KM 41 (1992)
pg.79-108


# REDUCED RANK REGRESSION AS AN APPROACH TO FACTOR ANALYSIS WITH FIXED FACTORS

**Rien van der Leeden**[1]
**Ab Mooijaart**[1]
**Jan de Leeuw**[2]

## Abstract

This paper examines an approach to fixed factor analysis. It means that we study a model in which the factors are fixed variables. Within this model, each individual can be characterized by a set of factor scores, which are model parameters. In contrast to the usual, random factor model, individual differences are in that way, part of the model. Fixed factor analysis may thus come closer to the aims of behavioral research, where one is often interested in describing the individuals in terms of a small number of factors or where it can be inappropriate to consider the individuals as a random sample from a well-defined population. Fixed factor models already exist for a long time but they have not become very popular. A reason for this is that under the assumption of multivariate normality, the method of maximum likelihood estimation fails. In the model discussed here restrictions are imposed that solve this problem. This is accomplished by defining the fixed factor model as a special version of the reduced rank regression model. All model parameters, i.e. factor scores, loadings and uniquenesses, are estimated by an alternating maximum likelihood algorithm. The proposed technique is illustrated with two real data examples.

**Key words:** Factor analysis, fixed factors, factor scores, reduced rank regression, individual differences, incidental parameters, maximum likelihood.

---

[1] Department of Psychometrics and Research Methodology
University of Leiden. PO Box 9555, 2300 RB Leiden. Tel. 071-273763
e-mail: `VanderLeeden@rulfsw.LeidenUniv.nl`

[2] Departments of Psychology and Mathematics
University of California at Los Angeles (UCLA)

## 1. Introduction

Most often the interest of factor analysts is in finding general structural laws, describing or explaining the observed relations among a set of variables by means of a small number of unobserved factors. The factor analysis model is usually embedded in a statistical framework which means that concepts such as replication and random sampling make sense. Indeed, the model with its so-called *structural* parameters, factor loadings and unique variances, focusses on structural aspects of the data. Although the factor model incorporates the individual scores on the common factors, the factors itself are considered to be random variables. Hence, individual differences are conceived as just 'error'. In fact, the so-called *factor scores* in the common factor model, can never be determined because they are not parameters in the model. If, for some reason, these scores are still desired, they have to be 'estimated' indirectly, utilizing ad-hoc procedures. The only way to derive real estimates of factor scores, is to let individuals add their factor scores as parameters to the model. Fixed factor models contain these *person* parameters.

In this paper we will discuss a factor analysis model with fixed factors, defined as a special case of a class of generalized linear models. This class of models arises from a multivariate reduced rank regression model with a generally parametrized residual covariance matrix and was discussed in De Leeuw, Mooijaart and Van der Leeden (1985) and in Van der Leeden (1990).

## 2. Historical context and further introduction

Factor analysis has had its origin in psychology. With the development of the product moment correlation coefficient at the end of the nineteenth century, Karl Pearson provided scientists with a powerful instrument to establish relationships between variables used to quantify the human behavior for the study of the mind. At first, the correlation coefficient merely served as a descriptive measure. Gradually however, with a rapidly increasing number of studied variables, the attention shifted to theories that could *explain* the observed relationships. Spearman (1904) was one of the first to formulate a theory to account for the intercorrelations of a number of testscores. He assumed a general ability, common to all the tests, and a specific ability, specific to each test. It was the first *common factor* model.

In the twenty years that followed, many authors, such as Burt, Pearson, Thomson, Garnett and Holzinger, made contributions to the 'factors of mind', which culminated in the 'multiple factor analysis' of Thurstone. Underlying these developments, the basic assumption always remained that the correlations among a number of observed variables could be explained by a smaller number of unobserved factors or latent variables, whereas the relations between those factors and the observed variables could be described by a linear model.

At the time of World War II, factor analysis was well established and widely employed throughout the U.S. Army for large-scale testing problems. In the 1950s and early 1960s, factor analysis was frequently used with the aim to bring order and meaning to the many relations among large numbers of variables. According to Harman (1967), its primary goal was to attain scientific parsimony and economy of description. Clearly, in these developments and applications the emphasis was mostly on structural aspects of the tecnique and it seemed very natural to assume that the factors were random variables.

The random factor model became very popular. Lawley (1940) and Rao (1955) connected the model with modern statistical theory. The first practical algorithms to estimate its parameters, i.e. loadings and uniquenesses, were presented by Jöreskog (1967). The method of maximum likelihood (ML) is currently the most popular, providing for consistent and efficient estimates of the structural parameters, with asymptotically a multivariate normal sampling distribution. Finally, the development of factor analysis even resulted in a very general approach to structural modelling of data (cf. Jöreskog, 1969).

This dominant 'structural perspective' has caused for instance, that factor scores received little attention in classical books about factor analysis, such as Thurstone (1947) and Harman (1960, 1967). It is also obscuring the fact that alternative factor models have been developed in which the factors are fixed variables, i.e. the factor scores are fixed unknown parameters which have to be estimated. This kind of *fixed* factor model, is attributed to Young (1941) and generalized by Lawley (1941).

Nevertheless, fixed factor models can not be found very often in the literature. A reason for this observation is that, apart from the fact that most factor analysts had a different conception of the technique and its aims, the fixed score model is more complicated from a statistical point of view. True, Lawley (1941) derived the likelihood equations, but Anderson and Rubin (1956) showed that the likelihood function is unbounded and maximum likelihood estimates do not exist. They also derived alternative estimates of the structural parameters based on the distribution of the covariances of the observed variables. These estimates appeared to be consistent and asymptotically normal, with covariances equal to those of the ML estimates under the random score model. Actually, they also showed that both the ML estimates of the structural parameters under the random score model and under the fixed score model, have the same asymptotic distribution. Anderson (1984) gives an extensive review regarding these topics.

From the results of Anderson and Rubin, it was concluded that the estimates of the structural parameters in the random score model could be applied in all cases, which made it somewhat futile to look for alternative estimation methods for the fixed score model. This fact tended to make the fixed score model even more unpopular, which is quite unfortunate because factor analysis models that incorporate individual factor score parameters, apart from structural parameters, may often come closer to the aims of social scientific research. Frequently, the interest is in the individual differences, i.e. in describing the individuals in terms of a smaller number of factors, and not only in the structure of the data. In fact, Young (1941) and Whittle (1952) even argued that the fixed score model is more suitable for most applications they knew. Indeed, if one is explicitly interested in individual scores or if it is inappropriate to consider the individuals as a random sample from a well-defined population, one has to apply the fixed score model.

Ultimately the distinction between the random and fixed factor model, and the greater popularity of the random model, is perhaps of a philosophical nature. The previously described developments make clgar that factor analysis has been dominated for a long time by the nomothetic approach that is familiar from psychophysics, focussing on general structural laws and considering individual differences as 'error'. The fixed factor model is more idiografic, it can be used to describe individuals succinctly. Hence, it comes closer to the spirit of many factor analytic studies in applied social sciences. It may be clear that the distinction between random and fixed factors has important implications for the recent, so-called 'factor score controversy', a discussion on the indeterminacy of factor scores, to which many authors made contributions. See e.g. Green (1976), Guttman (1955), Harris (1967), Heermann (1963), Schönemann (1971, 1973), Schönemann and Wang (1972), McDonald (1974), Mulaik (1976) and Mulaik and McDonald (1978), McDonald and Mulaik (1979). Obviously, within the fixed factor model this subject can be studied more clearly.

As was argued in a previous subsection, the method of maximum likelihood fails if it is applied directly to the fixed score model. The reason why is well understood (cf. Anderson, 1984). Intuitively this problem can be made clear by realizing that each individual adds her own factor scores to the set of parameters. For this reason, the factor score parameters are called *incidental*. However, if the number of individuals tends to infinity, the number of parameters tends to infinity too. Hence, one can imagine that there is actually too much freedom in the fixed score model.

One way out of the problem is to impose restrictions in the fixed score model and study the effect of these restrictions on the maximum likelihood estimates. The approach discussed in this paper is a restricted form of fixed factor analysis. In situations where the fixed factor model is preferable to the random model, and if the restrictions make sense, our model can be a useful addition to the literature.

## 3. The fixed factor model

Applying matrix notation, the fixed score model can be written as

$$\mathbf{Y} = \mathbf{FB} + \mathbf{E} \tag{1}$$

where $\mathbf{Y}$ is the $N$x$t$ matrix of observations, that is, we have $N$ observations on $t$ variables. $\mathbf{F}$ is the $N$x$s$ matrix containing the factor scores on $s$ factors and $\mathbf{B}$ is the $s$x$t$ matrix of factor loadings. We will assume that $\mathbf{E}$, the matrix of random error components, follows a matrix normal distribution with zero expectation and parameters $\mathbf{I}$ and $\mathbf{\Sigma}$, notated as $\mathbf{E} \sim N(\mathbf{0},\mathbf{I},\mathbf{\Sigma})$ or as $\mathrm{Vec}(\mathbf{E}') \sim N(\mathbf{0},\mathbf{I}_N \otimes \mathbf{\Sigma})$ (cf. Nel, 1977). $\mathrm{Vec}(\mathbf{E}')$ indicates the column vector in which $\mathbf{E}'$ is stacked column by column and $\otimes$ denotes the Kronecker product. Hence, each row of $\mathbf{E}$ is independent, normally distributed with dispersion matrix $\mathbf{\Sigma}$. Because we are discussing a factor model, $\mathbf{\Sigma}$ is diagonal, resulting from the usual assumption of mutually uncorrelated error components. Thus, the unknown, incidental parameters contained in $\mathbf{F}$, the loadings in $\mathbf{B}$, as well as the unique variances in $\mathbf{\Sigma}$ have to be estimated in model (1). This could be accomplished by maximizing the likelihood function. More conveniently, we could minimize f, the negative logarithm of this function, ignoring some irrelevant constants, given by

$$\mathrm{f}(\mathbf{B},\mathbf{F},\mathbf{\Sigma}) = N \log|\mathbf{\Sigma}| + \mathrm{tr}[(\mathbf{Y} - \mathbf{FB})\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{FB})'] \tag{2}$$

The problems that Anderson and Rubin have described regarding the minimization of [2], arise because we can always find $\mathbf{F}$ and $\mathbf{B}$ such that $\mathbf{\Sigma}$ is singular. For instance, choose $\mathbf{f}_{i.} = \mathbf{y}_{i.}$ and $\boldsymbol{\lambda}_{i.} = (1\ 0\ 0\ ...\ 0)'$. This yields $\boldsymbol{\sigma}_{ii} = 0$ and thus $|\mathbf{\Sigma}| = 0$. This results in $\log|\mathbf{\Sigma}| \to -\infty$ and thus $\mathrm{f}(\mathbf{B},\mathbf{F},\mathbf{\Sigma}) \to -\infty$, which means that the likelihood has no maximum. So, ML estimates do not exist, neither for the incidental parameters, nor for the structural parameters.

