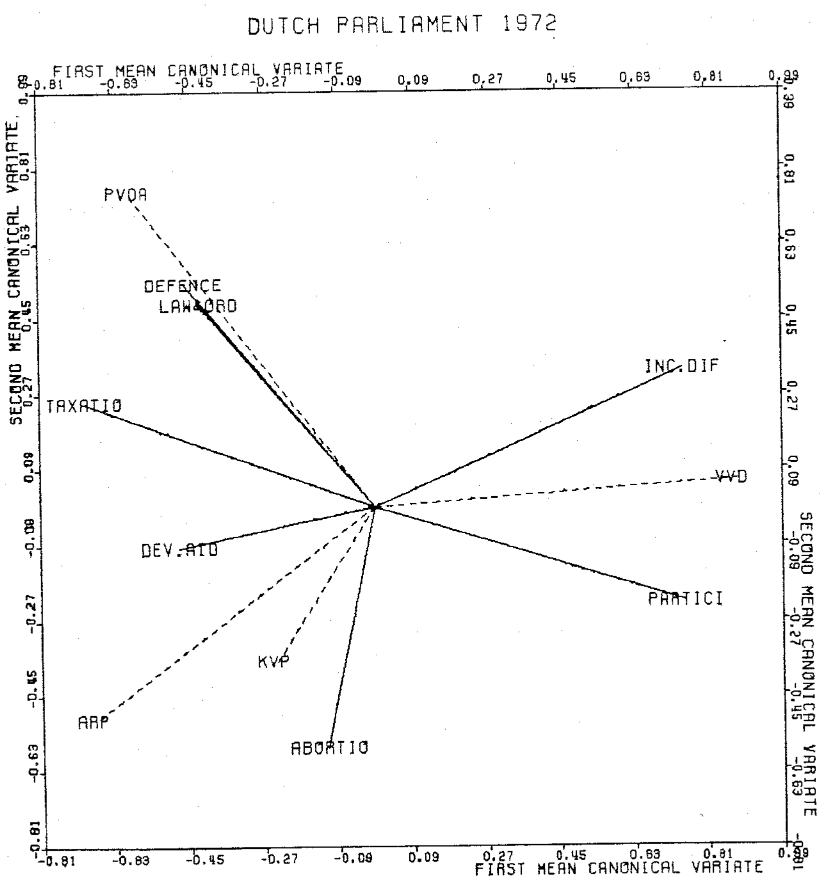

Figure 3. Component loadings: the projections of the optimally scaled variables in the mean canonical space (ordinal solution).

'development aid' seem to be less important. The vectors point in the direction of a high score; for the issues the meaning can be found in Table 1. The preference vectors point in the direction of antipathy. The preferences for the socialists (PvdA), conservatives (VVD) and the protestant christian democrats (ARP) discriminate most with regard to the more important issues. The preference for the catholics (KVP) looks less important, but lies most in the direction of the issue 'abortion'. The vectors can be extended to the other side (with the complementary meaning). If we do this for the ARP-preference we see that 'income differences much less' goes together with a 'great sympathy for the protestant christian democrats'. If an MP is pro liberal abortion, it is not clear whether he or she has a preference for the socialists or the conservatives. Knowing that an MP has a great sympathy for the protestant christian democrats does not give an unequivocal idea of the MPs opinion on the issue 'defence'. Both MPs pro and con 'maintaining strong Western armies' apparently have sympathy for the ARP. The preference for the protestant christian democrats lies between the preference for the socialists and the conservatives. They get sympathy from left and right.

The plot of the component loadings of the numerical solution (Fig. 4) shows that all issues and preferences are equally important. Only 'development aid' seems of minor importance. Again we see the triangle with angle points socialists, conservatives and christian democrats, but now the structure is more clear, as if the

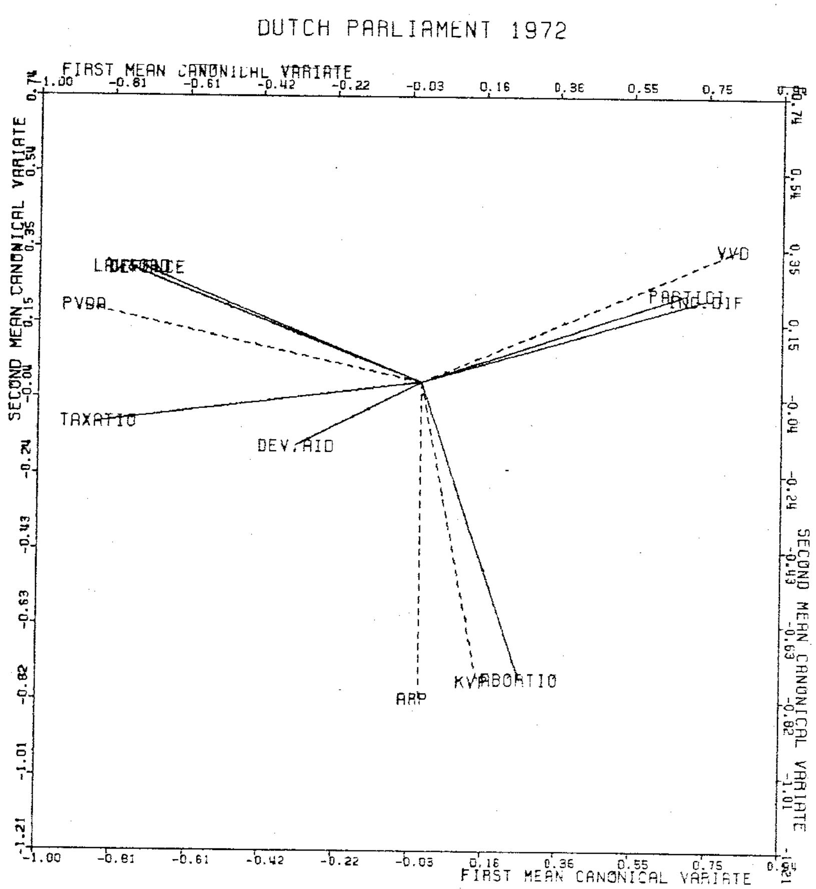

Figure 4. Component loadings: the projections of the optimally scaled variables in the mean canonical space (numerical solution).

numerical scaling options compel the MPs to make the contrasts more extreme. But in fact the contrasts we find in the numerical solution are the same as in the ordinal solution, although more pronounced.