In order to solve this problem restrictions can be studied either on $\mathbf{\Sigma}$ or on $\mathbf{F}$ and $\mathbf{B}$. That is, restricting $\mathbf{\Sigma}$ to be diagonal appears not to be sufficient. Whittle (1952) proposed to restrict $\mathbf{\Sigma}$ to be scalar, i.e. $\mathbf{\Sigma} = \sigma\mathbf{I}$. This restriction changes f into

$$[3] \qquad f(\mathbf{B},\mathbf{F},\mathbf{\Sigma}) = Nt \log|\sigma| + \sigma^{-1}\mathrm{tr}[(\mathbf{Y} - \mathbf{FB})'(\mathbf{Y} - \mathbf{FB})]$$

Hence, minimizing [3] only gives problems if $\mathbf{Y} = \mathbf{FB}$ and this situation will be very unlikely in practice. Actually, this restriction defines principal components analysis or singular value decomposition. Although this restriction solves the problem, it is not the solution we are looking for because the assumption that all error variances are equal is rather strong and certainly not self-evident in many cases.

Restricting $\mathbf{B}$, the matrix of factor loadings, could provide another type of solution. In confirmatory (random) factor analysis, for instance, $\mathbf{B}$ is usually restricted while $\mathbf{\Sigma}$ is kept diagonal. However, in fixed factor analysis, this will not solve the problem with respect to the incidental parameters, i.e. the factor scores, because the problem remains that it will often be possible to find $\mathbf{F}$ and $\mathbf{B}$ such that at least one column of $\mathbf{Y}$ is fitted exactly. As was explained above, the result of this is (at least) one vanishing error variance.

A third way of solving the problem of the unbounded likelihood, is to restrict the factor scores $\mathbf{F}$. In the next section, we will discuss a class of these restrictions, arising from the formulation of the fixed factor model as a reduced rank regression model with a diagonal error covariance matrix.

An attempt has been taken to solve the problem of ML estimation in *unrestricted* fixed factor analysis. McDonald (1979) gives a method of simultaneous estimation of factor scores and loadings. He defines $f_1(\mathbf{F},\mathbf{B})$ as the minimum of function [2] over all diagonal $\mathbf{\Sigma}$ and $f_2(\mathbf{F}, \mathbf{B})$ as the minimum of (2) over all positive definite $\mathbf{\Sigma}$. He suggests to find $\mathbf{F}$ and $\mathbf{B}$ so that the difference between $f_2(\mathbf{F}, \mathbf{B})$ and $f_1(\mathbf{F}, \mathbf{B})$ is maximized. Because the log likelihood ratio of two partial likelihoods is maximized, the resulting estimates are called *maximum likelihood ratio estimates*. McDonald shows that this log likelihood ratio is bounded below by zero, if and only if $\mathbf{\Sigma}$ is diagonal.

Thus, estimates of both structural and incidental parameters in the fixed factor model can be found minimizing a log likelihood ratio based on an alternative hypothesis. Without reference to such alternative hypothesis, as was pointed out by Anderson and Rubin (1956), the likelihood is unbounded.

However, McDonald shows that the maximum likelihood ratio estimates of the factor loadings and error variances in the fixed factor model are the same as the corresponding ML estimates in the random factor model. According to Anderson and Rubin (1956), this means that these estimates are consistent. It also appears however, that the estimates of the factor scores, the incidental parameters, are not consistent.

Because of these results McDonald states that, regarding the estimation in the fixed factor model, his findings are purely of theoretical interest . They mainly suggest the use of the structural ML estimates of the random model. In general, factor score estimates are arbitrary because the factor scores are not identified.

An additional discussion concerning maximum likelihood ratio estimation is given by Etezadi-Amoli and McDonald (1983).

## 4. A restricted fixed factor model

In this section we will discuss a fixed factor model with restrictions on the factor scores. Suppose $\mathbf{F}$ is decomposed as $\mathbf{XA}$, where $\mathbf{X}$ is a $N$x$r$ matrix of known constants and $\mathbf{A}$ is a $r$x$s$ matrix of weights. Hence we consider a model given by

$$\mathbf{Y} = \mathbf{XAB} + \mathbf{E} \tag{4}$$

If we let $\mathbf{M} = \mathbf{AB}$, then Equation 4 obviously defines a regression model with fixed regressors. If $\mathbf{M}$ is not of full rank and it holds that rank($\mathbf{M}$) $\leq s \leq \min(r,t)$, Equation 4 defines the so-calles *reduced rank regression* model. These kind of regression models are discussed in e.g. Tso (1981) and Anderson (1984). In the field of econometrics, model [4] is called a functional *errors-in-variables* model (cf. Geary, 1948; Tintner, 1946, 1952; Malinvaud, 1964; Gleser, 1981 and Kelly, 1984). From a regression point of view, such as in econometrics, it will be more likely to conceive $\mathbf{M} = \mathbf{AB}$ as defining a special kind of regression model. In this factor analysis context however, it is very natural to interpret the decomposition $\mathbf{F} = \mathbf{XA}$ as a restriction on the factor scores.

Several possibilities to impose linear restrictions are imaginable. For instance, we can think of $\mathbf{X}$ as a 'design' matrix. By assigning subjects equal rows in $\mathbf{X}$, we require these subjects to have the same factor scores. $\mathbf{X}$ may contain for instance, dummy variables coding some categorical (background) variables. If numerical variables are collected in the columns of $\mathbf{X}$, other columns could consist of powers of these columns, defining a non-linear trend in the factor scores (although the model still remains linear in its parameters). In our illustrations, we will make use of both possibilities.

An important implication of these restrictions is that the factors $\mathbf{F}$ are in the vector space spanned by the variables contained in $\mathbf{X}$. This explains why we can find real factor score estimates. Clearly, the factors are found as linear combinations of the columns of $\mathbf{X}$. The elements of the parameter matrix $\mathbf{A}$ act as weights and are found optimally. Hence, we do not find 'components' concerning the total observed space, as is the case with principal components analysis, but our factors are found within a part of this space. These factors serve as explaining variables regarding the observed interrelations among the y-variables.

A well known example of a (random) (single) factor model, in which the factor is found as a linear combination of some observed variables, is the so-called Multiple Indicators and MultIple Causes (MIMIC) model (Jöreskog & Goldberger, 1975). Jöreskog and Goldberger develop ML and other estimation procedures for a model in which multiple indicators, as well as multiple causes of a single latent variable are observed. In that way, model [4] with $s = 1$, i.e. with $\mathbf{A}$ equal to a column vector, can be considered as a version of the MIMIC model.

Taking $\mathbf{X}$ as a design matrix causes the sample to be divided into distinct groups. Within these groups, the subjects have identical scores on the x-variables, meaning that they also have an identical set of factor scores. If for instance, $\mathbf{X}$ contains dummy variables, coding the information from several explanatory (background) variables, this certainly makes sense. It is reasonable to assume that people of which we have obtained identical score profiles for the x-variables, can, to a certain extent, be considered homogeneous. Hence, in our fixed factor model, the assumption is that, for instance, people having the same religion, the same age, the same income, etc., will respond similarly to an underlying factor, found as an optimal linear combination of these variables.

This assumption seems to be even more reasonable, if we let the number of explaining x-variables increase, since this will reduce the size of the individual groups. It can be attractive, for instance, to add interaction variables to **X**, which possibility we have used in our illustrations. In the extreme case, row profiles in **X** become almost unique, so that virtually every individual is considered as a separate 'group', having its own unique set of factor scores.

## 5 . Estimation and identification

In this section we will develop a stepwise algorithm for the estimation of the parameters of our restricted fixed factor model. Assuming multivariate normality, we will maximize the (full information) likelihood function over the parameters contained in **A**, **B** and $\mathbf{\Sigma}$ utilizing an alternating procedure. Because of the similarity with alternating least squares techniques, we could call our algorithm an *alternating maximum likelihood* (AML) procedure, a terminology also used by De Leeuw (1989).

It should be noted that since we are considering a factor analysis model, $\mathbf{\Sigma}$, the covariance matrix of the error components, is restricted to be *diagonal*. In the algorithm described in the sequel, this restriction leads to an analytical solution for the parameters contained in $\mathbf{\Sigma}$. Since this feature makes the estimation proces not a standard maximum likelihood procedure, we will give a complete description of the algorithm.

At first however, we have to note that basically our model is not identified. It could be written as $\mathbf{Y} = \mathbf{XAZZ}^{-1}\mathbf{B} + \mathbf{E}$, with $\mathbf{Z} \neq \mathbf{I}$ equal to any square, nonsingular matrix of order $s$, so that **A** and **B** are not uniquely determined. Hence, $\mathbf{A}^* = \mathbf{AZ}$ and $\mathbf{B}^* = \mathbf{Z}^{-1}\mathbf{B}$ are proper solutions too. If we require the factor scores **XA** to be orthogonal, a rather common assumption in factor analysis, it can be shown that **A** and **B** are determined up to an orthogonal rotation. In that case it holds that $\mathbf{A'X'XA} = \mathbf{I}$, and so $\mathbf{A}^{*\prime}\mathbf{X'XA}^* = \mathbf{Z'A'X'XAZ} = \mathbf{Z'Z} = \mathbf{I}$, which shows that **Z** must be orthogonal. In the case of $s = 1$, the solutions for **A** and **B** will be determined up to a scalar.

Clearly, whenever we have found estimates of **A** and **B**, these matrices may be subject to orthogonal rotation. Compared to random factor analysis, this fact is hardly surprising. In the next subsections, the orthogonality retriction on **XA** appears to be very useful to simplify the estimation procedure. However, after estimation we could drop this restriction on the factor scores. This is allowed because **Z** may be any square, nonsingular matrix and it implies that, theoretically, estimates for **A** and **B** may also be subject to oblique rotation.

By the restriction $\mathbf{A'X'XA} = \mathbf{I}$ the variances of the factor scores are set equal to $N$. It can easily be shown that by dividing **A** by $\sqrt{N}$ and multiplying **B** by $\sqrt{N}$, an equivalent solution is obtained in which the factor scores have unit variances. In that case, the elements of **B** may be interpreted as corrrelations of the variables with the factors.

For the estimation of the parameters contained in **A**, **B** and $\mathbf{\Sigma}$, we have to minimize function f, given by Equation 2, rewritten as

[5] $$f(\mathbf{A},\mathbf{B},\boldsymbol{\theta}) = N \log|\mathbf{\Sigma}| + \mathrm{tr}[(\mathbf{Y} - \mathbf{XAB})\mathbf{\Sigma}^{-1}(\mathbf{Y} - \mathbf{XAB})']$$

We will write $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$, a common notation to express that the elements of $\mathbf{\Sigma}$ are functions of the unknown covariance parameters collected in the vector $\boldsymbol{\theta}$. Because we are considering a factor model, $\boldsymbol{\theta}$ will only hold parameters for the diagonal elements of $\boldsymbol{\Sigma}$.

First we will find an expression for $\hat{\mathbf{B}}$ in terms of $\mathbf{A}$ and $\boldsymbol{\theta}$. The partial derivative of f with respect to $\mathbf{B}$ is given by

$$\text{[6]} \qquad \partial f/\partial \mathbf{B} = 2\boldsymbol{\Sigma}^{-1}(\mathbf{A'X'Y} - \mathbf{A'X'XAB})$$

Hence, equating $\partial f/\partial \mathbf{B}$ to zero yields as an estimator for $\mathbf{B}$

$$\text{[7]} \qquad \hat{\mathbf{B}} = \mathbf{A'X'Y}$$

under the restriction $\mathbf{A'X'XA} = \mathbf{I}$. Notice that $\hat{\mathbf{B}}$ contains the coefficients of the regression of $\mathbf{Y}$ on $\mathbf{XA}$. Substitution of $\hat{\mathbf{B}}$ simplifies the optimization of f. Instead of minimizing $f(\mathbf{A},\mathbf{B},\boldsymbol{\theta})$, we can minimize the function $g(\mathbf{A},\boldsymbol{\theta})$, given by

$$\text{[8]} \qquad g(\mathbf{A},\boldsymbol{\theta}) = \max_{\mathbf{B}} f(\mathbf{A},\mathbf{B},\boldsymbol{\theta})$$

$$= N \log|\boldsymbol{\Sigma}| + \mathrm{tr}[(\mathbf{Y} - \mathbf{XAA'X'Y})\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{XAA'X'Y})']$$

which simplifies into

$$\text{[9]} \qquad g(\mathbf{A},\boldsymbol{\theta}) = N \log|\boldsymbol{\Sigma}| + \mathrm{tr}[\mathbf{Y'}(\mathbf{I} - \mathbf{XAA'X'})\mathbf{Y}\boldsymbol{\Sigma}^{-1}]$$

under the restriction $\mathbf{A'X'XA} = \mathbf{I}$.

Now the alternating optimization of $g(\mathbf{A},\boldsymbol{\theta})$ is started. In step 1, $\boldsymbol{\theta}$ is fixed and g (the -2 log likelihood) is minimized over the parameters in $\mathbf{A}$. In step 2, $\mathbf{A}$ is fixed and g is minimized over the parameters in $\boldsymbol{\theta}$. These steps are repeated and in each step, the parameters being held fixed, are updated with the results of the previous step. The procedure stops when a certain convergence criterium has been reached, i.e. when the estimates of both $\mathbf{A}$ and $\boldsymbol{\theta}$ have become stable.

*Step 1*

Find $\hat{\mathbf{A}}$ for fixed $\boldsymbol{\theta}$. In the case that $\boldsymbol{\theta}$ is fixed, minimizing g is equivalent to maximizing the function

$$\text{[10]} \qquad g^*(\mathbf{A}|\boldsymbol{\theta}) = \mathrm{tr}[\mathbf{Y'XAA'X'Y}\boldsymbol{\Sigma}^{-1}] = \mathrm{tr}[\mathbf{A'X'Y}\boldsymbol{\Sigma}^{-1}\mathbf{Y'XA}]$$

under the restriction $\mathbf{A'X'XA} = \mathbf{I}$.

Now let $\mathbf{X}$ be decomposed by a Gram-Schmidt orthogonalization as $\mathbf{X} = \mathbf{X}^*\mathbf{T}$, where the columns of $\mathbf{X}^*$ are orthogonal and $\mathbf{T}$ is an upper-triangular matrix. Function $g^*$ can now be written as

$$\text{[11]} \qquad g^*(\mathbf{A}|\boldsymbol{\theta}) = \mathrm{tr}[\mathbf{A'T'X^{*\prime}Y}\boldsymbol{\Sigma}^{-1}\mathbf{Y'X^*TA}]$$

where $\mathbf{A'T'TA} = \mathbf{I}$. If we define $\mathbf{P} = \mathbf{X^{*\prime}Y}\boldsymbol{\Sigma}^{-1}\mathbf{Y'X^*}$, $g^*$ simplifies into

[12] $$g^*(\mathbf{A}|\boldsymbol{\theta}) = \mathrm{tr}[\mathbf{A'T'PTA}]$$

This function has to be maximized for orthogonal **TA**. Because the matrix **A** has $s$ columns, $g^*$ is obviously maximized, if we let **TA** be the eigenvectors corresponding to the $s$ largest eigenvalues of **P**. Let **G** be these eigenvectors, then we can write as an estimator for **A**

[13] $$\hat{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{G}$$

It must hold that $\hat{\mathbf{A}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{A}} = \mathbf{I}$, which condition can easily be verified. Clearly, $\hat{\mathbf{A}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{A}} = \hat{\mathbf{A}}'\mathbf{T}'\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{T}\hat{\mathbf{A}} = \hat{\mathbf{A}}'\mathbf{T}'\mathbf{T}\hat{\mathbf{A}} = \mathbf{I}$, while $\mathbf{G} = \mathbf{TA}$ and $\mathbf{G}'\mathbf{G} = \mathbf{I}$.

*Step 2*

Find $\hat{\boldsymbol{\theta}}$ for given **A**. So function $g(\boldsymbol{\theta}|\mathbf{A})$ (Eq. 9) has to be minimized over $\boldsymbol{\theta}$. Now define $\mathbf{Q} = (1/N)\,\mathbf{Y}'(\mathbf{I} - \mathbf{XAA'X'})\mathbf{Y}$. For given **A**, this matrix can be computed and can be considered a (temporary) estimate of $\boldsymbol{\Sigma}$. Hence, minimizing $g(\boldsymbol{\theta}|\mathbf{A})$ is equivalent to minimizing the function

[14] $$g^{**}(\boldsymbol{\theta}|\mathbf{A}) = \log|\boldsymbol{\Sigma}| + \mathrm{tr}[\mathbf{Q}\boldsymbol{\Sigma}^{-1}]$$

Because we are considering a factor model, $\boldsymbol{\Sigma}$ is diagonal. This means that function $g^{**}$ reduces to

[15] $$g^{**}(\boldsymbol{\theta}|\mathbf{A}) = \sum_i \log(\sigma_{ii}) + \sum_i q_{ii}\,\sigma_{ii}^{-1}$$

Now the partial derivative of $g^{**}$ with respect to $\sigma_{ii}$ is given by

[16] $$\partial g^{**}/\partial\sigma_{ii} = \sum_i \sigma_{ii}^{-1} - \sum_i q_{ii}\,\sigma_{ii}^{-2}$$

Equating $\partial g^{**}/\partial\sigma_{ii}$ to zero yields as an estimator for $\sigma_{ii}$

[17] $$\hat{\theta}_i = \hat{\sigma}_{ii} = q_{ii}$$

It can easily be shown that for $\sigma_{ii} = q_{ii}$ the second order derivative of $g^{**}$ is positive, so for $\sigma_{ii} = q_{ii}$ the minimum of $g^{**}$ is attained. Note that this estimator is only valid if the uniquenesses are non-zero.

The estimation procedure can be summarized as:

(1) Take some starting values for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, i.e. the $\sigma_{ii}$'s.
(2) Estimate **A** for fixed $\boldsymbol{\theta}$, from $\hat{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{G}$. (*Step 1*)
(3) Compute $\mathbf{Q} = (1/N)\,\mathbf{Y}'(\mathbf{I} - \mathbf{XAA'X'})\mathbf{Y}$ and set $\hat{\sigma}_{ii} = q_{ii}$ (*Step 2*)
(4) Update $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}}$, and repeat step 1 and step 2 untill convergence.
(5) Estimate **B** from $\hat{\mathbf{B}} = \mathbf{A'X'Y}$.

In De Leeuw, Mooijaart and Van der Leeden (1985) and Van der Leeden (1990), the estimation procedure described above is discussed more generally. They consider a class of multivariate reduced rank regression models with a general parametrization of the residual covariances. This is accomplished by fitting a covariance structure model on $\mathbf{\Sigma}$. Hence, there is no simple analytical solution available for $\hat{\mathbf{\Sigma}}$. In step 2, such solution requires an iterative procedure. The fixed factor model discussed in this paper can, of course, be considered as a special case of this general setup.

One might consider using the LISREL program (cf. Jöreskog and Sörbom, 1984) (or similar programs like EQS (Bentler, 1989) or LISCOMP (Muthén, 1987)), as an alternative for the estimation procedure for fixed factor analysis. However, this does not appear to be a very fruitful approach. There are a few ways of thinking. First, treat $\mathbf{X}$ as a set of fixed variables, $\mathbf{Y}$ as a set of observed, stochastic variables and use a model with so-called 'fixed x'. In LISREL terminology, this means that one set of latent variables is set equal to $\mathbf{X}$, whereas the other set of latent variables (which are the factors in our model) are found as linear combinations of the variables in $\mathbf{X}$. In that way, if more than one single factor is involved, we have a generalized MIMIC model (see also Sections 4 and 7). This approach is the only way to estimate the regression parameters in $\mathbf{A}$ and $\mathbf{B}$, as well as the covariance parameters in $\mathbf{\Sigma}$, at the same time (it should be noticed here that with $\mathbf{\Sigma}$, we indicate the diagonal covariance matrix of the error components, and *not* the covariance matrix of the observed variables). However, computationally it has the disadvantage that we may have to deal with a huge covariance matrix as input for LISREL (e.g. see Example 1 in the sequel, in which more than 100 dummy variables are considered as columns of $\mathbf{X}$). Moreover, it is a questionable action to compute covariances from dummy variables coded by zeros and ones, containing possibly very small proportions. It will be questionable even when these dummy variables are being preprocessed, by estimating tetrachoric correlations under the assumption of normality of an underlying unmeasured continuous variable.

A second way of dealing with our model in LISREL, is treating it as a multi-sample covariance analysis. In that way the (product) matrix $\mathbf{XAB}$ defines a design matrix dividing the sample in a number of subsamples. This design matrix is partially unknown, because the columns of $\mathbf{X}$ that code the groups are weighted by the parameters contained in $\mathbf{A}$ and $\mathbf{B}$. These parameters can not be estimated in such multi-sample analysis, hence it will only give an approximation of the setup provided by our model, and only as far as the elements of $\mathbf{\Sigma}$ are concerned. Another disadvantage is that one will be confronted with a large number of possibly small, unstable subsamples. One may conclude that this approach is not very attractive.

A third approach could be to perform a LISREL analysis upon the residual covariances based on a least squares solution for $\mathbf{A}$ and $\mathbf{B}$. This is not equivalent to our full information maximum likelihood procedure, but resembles a restricted maximum likelihood procedure.

The alternative ways of dealing with our fixed factor model in LISREL (or similar programs) mentioned above, are treated in more detail in Van der Leeden (1990). It seems to us that the third approach is the most attractive.