We ignored the canonical weights in the above, preferring to examine the canonical or component loadings for instability of the weights arising from multicollinearity of the data. This instability also happens in multiple regression and linear CCA (see, for instance, Gnanadesikan, 1977, p. 22). The weights are used only for computing the canonical variates. In examining the correspondence between the optimally scaled variables and the canonical variates the canonical loadings suffice (Mulaik, 1972, p. 422).

5.3. *The mean canonical scores*

The *mean canonical scores* are the coordinates of each individual on the mean canonical variates. We plotted all MPs labelled by their party membership in the mean canonical spaces. Figure 5 is the plot of the ordinal solution and Fig. 6 of the numerical solution. We also depicted the rescaled variables as lines in the plots. The projections of the individual points on these lines approximately represent the individual opinions on the issues and the individual preferences. In Fig. 5 we see that the variable 'income differences' more or less separates the conservatives and democrats 70 from the rest of the members; the first group scores negative on 'income differences' and the rest positive. Some D'66-members are negative too. The issue 'abortion' separates the denominational parties (KVP, ARP, CHU, GPV, SGP) from the socialists (PvdA, PSP, PPR) and the liberals (D'66). The issue 'participation' divides the christian democrats (KVP, ARP, CHU) just like the issue 'taxation' does. The large spread of the conservatives (VVD) in the direction of the variable 'abortion' says that the conservatives differ very much in their opinion about abortion. In 1972 two-thirds of the MPs of the VVD had a liberal opinion about abortion and one-third wanted to prohibit abortion more or less. This seems rather peculiar for a conservative party. But the VVD was originally a liberal party. Traditionally many VVD members are pro liberal abortion. In Fig. 5 we left four MPs of the VVD out of the plot. They have such an extreme opinion on abortion and such a great antipathy for the ARP and KVP that they lie completely isolated in the left-hand lower corner. Had they been included the plot would have been so compressed that the labels would have been unreadable. The mean canonical scores of the four outliers are: $(-4{\cdot}112, -3{\cdot}104)$, $(-3{\cdot}251, -1{\cdot}948)$, $(-3{\cdot}645, -3{\cdot}924)$ and $(-4{\cdot}143, -0{\cdot}934)$. With regard to the preferences the MPs of the denominational parties (PVP, ARP, CHU, GPV, SGP) are mostly divided in their preference for the PvdA and the conservatives are divided in their preference for the KVP and ARP. Most of the spread for the socialists (PvdA, PSP, PPR) is also in the direction of the preference for the KVP and ARP. So both conservatives and socialists are divided in their opinion about the social democrats, but both are substantially less positive than the christian democrats themselves.

Figure 6 contains the mean canonical scores of the numerical solution. In both Fig. 5 and Fig. 6, the middle part of the plot is rather empty. When we investigated the canonical loadings we found that, especially for the preferences, the MPs only used the extreme rank orders. They either had a great sympathy or a great antipathy for the four largest parties. Therefore the middle part of the mean canonical space of both the numerical and ordinal solution is rather empty.

The most obvious difference of Fig. 6 from the previous plot is the fact that the

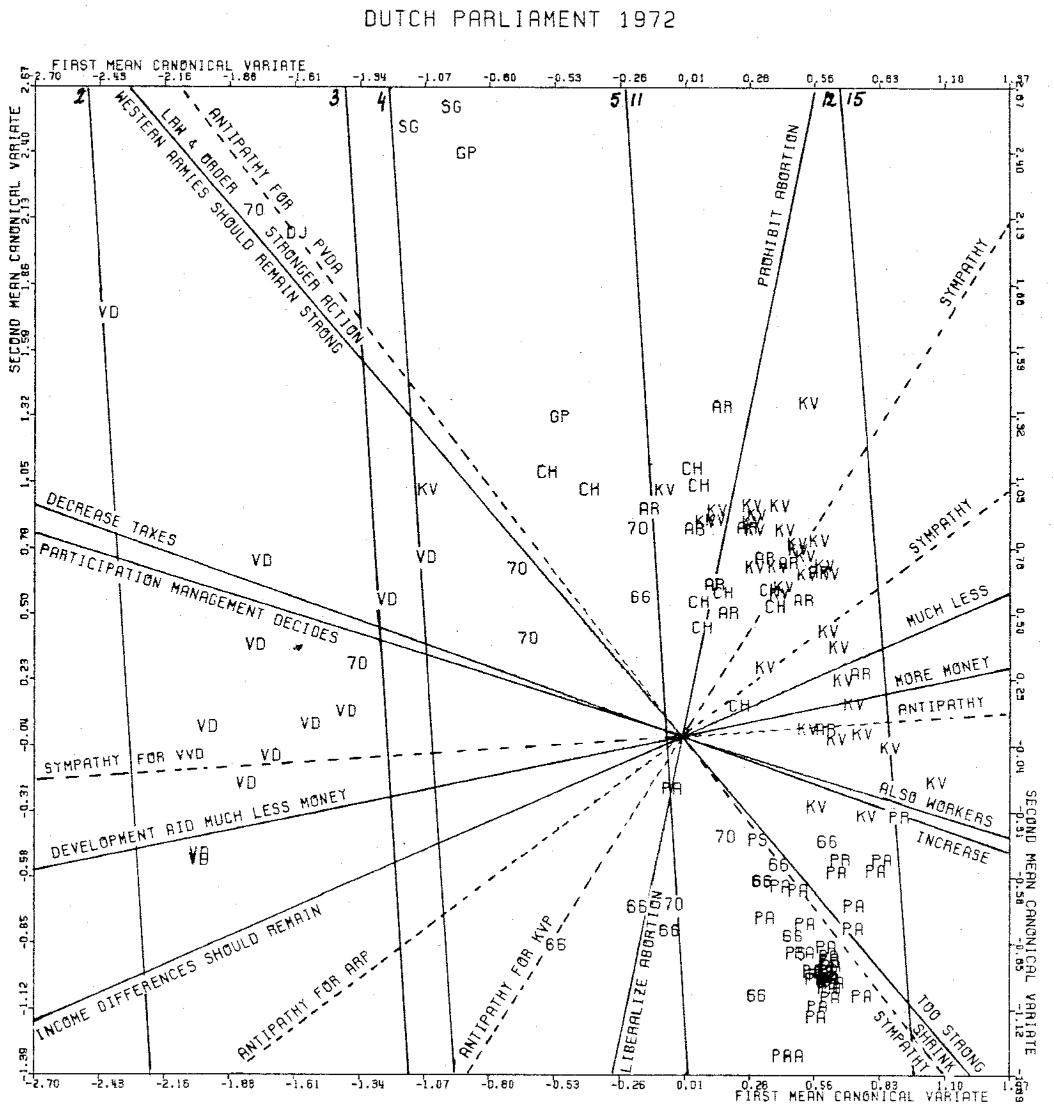

Figure 5. The MPs projected in the mean canonical space (ordinal solution). The individuals are labelled by their party membership (PA = PvdA. AR = ARP. KV = KVP. VD = VVD. CH = CHU. 66 = D'66. 70 = DS70. PS = PSP. PR = PPR. GP = GPV. SG = SGP. DJ). The lines through the origin are the directions of the optimally scaled variables. The parallel lines demonstrate the position of the category numbers of the preference for the VVD.

MPs are more regularly spread over Fig. 6. In an ordinal analysis the scores get a chance to grow close or extreme, which is not possible in a numerical analysis. So in an ordinal solution clustering and 'extremism' can be stronger than in a numerical solution of the same data. But the plot of the component loadings of the numerical solution was clearer than that of the ordinal solution. This is because in the numerical analysis the direction is the most important way to differentiate between the different MPs. In essence we find the same configuration in Fig. 5 as in Fig. 6, but especially on the item 'abortion' and the preference for the KVP and ARP the conservatives are not so divided. The christian democratic MPs still think differently

DUTCH PARLIAMENT 1972

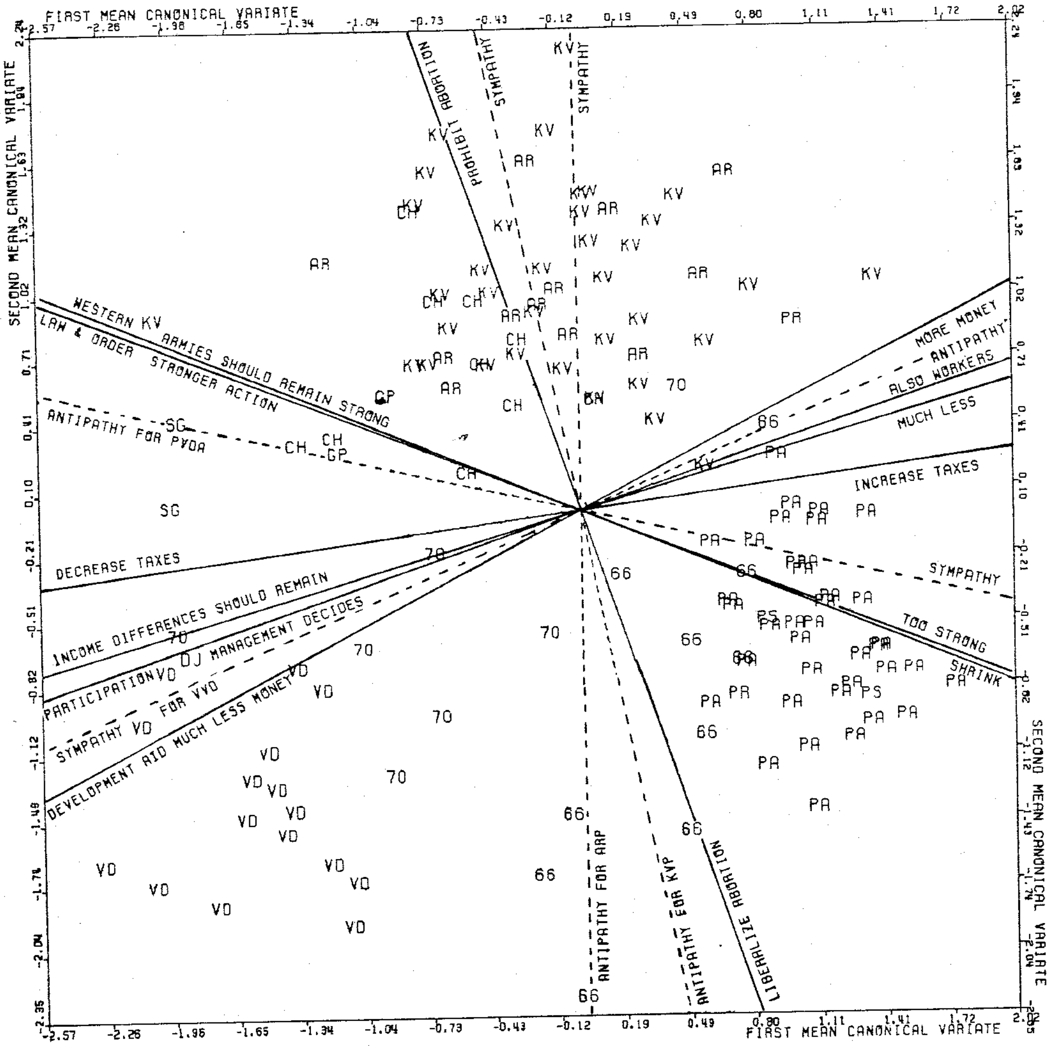

Figure 6. The MPs projected in the mean canonical space (numerical solution). The individuals are labelled by their party membership. The lines are the directions of the optimally scaled variables.

about 'law and order', 'defence' and the preference for the PvdA, but now also about 'taxation', 'income differences', 'participation' and the preference for the VVD. The christian democrats take a middle position between conservatives and socialists, except for matters concerning religion. The conservatives and socialists are clearly political opponents. Because there is no rescaling in the numerical case, we find the MPs in the configuration where we think they ought to be according to the general opinion. The nice thing about the ordinal solution is that it makes the position of the MPs much clearer because of the rescaling of the variables. In the following section we discuss the transformations of the variables.