## 6. Testing hypotheses

Obviously, it would be desirable to have the possibility of testing hypotheses about different models and to have some measure of the goodness-of-fit. Tests we could use are so-called likelihood-ratio tests. These are defined as follows. Suppose we want to test the null hypothesis $H_0 : \boldsymbol{\theta} \in \boldsymbol{\theta}_0$ against the alternative hypothesis $H_1 : \boldsymbol{\theta} \in \boldsymbol{\theta}_1$, where $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \subset \Omega$, the whole parameter space, and $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are disjoint. Most of the time it holds that $\Omega = \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1$. In those cases we may also write $H_1 : \boldsymbol{\theta} \in \Omega - \boldsymbol{\theta}_0$, which means that no specific alternative hypothesis is considered. Now considering the value $L(\boldsymbol{\theta}_0)$ of the likelihood function maximized under $H_0$, and the value $L(\Omega)$ of the likelihood function maximized over the whole parameter space $\Omega$, the likelihood-ratio is defined as

$$\lambda = L(\boldsymbol{\theta}_0)/L(\Omega) \tag{18}$$

Because $\boldsymbol{\theta}_0 \subset \Omega$, $L(\boldsymbol{\theta}_0) \leq L(\Omega)$ and so $0 \leq \lambda \leq 1$. The value $L(\boldsymbol{\theta}_0)$ is an indication of the 'likelihood' that $H_0$ is 'true', so if $\lambda$ is near 1, the null hypothesis is favored. If $\lambda$ is near 0, other parameter values than those restricted by $H_0$ are more likely and $H_0$ is not tenable. Based on this $\lambda$ a critical region can be constructed, and a likelihood-ratio test is defined. Wald (1943) showed that, for large sample sizes, -2 log $\lambda$ is chi-square distributed with degrees of freedom equal to the number of independent parameters estimated in $\Omega$, minus the number of independent parameters estimated in $\boldsymbol{\theta}_0$. So, it is possible to have a so-called most powerful test.

The likelihood-ratio test formulated without a specific alternative hypothesis, is able to test the goodness-of-fit of a particular model with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. To maximize the likelihood function over $\Omega$, requires the formulation of a so-called null model or fully saturated model which can be derived by letting all parameters unconstrained.

In our case however, specifying an unconstrained model, causes several problems. It requires a matrix $\mathbf{X}$ which is square (order $N$ x $N$) and of full column rank, and at the same time, a matrix $\boldsymbol{\Sigma}$ which is completely free. If $\mathbf{X}$ is square and of full rank, a special case would be if $\mathbf{X} = \mathbf{I}_N$, each individual observation has its own set of model parameters. In that case, however, the likelihood function is unbounded and ML estimates do not exist. So, testing against a model with $\mathbf{X}$ being made 'almost' $\mathbf{I}_N$, or with $\mathbf{X}$ being completed with random columns until 'almost' order ($N$ x $N$) with full column rank, does not seem to accomplish a suitable null model. Besides, this solutions may result in severe computational problems.

Testing against a model with a free error covariance matrix $\boldsymbol{\Sigma}$, which is not a problem in particular and perhaps the most realistic idea, will only provide a test of a specific covariance structure of $\boldsymbol{\Sigma}$. It will not gain insight into the fit of a particular model as a whole. In that case, we will not have the possibility of testing the model constraints imposed by $\mathbf{AB}$, i.e. the rank restrictions, against an absolute null model.

Clearly, a suitable null model is not easily formulated and a proper test of the goodness-of-fit of a particular model is not available. However, what we can do using likelihood-ratio tests, is creating a 'model-hierarchy' of nested models and test these various models against each other, to establish which model is the most appropriate. In that way we consider null hypothesis $H_0 : \boldsymbol{\theta} \in \boldsymbol{\theta}_0$, against alternative hypotheses of the form $H_1 : \boldsymbol{\theta} \in \boldsymbol{\theta}_1$, where $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \subset \Omega$ and $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are disjoint, but where it holds that $\Omega \neq \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1$. So the model under $H_1$ is conceived as some implicit null model and likelihood-ratio tests can be constructed in a similar way as described above.

Nested models can be created by adding columns to $\mathbf{X}$, or by choosing different rank restrictions. For instance, we can add columns with variables defining interactions of variables already conained in $\mathbf{X}$. Choosing a larger $s$ gives a solution with more factors. Relaxing the rank restrictions will allow a larger number of linear combinations of original predictor variables to act as predictors, and so on. Applying likelihood-ratio tests, we can decide if the addition of parameters in these ways provides for a more suitable model. In our illustrations we will make use of such model-hierarchy.

Wald statistics (cf. Buse, 1982), which could provide an alternative way of testing, are not applicable in our model. Using Wald tests one is able to evaluate hypotheses of the kind whether certain parameters can be constrained to zero, instead of letting them unconstrained during estimation. In our case, however, testing such hypotheses is not very useful for two reasons. First, since we consider a factor model, the error covariance matrix $\boldsymbol{\Sigma}$ is restricted to be diagonal, whereas its elements are unrestricted. Hence, we are not interested in constraint (or zero) diagonal elements of $\boldsymbol{\Sigma}$, and we are never considering any unconstraint off-diagonal elements of $\boldsymbol{\Sigma}$. Second, in our model, the other set of parameters in matrices $\mathbf{A}$ and $\mathbf{B}$ is meant to be unconstrained at any time. We do not have the possibility to impose constraints on these parameters, except for the rank restriction. This restriction, however, is not accomplished by the estimation of model parameters, but chosen by the user.

## 7. Relations with other models

The general linear model in Equation 4 can be considered defining a broad class of multivariate analysis techniques. The fixed factor model discussed in this paper is formulated as a submodel of this class. For our purposes, the residual covariance matrix $\boldsymbol{\Sigma}$ is restricted to be diagonal, which results from the assumption of mutually uncorrelated uniquenesses, one of the basic assumptions in factor analysis. The introduction of the fixed matrix $\mathbf{X}$, which in general solves the problem of the unbounded likelihood, can be interpreted as a restriction on the factor scores, causing (small) groups of individuals to have the same set of factor scores. However, manipulating $\mathbf{X}$ and/or structuring $\boldsymbol{\Sigma}$ in different ways, results in a variety of other model interpretations, which we will briefly summarize in this section. A comprehensive discussion of these submodels, including their consequences for parameter estimation is given in Van der Leeden (1990).

If $\boldsymbol{\Sigma}$ is restricted as $\boldsymbol{\Sigma} = \sigma\mathbf{I}$, unknown, diagonal but with equal elements, the technique is equivalent to *redundancy* analysis (cf. Van den Wollenberg, 1977; Israëls, 1987). This restriction is a special case of the general structure $\boldsymbol{\Sigma} = \sigma\boldsymbol{\Sigma}_0$, where $\sigma$ is unknown and $\boldsymbol{\Sigma}_0$ is a known matrix, which states that $\boldsymbol{\Sigma}$ is proportional to $\boldsymbol{\Sigma}_0$ with respect to a parameter $\sigma$. For instance, Gleser (1981) and Kelly (1984) have studied multivariate 'errors-in-variables' regression models with (non)homogeneous error variances, that incorporate this restriction. Here, there is also an interesting relationship with applications to *longitudinal* data. For instance, $\boldsymbol{\Sigma}$ could handle a serial correlation problem arising from, say, a Markov process, in which $\boldsymbol{\Sigma}_0$ is completely defined by the Markov parameter, i.e. $\boldsymbol{\Sigma}_0$ is proportional to the unknown, homogeneous error variance $\sigma$ (cf.

Intriligator, 1978). We could also consider a block-diagonal $\mathbf{\Sigma}$, or specify certain patterns of free and fixed elements in the error covariances.

If it is assumed that $\mathbf{\Sigma}$ is completely free and unknown, the technique relates to *canonical correlation* analysis. For instance, Tso (1981) and Izenman (1975) studied reduced rank regression models under this assumption and stipulated this relationship. If the rank of $\mathbf{XA}$ is restricted to one, we have a version of the so-called MIMIC model. MIMIC with an unconstrained $\mathbf{\Sigma}$ has been discussed by Hauser and Goldberger (1971). Bagozzi, Fornell and Larcker (1981) discuss the relationship between the canonical correlation model and MIMIC within the framework of linear structural relations models.

A special kind of models emerge if $\mathbf{\Sigma}$ is completely known. Such models are considered for instance, in econometrics (cf. Tintner, 1946, 1952; Geary, 1948 and Malinvaud, 1964). These authors study 'errors-in-variables' models to describe economic systems, where in some cases the variances of the all disturbances appear to be known.

In an earlier section of this paper, we have mentioned some possible choices for $\mathbf{X}$. Especially we have emphasized the interpretation of $\mathbf{X}$ as a design matrix. Also numerical variables could be collected in the columns $\mathbf{X}$. Different choices for $\mathbf{X}$, combined with certain structures of $\mathbf{\Sigma}$, again yield other model interpretations and applications. For instance, if $\mathbf{X}$ consists of only one column containing the scores on a categorical variable, and $\mathbf{\Sigma}$ is unconstrained, we have a model similar to *discriminant* analysis. If $\mathbf{X}$ contains a set of predictor variables and we are studying a structural relations model fitted on 'meaningful' residual covariances in $\mathbf{\Sigma}$, we have a version of *covariance* analysis. In that case, $\mathbf{\Sigma}$ is interpreted as the covariance matrix of the variables in $\mathbf{Y}$, adjusted for the effects of the x-variables. And so on.

## 8. Example 1 : STIMEZO data

In this section we will illustrate our fixed factor model with a real data example. The data for this illustration come from a survey, held in 1974, among 575 respondents (see Veenhoven and Hentenaar, 1975). We will call them STIMEZO data. In this survey, people were asked to give their opinion with respect to statements about several controversial issues, such as abortion, capital punishment, euthanasia, etc. Also some background information about the respondents was recorded. We analyzed data obtained from 535 respondents, that contained no missing values. The variables we have chosen for our analysis, are described in Table 1.

We will analyze the variables concerning capital punishment (CP) and abortion (AB) with the fixed factor model. Previous knowledge about the data made us decide to take for $\mathbf{X}$ a design matrix constructed from the background variables REL, POL and EDU. The categories of these variables were coded with dummy variables, i.e. the columns of $\mathbf{X}$ contain ones and zeros. Also several interaction variables were calculated and added as dummy variables to $\mathbf{X}$.

Given the set of CP and AB variables, we will have to determine which model is appropriate to find factors from the columns of $\mathbf{X}$, giving a satisfactory explanation of the interrelations among the CP and AB variables. For instance, one can think of models with or without interactions, incorporating less than three background variables, etc.