5.4. *The category quantifications*

To get an impression of the transformations of the variables, we give a plot of the original category numbers against the category quantifications (Fig. 7). The points are connected to show the monotone transformation of each variable (see Section 4.2). For the issue 'development aid' we see for instance that the MPs who agree with

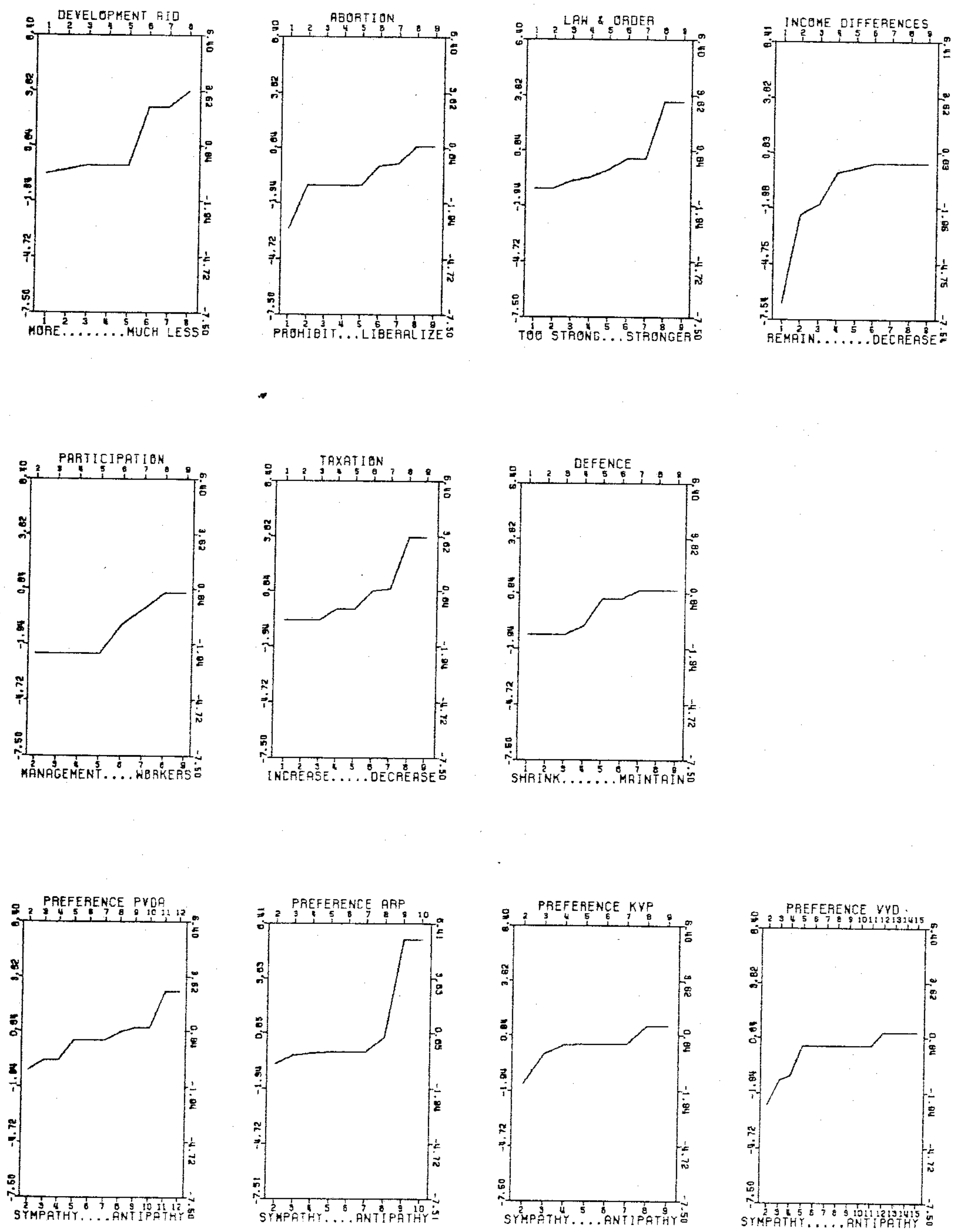

Figure 7. The transformations of the variables (ordinal solution). Original category numbers (horizontal) against category quantifications (vertical).

an increase in development aid are not distinguishable from MPs who have a neutral opinion about it. Scores 1 to 5 are quantified in nearly the same way. We also see that the MPs who want to spend much less money on aid to developing countries are not differentiated either, because scores 6 to 8 are also quantified in nearly the same way. The essence of the scoring is not in being pro or con, but in being neutral/pro or con. The ideas of the christian democrats are not very near to those of the conservatives, although they used the middle of the scale, but are very near to the ideas of the socialists. We could not have concluded this from the numerical solution. Regarding the issue 'income differences' we see that all high scores are quantified in the same way, which means that there is no differentiation in wanting to decrease income differences. It does not give information whether one has a score 4 or 9 on this issue. The opinion of the christian democrats (KVP, ARP, CHU) and the socialists (PvdA, PSP, PPR) is opposed to the opinion of the conservatives (VVD, DS70). In contrast with the high scores, there is a great differentiation in the low scores of the issue 'income differences'. The extent to which a conservative MP wants to keep the income differences as they are, determines his or her place in Fig. 5. From Fig. 6 we know that the christian democratic MPs score mostly in the middle on nearly all variables, but the 'real political colour' is revealed by a non-linear treatment of the data. We give one more example of this phenomenon. If we look at the category quantifications of the preference for the VVD (Fig. 7 and Table 3) we see in fact a four step transformation. The real difference is between (2), (3, 4), (5 to 11) and (12 to 15). Lines perpendicular to the preference vector of the VVD are depicted on the correct positions in Fig. 5. These positions agree with the transformed value of the original category numbers on the corresponding vector. Table 3 shows the coordinates of these positions obtained by multiplying the category quantifications with the component loadings.