**Table 1.** Description of the variables in the analysis

- Three statements about CAPITAL PUNISHMENT

| | |
|---|---|
| CP1 | Taking hostages should be punishable by death. |
| CP2 | Murder should be punished by death. |
| CP3 | In times of war killing people is justifiable. |

– Four statements about ABORTION

| | |
|---|---|
| AB1 | It is the woman's right to have an abortion if she wants it. |
| AB2 | Medical prctitioners who perform abortion are not better than murderers. |
| AB3 | People who agree with abortion have little respect for life. |
| AB4 | Abortion is justifiable under no circumstances. |

The previous statements have the response from (1) = agree completely to (5) = disagree completely. The responses of AB1 are reordered, in order to get a positive correlation with AB1, AB2 and AB3.

– Three BACKGROUND variables

| | | | |
|---|---|---|---|
| REL | Religion, with categories | | |
| | PRO | (1) | Protestant |
| | REF | (2) | Reformed |
| | RC | (3) | Roman Catholic |
| | NON | (4) | None |
| POL | Political preference, with categories | | |
| | LEF | (1) | Left |
| | DEN | (2) | Denomination |
| | LIB | (3) | Liberal |
| | RI | (4) | Right |
| | NON | (5) | None |
| EDU | Educational level, with categories | | |
| | A | (1) | LO, VGLO (*) |
| | B | (2) | ULO |
| | C | (3) | VHMO |
| | D | (4) | Professional training or university |

(*) LO, VGLO, ULO and VHMO are abbreviations of typical Dutch schooltypes. In our opinion it is therefore irrelevant to translate them. The four categoriesof EDU range from elementary school to university. They are denoted by A, B, C and D, with D indicating the highest level.

By manipulation of the design matrix, a 'model-hierarchy' was studied. The results are summarized in Table 2, in which a notation is used that is custom in hierarchical loglinear modeling. We have indicated the variables REL, POL and EDU, with [1], [2] and [3] respectively. For instance, a model which incorporates POL and EDU and a first-order interaction between both variables, is denoted by [23], and so on (cf. Fienberg, 1980). For each model in this hierarchy, Table 2 gives which interactions are involved, the -2 log L values and the total number of parameters to be estimated.

As we have already explained in a previous section, we can decide if one model is more appropriate than another in the hierarchy of nested models, by evaluating the difference between the

**Table 2.** -2 log L values and the total number of parameters to be estimated for various factor models as applied to the data of Table 1; REL=[1], POL=[2] and EDU=[3]

| model abbreviation | -2 log L | # parameters |
|---|---|---|
| [1] | 11956.85 | 23 |
| [2] | 11566.32 | 25 |
| [3] | 12343.65 | 23 |
| [1][2] | 11457.04 | 31 |
| [12] | 11418.66 | 53 |
| [1][3] | 11808.98 | 29 |
| [13] | 11667.78 | 47 |
| [2][3] | 11461.90 | 31 |
| [23] | 11363.41 | 53 |
| [1][2][3] | 11348.83 | 37 |
| [12][13][23] | 11175.35 | 99 |
| [123] | 10989.22 | 141 |

corresponding -2 log L values. Under the assumption of normality, this difference is chi-square distributed with degrees of freedom equal to the difference between the number of parameters to be estimated under each model. From Table 2, it appears that all $\chi^2$ values, based on the difference between the last model and the other, less inclusive, models are significant. So, taking for **X** all background variables and adding all interactions, is a meaningful thing to do. Thus, in the sequel, we will consider the most comprehensive model described in Table 2.

In Table 3, the varimax rotated factor matrix of the two factor solution, corresponding to model [123] is presented. This means that for this solution, we have restricted the rank of **X** to be equal to two, i.e. our algorithm has found two optimal linear combinations of the columns of **X**, which will account for the observed interrelations among the CP and AB variables.

**Table 3.** Varimax rotated factor loadings (covariances) for the two factor solution for the variables concerning capital punishment (CP) and abortion (AB)

| variables | F1 | F2 |
|---|---|---|
| CP1 | .066 | -.662 |
| CP2 | .007 | -.644 |
| CP3 | .200 | -.228 |
| AB1 | .776 | .098 |
| AB2 | .648 | -.166 |
| AB3 | .881 | -.122 |
| AB4 | .719 | -.164 |

From Table 3, it becomes clear that the four abortion variables are well explained by the first factor, while the second factor deals with capital punishment. It appears that the solution of model [123] explains 22.9% of the total observed variance. The variable CP3 does not fit so well in the solution.

An explanation for this fact could be that 'killing people in times of war' is conceived by the respondents as a somewhat different issue than punishing people by death for some committed crime.

In the preceding section, it was explained that the linear restrictions which arise from the premultiplication with **X**, cause the sample to be subdivided into as many groups as there are different rows in **X**.This grouping can be attractive for interpretation. In this case, combining the three background variables, yield a 4 x 5 x 4 three-dimensional array. Thus, the individuals could be subdivided into eighty different groups. However, not all combinations really existed in the data. After the removing of empty cells, 63 combinations of the categories of REL, POL and EDU were left for interpretation. Speaking in terms of the analysis of variance, we have main effects, as well as first and second order interactions.

The factor score estimates for individuals in the same group are equal. One way to interpret these scores is to make plots with the 63 groups. A plot with all 63 points however, would be rather complex. For instance, one can find four different points for each educational level together with the combination Reformed/Right and this will probably make it obscure to see which effects really exist. Therefore, to simplify the interpretation, we have plotted the scores corresponding to the *centroids* of the categories of each background variable. For instance, the combinations just mentioned are then reduced to one point, Reformed/Right, in the plot where all categories of the variable EDU have been taken together. In fact, by doing so, the interaction effect of that variable is reduced. For each pair of background variables, the centroids over the remaining third variable are plotted in Figure 1, 2 and 3.

As we have already described, one can identify two factors: the first factor deals with abortion, the second with capital punishment. To interpret the first factor, we consider Figure 1.
Figure 1, the plot of the centroids over educational categories, shows, on the first dimension, points which projections can be ordered from 'very religious' (Reformed, Protestant) and political 'right', through religious and 'no political preference', towards non-religious and political 'left' or 'liberal'. This is also illustrated in Table 4, in which the mean scores on the abortion variables and the scores on the first factor are given for each group defined by a combination of REL and POL. In this table a high score on the abortion variables indicates a pro-abortion opinion.

**Table 4.** Mean scores on the abortion variables, scores on the first factor and frequencies

| | RI | DEN | NON | LEF | LIB |
|---|---|---|---|---|---|
| REF | 1.57<br>-2.64<br>11 | 3.08<br>-.87<br>33 | 3.11<br>-.86<br>7 | 3.25<br>-.65<br>4 | 3.25<br>-.76<br>5 |
| PRO | 1.75<br>-2.55<br>5 | 2.77<br>-1.25<br>28 | 3.46<br>-.25<br>24 | 3.73<br>.11<br>31 | 3.90<br>.31<br>12 |
| RC | 2.75<br>-1.39<br>1 | 2.76<br>-1.26<br>60 | 3.64<br>-.08<br>46 | 3.95<br>.33<br>31 | 4.02<br>.56<br>26 |
| NON | ---<br>---<br>--- | 4.88<br>1.37<br>2 | 4.14<br>.65<br>65 | 4.34<br>.82<br>102 | 4.52<br>1.04<br>42 |

From Table 4, it is evident that the mean scores for all groups defined by these two background variables and the corresponding factor scores on the first factor, can be ordered almost perfectly in a double monotone way. It is also possible to make an almost perfect monotone ordering of these scores over all cells. These facts support the conclusion that on the first factor, at the left side of the plot in Figure 1, one finds individuals who have a negative opinion towards abortion, are politically at the'right' and religious. At the right side of the plot, one finds individuals who have a positive opinion towards abortion, are politically at the 'left' or 'liberal' and non-religious.

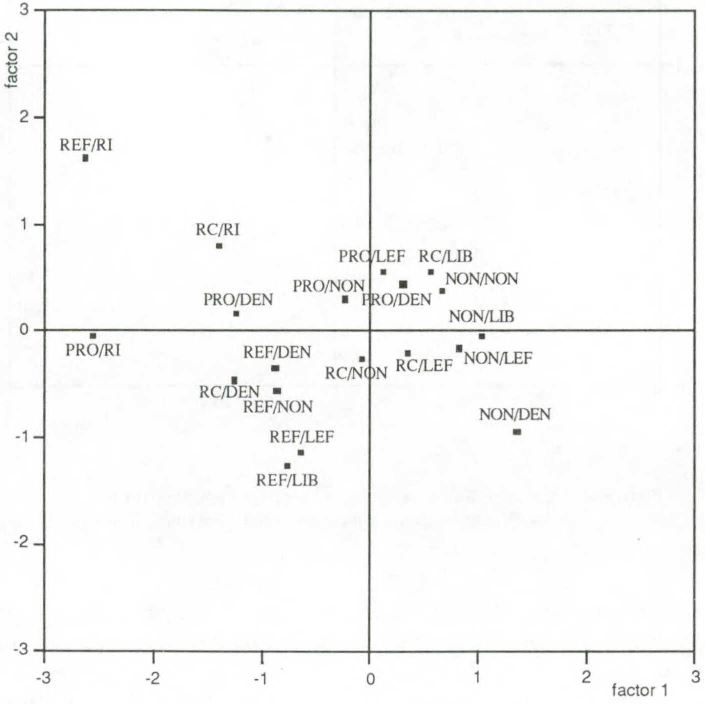


**Figure 1.** Centroids for the categories of the variable educational level (EDU); combinations of religion (REL) and political preference (POL)

To interpret the second factor, we consider Figure 2 and 3. From these figures it becomes clear that the educational level is most important to the second factor. One can see that for both background variables POL and REL, the centroid points are roughly ordered from low educational level at the top, towards high educational level at the bottom of the plot. On the contrary, the corresponding categories of the other variables have no clear ordering on the second dimension. It appears that

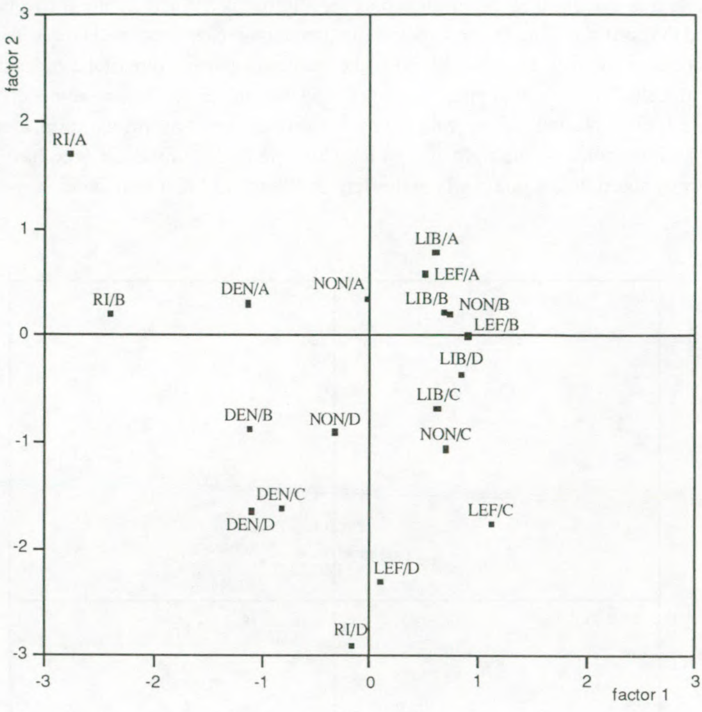


**Figure 2.** Centroids for the categories of the variable religion (REL); combinations of political preference (POL) and educational level (EDU)

people with a higher educational level are much more against capital punishment than those with a lower educational level. It also appears that the centroid points of Figure 2 and 3 can only be ordered over all cells according to the mean score on the CP variables. Such ordering will be less perfect than the one on the abortion factor. Certainly, a double monotone ordering cannot be accomplished.

The horizontal spread of the points in Figure 2 and 3 can be explained by an interaction with REL and POL. Roughly, one finds again individuals who are religious and political 'right' at the left side of these plots, versus individuals who are non-religious and political 'left' or 'liberal' at the right side.

Concluding one can say that on the second dimension, that is, on the second factor, at the top of the plot one finds individuals with a low educational level who favour capital punishment, while at the bottom of the plot, individuals are located with a high educational level who have a strong negative opinion towards capital punishment.

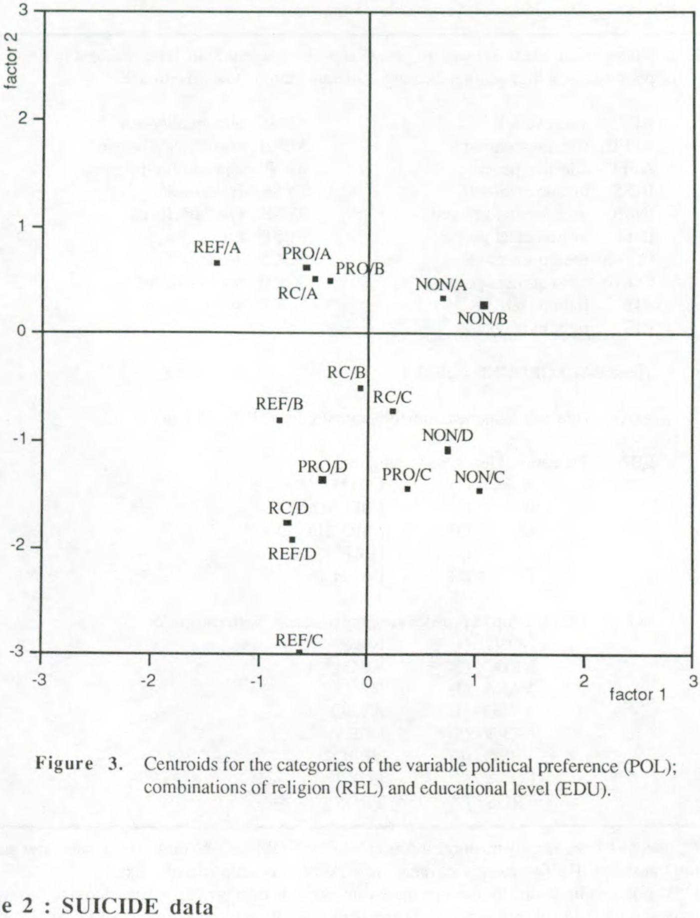


**Figure 3.** Centroids for the categories of the variable political preference (POL); combinations of religion (REL) and educational level (EDU).

## 9. Example 2 : SUICIDE data

For a second example serving to illustrate our fixed factor model, data have been analyzed collected with a 63-item suicide-attitude questionnaire. This questionnaire has been constructed and used by Diekstra and Kerkhof in a large-scale study on attitudes towards suicide (Diekstra & Kerkhof, 1989). The data we have used are the results of the administration of this questionnaire in 1975 to a sample from the population of Nijmegen, a Dutch, medium size town. We have called them SUICIDE data. In total, the sample consisted of 712 subjects. After removing the respondents with missing data, 545 subjects were left for our analysis.

We have analyzed 19 attitude scales that Diekstra and Kerkhof constructed from the original 63 items. Each scale is recoded to 5 categories. The scales combine so-called 'referents' and 'attitude-components'. The theory underlying these terms comes from the social learning theory of suicidal behavior (cf. Diekstra, 1985). In this theory, it is stated that suicidal behavior is chosen and goal directed. Whereas the suicidal person may consider suicide the only way left to solve a problem, the

**Table 5.** Description of the variables in the analysis

- Nineteen attitude scales with respect to attitudes towards SUICIDE, recoded to five-point scales, a high score indicating a tolerant attitude towards suicide.

| | | | |
|---|---|---|---|
| AFFS | affective-self | ABNS | abnormality-self |
| AFFB | affective-beloved | ABNB | abnormality-beloved |
| AFFP | affective-people | ABNP | abnormality-people |
| INSS | instrumental-self | FYSS | fysical-self |
| INSB | instrumental-beloved | FYSB | fysical-beloved |
| INSP | instrumental-people | FYSP | fysical-people |
| CONS | consequences-self | SOCS | social-self |
| CONP | consequences-people | SOCB | social-beloved |
| RIS | right to-self | SOCP | social-people |
| RIS | right to-people | | |

- Three BACKGROUND variables

AGE Age as a numerical variable, ranging from 16 to 71 years

EDU Educational level, with categories

| | | |
|---|---|---|
| A | (1) | LO (*) |
| B | (2) | LBO, MAVO |
| C | (3) | MBO, HAVO |
| D | (4) | HBO |
| E | (5) | University |

BO Membership of broadcasting organization, with categories

| | | |
|---|---|---|
| NON | (1) | None |
| KRO | (2) | KRO (**) |
| VARA | (3) | VARA |
| AVRO | (4) | AVRO |
| NCRV | (5) | NCRV |
| VPRO | (6) | VPRO |
| EO | (7) | EO |
| TROS | (8) | TROS |

(*) As we have already mentioned concerning the STIMEZO example, these categories are not translated. The five categories range from elementary school to university.
(**) It is also irrelevant to translate these categories. If membership is considered to be an indicator for political preference, one can say that members of the religious KRO, NCRV and EO will favor parties ranging from 'centre' (KRO) to the Right (NCRV, EO). AVRO and TROS can be associated with the Liberal parties and members of the socialistically oriented VARA and modern VPRO will favor parties of the Left.

consequences of this behavior however, may be very harmful for close relatives or other persons in the direct or more wider surroundings. This implies that people may differ in their opinions concerning suicide, depending upon whether the supposed suicidal person would be the respondent him -or herself, or someone else, e.g. their most beloved or people in general. It is this differentiation that is indicated by the term 'referents'. As 'attitude-components' one can distinguish between affective, cognitive and instrumental components. This so-called three components model of attitude is well known in social psychology (see e.g. Schuman and Johnson, 1976).

An example of a scale that is considered to be cognitive, is the question concerning 'the right to commit suidide'. Combining this scale with the 'referents' yields the following three scales: 'Do you think / you / your most beloved / people in general / have the right to commit suicide?', which must be scored 'always, mostly, sometimes, mostly not or never'.

In the sequel we will denote the scales by abbreviations such as 'affective-beloved' (AFFB), 'instrumental-self' (INSS), 'right to-people' (RIP), etc. It should be noted that not all possible combinations of 'referents' and 'attitude-components' are included. According to Diekstra and Kerkhof these missing combinations are neither possible nor meaningful, given the contents of the items. For further details concerning this questionnaire and the underlying theory we refer to Diekstra and Kerkhof (1989).

Apart from the 19 scales, we have also used for this illustration the background variables age (AGE), educational level (EDU) and membership of a broadcasting organization (BO). In the Dutch broadcasting system a number of broadcasting organizations or 'unions' operate at the same time. Each of these unions is having her own identity and people can obtain membership of them. Because of this identy, membership of a certain broadcasting union can be considered an indicator of political preference. In Table 5 a brief description is given of the attitude scales and the background variables.

We have analyzed the 19 scales with our fixed factor model, constructing the factors from a design matrix **X**, containing the variables AGE, EDU and BO. What distinguishes this example from the previous one, is that AGE is collected in **X** as a numerical variable, that is, as one column. Also the second power of this variable is added. In this way we can fit the factor scores with a polynomial of second degree. Analyses using a design matrix containing the third power of the variable AGE, were also carried out, but not reported here. Although, according to the likelihood-ratio tests, the addition of this variable appeared to be a significant contribution, the results of these analyses appeared to be rather unstable. It was observed that the polynomials of third degree, describing the relationship between AGE and the factor scores, were too much determined by outliers, i.e. subjects showing extreme responses. Interpretation of these curves would therefore be very unreliable.

As in the first example, a model-hierarchy was studied. Various two-factor models were applied to the data in order to establish which model would be the most appropriate. According to the different likelihood-ratio tests, it appeared (again) that the most comprehensive model [1234], i.e. the model incorporating all interactions, was most suitable. In Table 6, the various models from this hierarchy are summarized.

In the sequel, we will focus upon the interaction between AGE and the other background variables, and the square relationship between factor scores and AGE, defined by the variable AGE2. In order to stress the importance of the other background variables EDU and BO and their interaction with AGE for factor construction, both factor matrices corresponding to the solutions for model [12] and [1234] are given in Table 7. It should be noted that the factor loadings in this table are not varimax rotated, because it appeared that such rotation did not result in improved interpretation of both models. Instead, we chose to maximize the loadings on the first factor.