Table 3. Preference for VVD, category quantifications, component loadings, and category coordinates. Ordinal solution

Category quantifications		Component loadings		Coordinates in mean canonical space	
2	−2·514			−2·220	−0·160
3	−1·252			−1·110	−0·080
4	−1·057	0·883	0·064	−0·933	−0·067
5	0·433			0·382	0·028
6	0·433			0·382	0·028
7	0·433			0·382	0·028
8	0·433			0·382	0·028
9	0·433			0·382	0·028
10	0·433			0·382	0·028
11	0·433			0·382	0·028
12	1·091			0·963	0·069
15	1·095			0·964	0·070

We find that all nuances from sympathy (5) to dislike (11) are gone. From Fig. 6 we should have concluded that the christian democrats are really more positive about the VVD than the socialists, but from Fig. 5 we can see that this is hardly true. So we get more information out of the data if we consider the fact that the scores are only rank orders.

For all the category quantifications in Fig. 7 we could draw lines perpendicular to the vectors to illustrate how the category quantifications correspond to the positions of the MPs in the mean canonical space. But we leave this as an exercise for the reader. Without having the exact numbers, the lines can be constructed by combining Figs 3 and 7.

5.5. *Conclusion of the CANALS analyses*

In both analyses we find the two contrasts that are basic in Dutch politics: 'left–right' and 'denominational–non-denominational'. The left–right contrast corresponds more or less to the different opinions on taxation, participation, income differences, law and order and defence, and is associated with the preferences for the VVD and the PvdA. The denominational–non-denominational contrast corresponds to the different opinions on abortion and the preferences for the KVP and ARP. In the ordinal analysis of the data we find in essence the same configurations as in the numerical analysis, but the numerical solution only confirms the common sense ideas about the different parties. The ordinal solution shows political nuances that were rather unclear. The conservatives (VVD, DS70) are further away from the christian democrats (KVP, ARP, CHU) than the socialists (PvdA, PSP, PPR) are. The conservatives are very divided in their opinion about abortion. The preference of the christian democrats for the VVD equals the preference of the socialists for the VVD. There exists a real difference in opinion on development aid between the conservatives on one side and the christian democrats on the other side. There is no difference in opinion between socialists and christian democrats on this subject. The ordinal scaling of the data reveals the nuances, although both numerical and ordinal solutions show the main tendencies. In both analyses, we can say that the issues and the preferences are predictable from each other rather well. The ordinal solution is the better one, but we knew that beforehand, because more freedom will always lead to a better fit.

6. Bootstrap: investigating stability of results

To investigate the stability of the CANALS results we did a *bootstrap* study on the data of the Dutch Parliament. The bootstrap technique (Efron, 1979) is related to the more familiar jackknife technique (Miller, 1974). Both techniques, based on the idea that the data are a sample from an unknown population, are used to examine the *stability* of results. For this purpose the bootstrap technique employs *random samples with replacement* from the original data. The size of the bootstrap samples is equal to the size of the original sample, which is the total number of objects. The bootstrap technique repeats the analysis method applied to the original data on the bootstrap samples, so that we get replications of all the original results. The replications are used to estimate *confidence intervals* and *bias* of the corresponding results from the original analysis (Gifi, 1981).

6.1. *Stability of the CANALS scaling of the Dutch Parliament data*

We have seen in Section 5.4 that the ordinal CANALS analysis gives us the transformations of the variables (Fig. 7). The bootstrap technique examines the stability of the scaling results from the ordinal CANALS analysis of the Parliamentary data. We do not start from a sample, but from the population, if you like to call it that. Therefore the relation between the original data and the bootstrap samples is straightforward. The nearer the bootstrap mean of a certain category quantification is to the original value of the ordinal analysis, the better. The smaller the variance of this quantification, the better.

We took 30 samples of size 137, with replacement, from the data of the Dutch Parliament. On all these samples we did a CANALS analysis with ordinal restrictions. From all the quantifications for each category we computed the *mean*, the *range* and the *variance*. The range corresponds to the 95 per cent confidence intervals, that is the intervals in which 95 per cent of the quantifications lies. Figure 8 shows the range of all category quantifications (vertical lines), the bootstrap means (×) and the values of the original analysis (o) for both the issues (Fig. 8*a*) and the preferences (Fig. 8*b*). All plots are on the same scale. It can be seen that the range of the category quantifications is generally much larger at the lower or the higher category numbers than at the middle ones. Furthermore some variables spread obviously less in the category quantifications than other variables.

For instance, the preference for the VVD or the issue 'defence' are both very stable. Very stable is used only relatively, in comparison with the other variables, not absolutely. The shape of the curves, whether linear, convex or concave seems generally speaking stable, although some categories have a large range. As might be expected, the frequencies of the different categories have a great influence on the range of the category quantifications. The variance of each category quantification is low for the higher frequencies above two or three and is comparatively large for the lower frequencies. This means that the badly filled categories, which are the top and bottom ones, have a large chance of getting a very high or low quantification and that we should not take this quantification too seriously. The instability is almost completely caused by the low frequency of the corresponding category. Because the variance does not always go along with the range, we checked this by plotting the two against each other. The figure, not given here, forms a nearly straight line of 45 degrees with one outlier. Category 9 of the issue 'taxation' has frequency one, but has a very small range too.