Clearly, the factor loadings in Table 7 show that the addition to **X**, of the variables EDU and BO and their interactions with the variable AGE, results in a more convincing model. The overall loadings of model [12] are much lower compared to those of model [1234]. In fact, for model [12] one can hardly speak of the existence of a second factor. The improvements attained by model

**Table 6.** -2 log L values and the total number of parameters to be estimated for some factor models applied to the data of Table 5; AGE=[1], AGE2=[2] (*), EDU=[3] and BO=[4]

| model abbreviation | -2 log L | # parameters |
|---|---|---|
| [12] | 33316.38 | 57 |
| [3] | 33488.73 | 61 |
| [4] | 33262.71 | 67 |
| [1][3] | 33086.78 | 63 |
| [12][3] | 33056.59 | 65 |
| [13] | 33054.80 | 71 |
| [123] | 32965.47 | 81 |
| [1][4] | 33071.00 | 69 |
| [12][4] | 33067.90 | 71 |
| [14] | 33042.71 | 83 |
| [124] | 32998.50 | 99 |
| [3][4] | 33085.57 | 75 |
| [34] | 32975.77 | 113 |
| [1][3][4] | 32902.99 | 77 |
| [12][3][4] | 32881.50 | 79 |
| [13][14][34] | 32725.43 | 137 |
| [123][124][34] | 32592.62 | 161 |
| [134] | 32524.34 | 175 |
| [1234] | 32286.49 | 233 |

(*) AGE2 indicates the second power of the variable AGE

**Table 7.** Factor loadings (covariances) for the two-factor solutions concerning the models [12] and [1234]

| | [12] | | [1234] | |
|---|---|---|---|---|
| variables | F1 | F2 | F1 | F2 |
| AFFS | .347 | -.041 | .446 | -.183 |
| AFFB | .319 | -.017 | .428 | -.143 |
| AFFP | .184 | -.092 | .303 | -.146 |
| ABNS | .334 | .042 | .573 | .234 |
| ABNB | .198 | -.082 | .314 | .018 |
| ABNP | .271 | .046 | .603 | .061 |
| CONS | .258 | .033 | .477 | .326 |
| CONP | .240 | .132 | .470 | .373 |
| RIS | .169 | -.022 | .289 | .009 |
| RIP | .348 | .075 | .626 | .096 |
| INSS | .124 | .068 | .208 | -.044 |
| INSB | .056 | .064 | .259 | .137 |
| INSP | .237 | .106 | .405 | .051 |
| FYSS | .385 | -.025 | .577 | -.212 |
| FYSB | .241 | .020 | .450 | -.225 |
| FYSP | .020 | -.028 | .249 | -.065 |
| SOCS | .391 | -.178 | .364 | -.420 |
| SOCB | .365 | -.051 | .362 | -.375 |
| SOCP | .148 | .052 | .194 | -.053 |

[1234] are also reflected in the total amount of variance accounted for by the two factors, which increased from 4.8% for model [12] to 13.9% for model [1234].

Interpretation of the solution for model [1234] yields a general first factor on which all variables have positive, moderately high loadings. The scales dealing with 'the right to commit suicide for people in general' and the question 'whether people who commit suicide are abnormal' have the highest loadings. The second factor roughly distinguishes between the affective, fysical and social components on the one hand and the abnormality-self and consequences components on the other hand.

An interpretation of this kind corresponds with the findings of Diekstra and Kerkhof, who also reported a first factor dealing with general tolerance towards suicide. Our second factor can be imagined as making a distinction between emotional and rational arguments, which topics are separated into further detail in the remaining factors of the solution of Diekstra and Kerkhof.

In terms of factor scores it means that a high score on the first factor indicates tolerance towards suicide. A high score on the second factor means more emphasis on rational rather then on emotional aspects concerning suicide.

Another way to illustrate the importance of the interaction of the background variables is shown in Figure 4. In this plot the factor scores of the first factor of model [12] are plotted against the variable AGE, together with the centroids for all combinations of the categories of EDU and BO. That is, for each group of respondents, defined by a combination of categories of EDU and BO, the

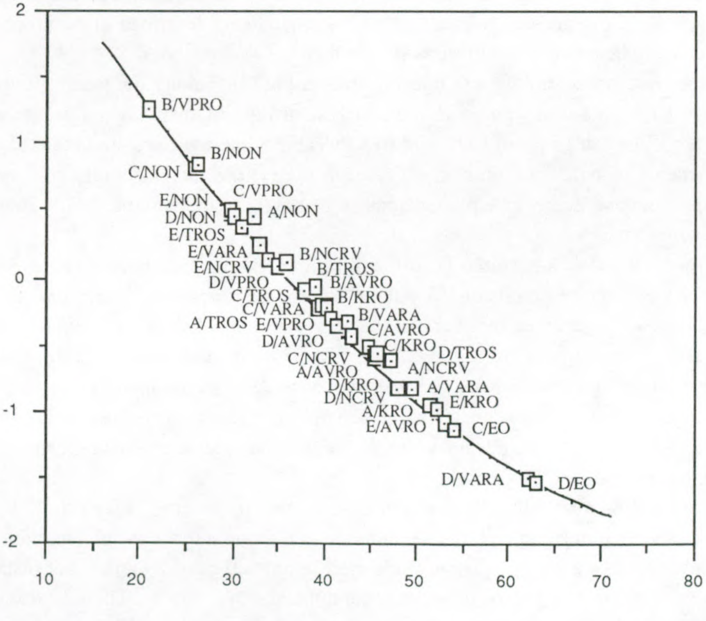


**Figure 4.** Factor scores of the first factor of model [12], plotted against AGE and centroids for all combinations of the categories of EDU and BO

**Table 8.** Contingency table of the variables EDU and BO

| | | BO | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NON | KRO | VARA | AVRO | NCRV | VPRO | EO | TROS | |
| EDU | A | 33 | 28 | 8 | 34 | 6 | - | - | 19 | 128 |
| | B | 79 | 43 | 17 | 46 | 10 | 1 | - | 18 | 214 |
| | C | 49 | 31 | 6 | 16 | 13 | 8 | 2 | 10 | 135 |
| | D | 18 | 11 | 2 | 7 | 1 | 7 | 1 | 3 | 50 |
| | E | 6 | 3 | 3 | 1 | 1 | 3 | - | 1 | 18 |
| | | 185 | 116 | 36 | 104 | 31 | 19 | 3 | 51 | 545 |

mean age and mean factor score serve as the coordinates for a point in Figure 4. The number of respondents in each of these groups is shown in Table 8, which presents the contingency table of EDU and BO.

In Figure 4, the existence of interaction is shown in two ways. First, if the categories of the variable BO are arranged along the curve, given one category of EDU and vice versa, one derives a different ordering for each combination. Second, for each group of respondents defined by a combination of categories of EDU and BO, the variable AGE is differently distributed. Looking at Table 8 however, it appears that we have to be somewhat careful about these arguments, because one observes really small groups, that might be very unstable. As far as interpretation is concerned, we can conclude that in general, over all groups of respondents described in the preceding, older individuals tend to be less tolerant towards suicide then the younger ones.

For further interpretation, we will consider model [1234]. Taking the small group sizes into account, we have focussed upon the eight largest groups of respondents as determined by combinations of the categories of EDU and BO. In Table 8, these groups are indicated by printing their frequencies in bold face numbers. For each factor, the factor scores, which will be on polynomials of second degree, are plotted against the values of the variable AGE. These plots are given in Figures 5 and 6.

From Figures 5 and 6, it becomes clear that different relationships hold between factor scores and AGE for the different groups of repondents. For the first factor (Figure 5), we see that some polynomials show the general trend. For instance, for the groups C/NON and B/KRO, it holds that with increasing age, people tend to have lower factor scores, indicating less tolerance towards suicide. For some groups, such as C/KRO, we observe a slight increase in tolerance with increasing age but only to the age of approximately 35 years. In the remaining groups, e.g. B/AVRO and B/NON, there appears to be an increase in tolerance starting around an age of somewhere between 40 and 50 years.

For the second factor (Figure 6), one can observe that in general, older people tend to have higher factor scores indicating a shift to a more rational attitude towards suicide, with increasing age. For some groups we see a change towards more emotional considerations. For instance, in the B/KRO group there is a switch somewhere around the age of 45 years. There is also one group, A/AVRO, in which the respondents show a more emotional view with increasing age up to approximately 45 years, and then shift to rational arguments.

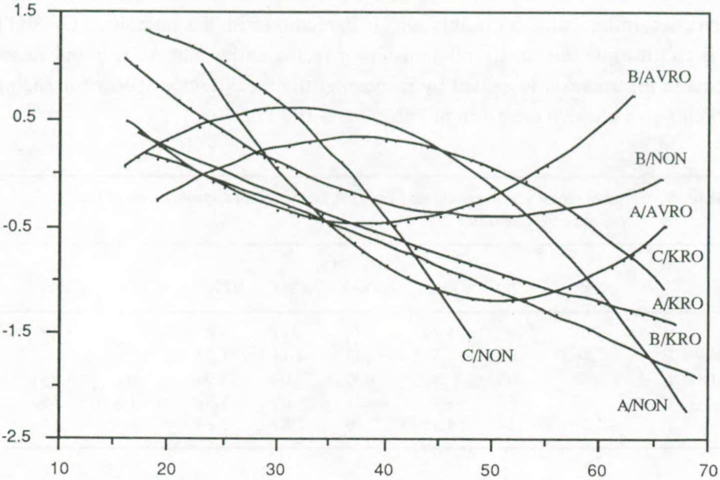


**Figure 5.** Factor scores of the first factor of model [1234], plotted against AGE for the eight largest groups of respondents

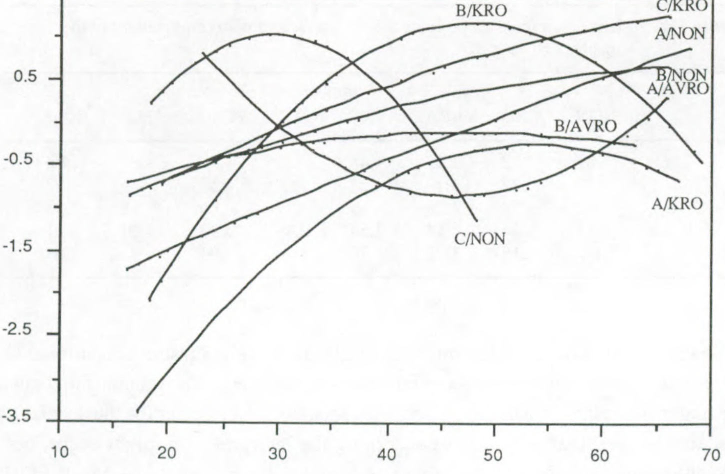


**Figure 6.** Factor scores of the second factor of model [1234], plotted against AGE for the eight largest groups of respondents

Although the curves in Figure 5 and 6 make it possible to do some interesting, speculative interpretation concerning attitudes towards suicide in relation with the variables AGE and BO, they mainly serve to illustrate the square relationship of factor scores and AGE in our factor model. More systematic information is gained by inspecting the mean factor scores for each occuring group, for both factors, which are given in Table 9 and 10.

**Table 9.** Mean factor scores concerning the first factor for all combinations of the variables EDU and BO

| | | BO | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | NON | KRO | VARA | AVRO | NCRV | VPRO | EO | TROS |
| | A | -.01 | -.95 | -.42 | -.76 | -.51 | -.- | -.- | -.29 |
| | B | .46 | -.72 | .38 | -.23 | -1.11 | 1.25 | -.- | -.37 |
| EDU | C | .86 | -.08 | .72 | -.23 | -.94 | 1.94 | -.50 | -.95 |
| | D | .87 | -.50 | -.40 | -.21 | .39 | 1.02 | -.12 | -.36 |
| | E | 2.20 | .34 | 1.45 | .46 | 1.87 | 3.64 | -.- | .44 |

From Table 9 it becomes clear that higher educated people show more tolerance towards suicide. Also people who are not a member of any broadcasting organization at all, or have a membership of the more progressive VARA and VPRO, tend to be more tolerant. People having a lower educational level and people who favour religious and politically 'right winged' broadcasting organizations, NCRV and EO, or the liberally oriented TROS and AVRO appear to be more restrictive.