One further thing to investigate is the position of the bootstrap mean relative to the original category quantification of the ordinal analysis. We therefore plot the bootstrap means connected for each variable, as done before with the category quantifications of Fig. 7. Figure 9 shows the means and the original category quantifications in one plot. Firstly we see that the bootstrap means (dotted lines) are also monotonically increasing. This is because the mean of several monotonically increasing lines is still monotonically increasing. Secondly we see that the sharp angles in the original curves are flattened in the bootstrap curves. In fact the bootstrap transformations are more linear than the original transformations of the ordinal analysis. For the issue 'participation' we find a real difference in the lower categories. The ties in the ordinal solution are apparently not very stable, which may be due to the low frequencies of categories 2 and 3, namely two. These are conservative calvinistic MPs and one MP of the VVD who have strong objections to worker participation. The transformations of several variables seem wonderfully stable. Even the tie of 6 and 7 of the issue 'development aid' is still there in the bootstrap means. The original value of this tie block is somewhat higher than the bootstrap value, but the main shape of the transformation still exists. As mentioned before, the concave or convex shape is preserved in all cases. For instance, in the case of preference for the ARP, the curve stays low till score 8 and then suddenly rises. There is no real difference between having sympathy for the ARP and having a little

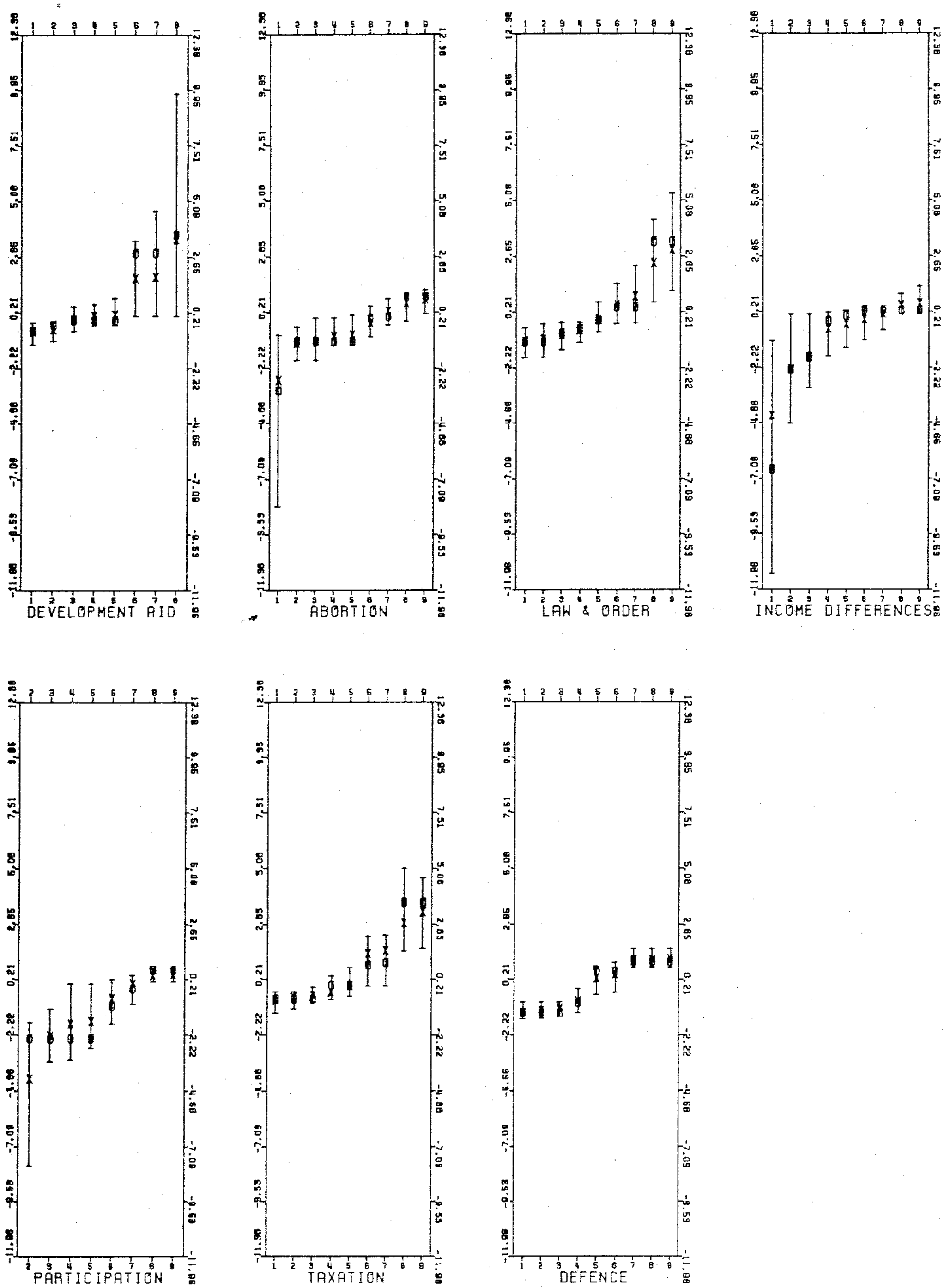

Figure 8(a). The range of the category quantifications (vertical lines), the bootstrap means (×) and the original values (o) for the issues.

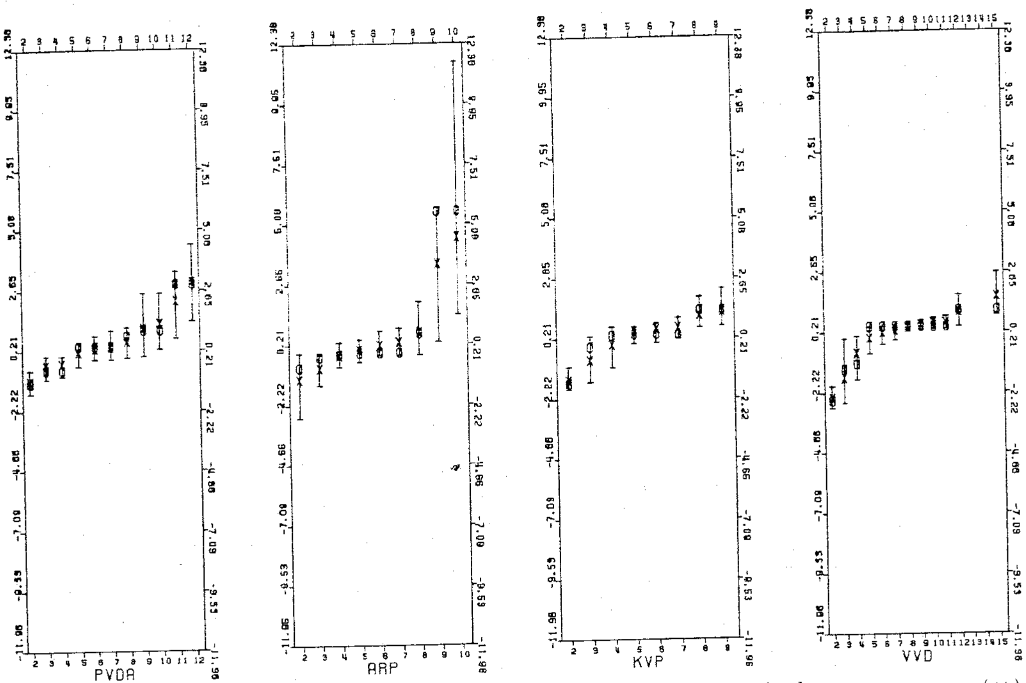

Figure 8(*b*). The range of the category quantifications (vertical lines), the bootstrap means (×) and the original values (o) for the preferences.

aversion, at least not in combination with the issues. The real difference lies between the people with great antipathy and the rest.

6.2. *Conclusion of the bootstrap study*

The bootstrap on the category quantifications of the Dutch Parliament shows that the basic shape of the transformations is stable. It also reveals that the quantifications of the categories with low frequencies are very sensitive under random selection of the individuals. We found that using the 95 per cent confidence intervals gives a good impression of the stability of the transformations of the variables. The bootstrap means are a nice tool for comparing the original values of the parameters with the bootstrap values.

The quantifications of the lowest or the highest category for nearly all variables are unstable due to low frequencies. The categories between the extremes are more stable. The issue 'defence' and the preference for the VVD are more stable than the other variables. For the other variables the bootstrap details differ more or less from the original transformations. For instance, where there are ties in the original ordinal transformations, in most cases they are not found in the mean bootstrap solutions. However the fact that the main shape of the ordinal transformations is preserved in the bootstrap transformations means that the data behave in a rather 'consistent' manner.

7. Conclusion

The non-linear canonical correlation technique gives new possibilities in data analysis when we are interested in the relation between two sets of variables. The greatest

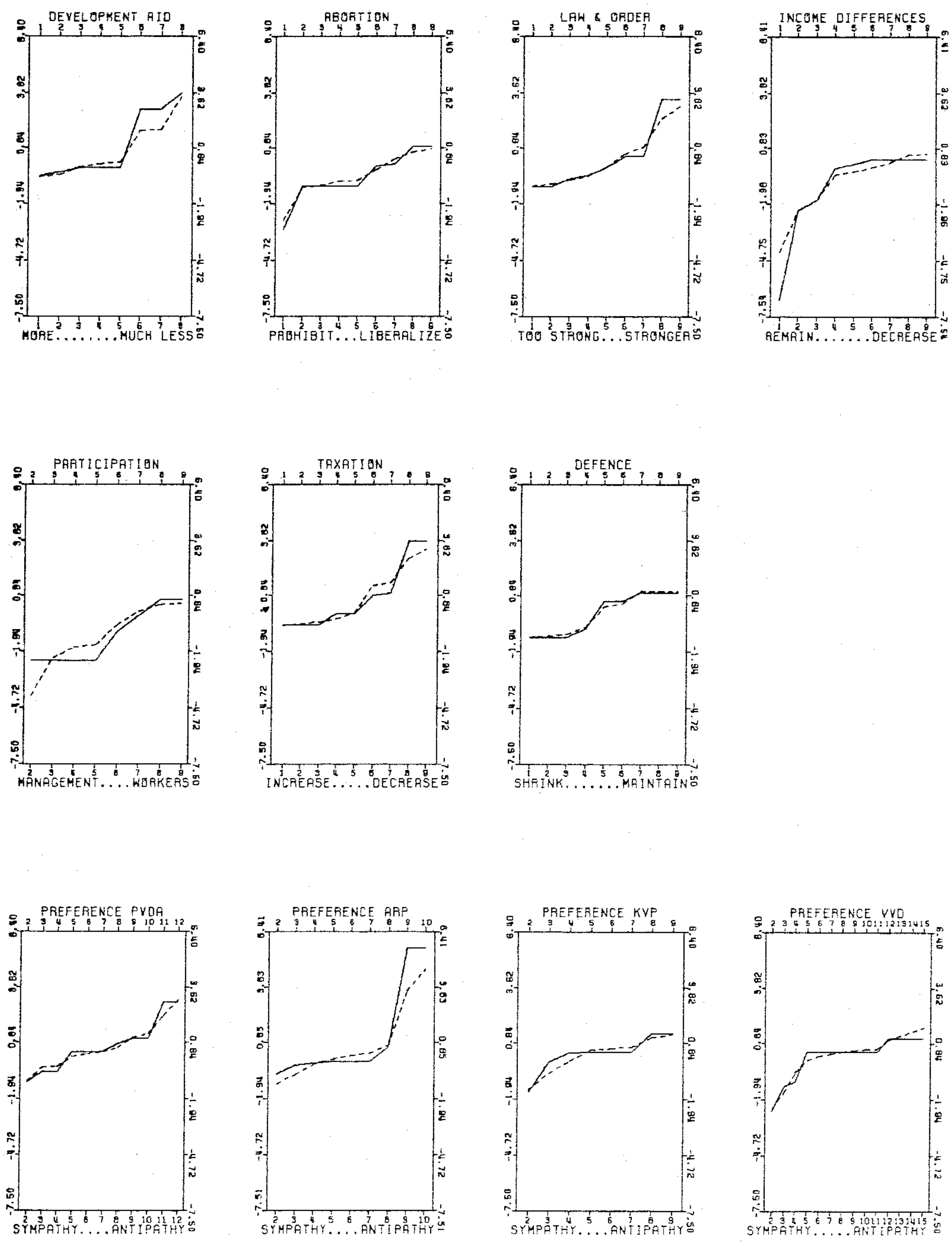

Figure 9. The transformations of the variables (——) and the bootstrap means (- -). Ordinal solution.

advantage of this technique is the fact that we are free to choose the measurement level of each variable separately, so that we do not have to impose strong restrictions on the analysis unnecessarily. The CANALS technique is a real extension and improvement of the technique proposed by Young *et al.* (1976). The alternating least squares algorithm of the CANALS program is a very nice way of avoiding eigenvalue computation and computing results rather quickly and precisely. Another advantage of the CANALS technique is that missing data can be treated in a reasonable way. Many techniques do not have facilities for missing data which are optimal. Missing observations can easily cause unique data patterns that dominate the solution, but with awareness of this phenomenon, it is easy to respond adequately.