**Table 10.** Mean factor scores concerning the second factor for all combinations of the variables EDU and BO

| | | BO | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | NON | KRO | VARA | AVRO | NCRV | VPRO | EO | TROS |
| | A | -.87 | -.52 | -2.23 | -.48 | .31 | -.- | -.- | -1.30 |
| | B | -.42 | .24 | 1.17 | -.23 | 1.00 | .52 | -.- | .12 |
| EDU | C | .63 | .58 | 1.08 | .45 | .27 | .04 | .71 | -.38 |
| | D | .87 | -.20 | .14 | 1.54 | 2.49 | .48 | -3.00 | -.21 |
| | E | .71 | -.15 | 1.31 | 1.90 | 1.76 | .01 | -.- | -1.04 |

Table 10 shows that for some broadcasting organizations, with increasing educational level there appears to be a shift from emotional to more rational points of view. The relationship between mean factor score and broadcasting organization does not seem to be very clear for the second factor. We have to realize however, that difficulties concerning the interpretation might occur, because this factor is less important then the first factor and because for several combinations of EDU and BO, we observe very small groups of respondents. So we have to be careful not to make our interpretation too excessive.

## 10. Summary and discussion

Throughout the literature we can observe two basically different factor analysis models. Most popular is the random score model, in which the factors are considered random variables. The structural parameters of the model, loadings and uniquenesses, provide structural infomation of the data. Individual differences, translated in factor scores, can only be studied utilizing ad-hoc procedures (see e.g. McDonald and Burr, 1967). In fact, factor scores can never be estimated in the usual sense, because they are not incorporated as parameters in the random score model. Geometrically, this can be understood by realizing that the set of common and unique factors spans a higher dimensional space then the observed variables. This results in an indeterminacy of the common factors, because the only real way to estimate the factors would be to define a regression from the factors on the observed variables. However, whenever which set of basis vectors of the observed space will not be sufficient to describe the higher dimensional space that contains the factors (cf. Mulaik, 1972).

An alternative model is the fixed score model, in which the factors are fixed quantities, i.e. additional, incidental parameters. These parameters can be used to study individual differences. However, assuming the factors to be fixed variables does not solve the indeterminacy of the factor scores. In fact, we may increase this problem because we are introducing more and more parameters. Indeed, the fixed score model is more complicated from a statistical point of view than the random score model and the ML method fails if it is applicated in the usual way. In some way, we have to impose restrictions on that parameters of the fixed score models.

The method proposed in this paper is based on linear restrictions on the factor scores, formulating the model as a special case of a general reduced rank regression model. The factors are constructed as optimally found linear combinations of the variables contained in a known matrix, which could be interpreted as a design matrix. Which variables are to be collected in the columns of this matrix must be chosen by the user. One could think of categorical (background) variables, coded by dummy variables or just in raw form, observed numerical variables, and so on.

Geometrically, our restrictions imply that the factors are in (a part of) the space spanned by the observed variables. Clearly this is the reason that our factors are *measurable*, although they are not directly observed, to use the terminology of Bentler (1980). In his view, our factors can better be called *unmeasured*, rather than *latent* variables, as is the correct designation of common factors in the random model. Therefore, one might have the opinion that our model is not defing a real factor model. However, because our factors are still not directly observed, we think this remains rather arbitrarily. During the development of the random model there have been models proposed in which the factors were also not real latent variables. For instance, in traditional *image* (factor) analysis (Guttman, 1953), the factors are defined as linear combinations of observed variables. Also the MIMIC model, mentioned earlier in this paper, is an example of a factor model with an unmeasured factor.

In our opinion, it is more important to consider the usefulness of our model, rather than its philosophical aspects. In many behavioral research sitiations where factor analysis is applicated, the interest is in the interrelations among the variables, as well as in the individual differences. Like we have illustrated with the examples, to those cases, we think that the method proposed here can have a contribution.

ontvangen 15-10-1992
geaccepteerd 17- 4 -1992

## References


Anderson, T.W. (1984). The Wald Memorial Lectures. Estimating linear statistical relationships. *Annals of Statistics*, **12**, 1-45.

Anderson, T.W., & Rubin, H. (1956). Statistical inference in factor analysis. In Jerry Neyman (Ed.), *Proceedings of the Third Berkely Symposium on Mathematical Statistics and Probability*, **5**, 111-150. University of California Press.

Bagozzi, R.P., Fornell, C., & Larcker, D.F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*, **16**, 437-454.

Bentler, P.M. (1980). Multivariate analysis with latent variables: causal modeling. *Annual Review of Psychology*, **31**, 419-456.

Bentler, P.M. (1989). *EQS Structural equations program manual*. Los Angeles, CA: BMDP Statistical Software, Inc.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: an expository note. *The American Statistician*, **36**, 153-157.

De Leeuw, J. (1989). *Fitting reduced rank regression models by alternating maximum likelihood.* UCLA Statistics Series #35. Departments of Psychology and Mathematics, University of California at Los Angeles.

De Leeuw, J., Mooijaart, A., & Van der Leeden, R. (1985). *Fixed factor score models with linear restrictions* . Research Report 85-06. Department of Data Theory, University of Leiden.

Diekstra, R.F.W. (1985). Suicide and suicide attempts in the European Economic Community: An analysis of trends with special emphasis upon trends among the young. *Suicide and Lifethreatening Behavior*, **15**, 15-22.

Diekstra, R.F.W., & Kerkhof, A.J.F.M. (1989). Attitudes towards suicide: The development of a suicide-attitude questionnaire (SUIATT). In R.F.W. Diekstra, R. Maris, S. Platt, A. Schmidtke, & G. Sonneck (Eds.), *Suicide and its prevention: The role of attitude and imitation* (pp.91-107). Leiden : E.J. Brill.

Etezadi-Amoli, J., & McDonald, R.P. (1983). A second generation nonlinear factor analysis. *Psychometrika*, **48**, 315-342.

Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*, 2nd ed. Cambridge, MA: MIT Press.

Geary, R.C. (1948). Studies in relations between economic time series. *Journal of the Royal Statistical Socitey*, **10**, 140-158.

Gleser, L.J. (1981). Estimation in a multivariate 'errors in variables' regression model: large sample results. *Annals of Statistics*, **9**, 24-44.

Green, B.F. (1976). On the factor score controversy. *Psychometrika*, **41**, 263-266.

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, **18**, 277-296.

Guttman, L. (1955). The determining of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, **8**, 65-81.

Harman, H.H. (1960). *Modern Factor Analysis* . Chicago: University of Chicago Press.

Harman, H.H. (1967). *Modern Factor Analysis 2nd ed.*. Chicago: University of Chicago Press.

Harris, C.W. (1967). On factors and factor scores. *Psychometrika*, **32**, 363-379.

Hauser, R.M., & Goldberger, A.S. (1971). The treatment of unobservable variables in path analysis. In H.L. Costner (Ed.), *Sociological Methodology* (pp. 81-177). San Francisco: Jossey-Bass.

Heerman, E.F. (1963). Univocal or orthogonal estimators of orthogonal factors. *Psychometrika*, **28**, 161-172.

Intriligator, M.D. (1978). *Econometric Models, Techniques, and Applications*. Amsterdam: North-Holland Publishing Company.

Israëls, A. (1987). *Eigenvalue Techniques for Qualitative Data.* Leiden: DSWO Press.

Izenman, A.J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, **5**, 248-264.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443-482.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, **34**, 183-202.

Jöreskog, K.G., & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, **70**, 631-639.

Jöreskog, K.G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods*. Mooresville, IN: Scientific Software.

Kelly, G. (1984). The influence function in the errors in variables problem. *The Annals of Statistics*, **12**, 87-100.

Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, **60**, 64-82.

Lawley, D.N. (1941). Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh*, **61**, 176-185.

Malinvaud, E. (1964). *Statistical Methods of Econometrics*. Amsterdam: North Holland.

McDonald, R.P. (1974). The measurement of factor indeterminacy. *Psychometrika*, **39**, 203-222.

McDonald, R.P. (1979). The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology*, **32**, 212-228.

McDonald, R.P., & Burr, E.J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika*, **32**, 381-401.

McDonald, R.P., & Mulaik, S.A. (1979). Determinacy of common factors: a nontechnical review. *Psychological Bulletin*, **86**, 297-306.

Mulaik, S.A. (1972). *The Foundations of Factor Analysis*. New York: McGraw-Hill.

Mulaik, S.A. (1976). Comments on 'The measurement of factorial indeterminacy'. *Psychometrika*, **41**, 249-262.

Mulaik, S.A., & McDonald, R.P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika*, **43**, 177-192.

Muthén, B.O. (1987). *LISCOMP: Analysis of linear structural equations using a comprehensive measurement model*. Mooresville, IN: Scientific Software.

Nel, H.M. (1977). *On distributions and moments associated with matrix normal distribution*. Technical Report #24, Department of Mathematical Statistics, University of the Orange Free State, Bloemfontein, South Africa.

Rao, C.R. (1955). Estimation and test of significance in factor analysis. *Psychometrika*, **20**, 93-111.

Schönemann, P.H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, **36**, 21-30.

Schönemann, P.H. (1973). Erratum. *Psychometrika*, **38**, 149.

Schönemann, P.H., & Wang, M.M. (1972). Some new results on factor indeterminacy. *Psychometrika*, **37**, 61-91.

Schuman, H., & Johnson, M.P. (1976). Attitudes and behavior. *Annual Review of Sociology*, **2**, 161-207.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, **15**, 201-293.

Thurstone, L.L. (1947). *Multiple Factor Analysis* . Chicago: University of Chicago Press.

Tintner, G. (1946). Multiple regression for systems of equations. *Econometrica*, **14**, 5-36.

Tintner, G. (1952). *Econometrics*. New York: Wiley.

Tso, M.K.-S. (1981). Reduced rank regression and canonical analysis. *Journal of the Royal Statistical Society*, **43**, 183-189.

Van den Wollenberg, A.L. (1977). Redundancy analysis. An alternative for canonical analysis. *Psychometrika*, **42**, 207-219.

Van der Leeden, R. (1990). *Reduced rank regression with structured residuals*. Leiden: DSWO Press.

Veenhoven, R., & Hentenaar, F. (1975). *Nederlanders over abortus* [The Dutch about abortion]. Stimezo-onderzoek 3 [Stimezo-Research #3].

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of variables is large. *Transactions of the American Mathematical Society*, **54**, 426-482.

Whittle, P. (1952). On principal components and least squares methods of factor analysis. *Skandinavisk Aktuarietidskrift*, **35**, 223-239.

Young, G. (1941). Maximum likelihood estimation and factor analysis. *Psychometrika*, **6**, 49-53.