The application of CANALS shows us that, in case of the Parliamentary data, a numerical and an ordinal treatment of the variables give analysis results which are very similar. However, it shows us too that the fine nuances in the scoring system get a chance to come to the surface when treating the data on an ordinal measurement level. Finally the bootstrap study of the data of the Dutch Parliament gives some evidence of the stability of the transformations of the variables.

Acknowledgements

This research was supported in part by the SWOV (Stichting Wetenschappelijk Onderzoek Verkeersveiligheid).

The authors gratefully thank Willem Heiser for his helpful comments and Steef de Bie for the work he did on the bootstrap study.

References

Adelman, I., Geier, M. & Morris, C. T. (1969). Instruments and goals in economic development. *American Economic Review*, **59**, 409–426.

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Barcikowski, R. S. & Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights and canonical variate-variable correlations. *Multivariate Behavioral Research*, **10**, 353–364.

Bartlett, M. S. (1948). Internal and external factor analysis. *British Journal of Psychology, Statistical Section*, **1**, 73–81.

Buchanan, T. (1979). Canonical Analysis in the Behavioral Sciences. University of Wyoming: Center for Behavioral Studies.

Burg, E. van der & Leeuw, J. de (1978). How to use CANALS. Datatheory, University of Leiden.

Daalder, H. & Geer, J. P. van de (1977). Partijafstanden in de Tweede Kamer. *Acta Politica*, **12**, 289–345.

Daalder, H. & Rusk, J. G. (1972). Perceptions of party in the Dutch Parliament. In S. C. Patterson & J. C. Wahlke (eds), *Comparative Legislative Behaviour: Frontiers of Research*. New York: Wiley.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Finney, D. J. (1956). Multivariate analysis and agricultural experiments. *Biometrics*, **12**, 67–71.

Geer, J. P. van de (1971). *Introduction to Multivariate Analysis for the Social Sciences*. San Francisco: Freeman.

Geer, J. P. van de & Man, H. de (1974). Analysis of responses to issue statements of the Dutch Parliament 1972. Datatheory, University of Leiden. Internal report RB-001-74.

Gifi, A. (1980). Niet-lineaire multivariate analyse. Datatheory, University of Leiden.

Gifi, A. (1981). Non-linear multivariate analysis. Datatheory, University of Leiden.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley.

Hooper, J. W. (1959). Simultaneous equations and canonical correlation theory. *Econometrica*, **27**, 245–256.

Horst, P. (1961). Relations among *m* sets of measures. *Psychometrika*, **26**, 129–149.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321–327.

Kroonenberg, P. M. & Leeuw, J. de (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, **45**, 69–97.

Leeuw, J. de (1973). Canonical analysis of categorical data. Dissertation, University of Leiden.

Leeuw, J. de (1977). Normalized cone regression. Datatheory, University of Leiden. Mimeo.

Leeuw, J. de & Rijckevorsel, J. J. L. van (1980). HOMALS & PRINCALS, some generalizations of principal components analysis. In E. Diday *et al.* (eds), *Data Analysis and Informatics*. Amsterdam: North Holland.

Leeuw, J. de, Young, F. W. & Takane, Y. (1976). Additive structure analysis in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 471–503.
Meridith, W. (1964). Canonical correlations with fallible data. *Psychometrika*, **29**, 55–65.
Miller, R. G. (1974). The jackknife: A review. *Biometrika*, **61**, 1–15.
Mukherjee, B. N. (1966). Application of canonical analysis to learning data. *Psychological Bulletin*, **66**, 9–21.
Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York: McGraw-Hill.
Norman, W. T. (1965). Double-split cross-validation: An extension to Mosier's design, two undesirable alternatives, and some enigmatic results. *Journal of Applied Psychology*, **49**, 348–357.
Stewart, G. W. (1973). *Introduction to Matrix Computations*. New York: Academic Press.
Stewart, D. K. & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, **70**, 160–163.
Thomson, G. H. (1940). Weighting for battery reliability and prediction. *British Journal of Psychology, Statistical Section*, **30**, 357–366.
Thomson, G. H. (1947). The maximum correlation of two weighted batteries. *British Journal of Psychology, Statistical Section*, **1**, 27–34.
Thorndike, R. M. & Weiss, D. J. (1973). A study of the stability of canonical correlations and canonical components. *Educational and Psychological Measurement*, **33**, 123–134.
Tinter, G. (1946). Some applications of multivariate analysis to economic data. *Journal of American Statistical Association*, **41**, 472–500.
Waugh, F. V. (1942). Regression between sets of variables. *Econometrica*, **10**, 290–310.
Weiss, D. J. (1972). Canonical correlation analysis in counseling psychology research. *Journal of Counseling Psychology*, **19**, 241–252.
Wood, D. A. & Erskine, J. A. (1976). Strategies in canonical correlation with application to behavioral data. *Educational and Psychological Measurement*, **36**, 861–878.
Young, F. W., Leeuw, J. de & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 505–529.

Received 12 January 1982; revised version received 27 August 1982

Requests for reprints should be addressed to Eeke van der Burg, Department of Datatheory, Faculty of Social Sciences, University of Leiden, Hooigracht 15, 2312 KM Leiden, The Netherlands.
Jan de Leeuw is also at the above address.