PSYCHOMETRIKA—VOL. 56, NO. 3 433–447
SEPTEMBER 1991

FITTING LONGITUDINAL REDUCED-RANK REGRESSION MODELS BY ALTERNATING LEAST SQUARES

CATRIEN C. J. H. BIJLEVELD

DEPARTMENT OF PSYCHOMETRICS
LEIDEN UNIVERSITY

JAN DE LEEUW

DEPARTMENTS OF PSYCHOLOGY AND MATHEMATICS
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

An alternating least squares method for iteratively fitting the longitudinal reduced-rank regression model is proposed. The method uses ordinary least squares and majorization substeps to estimate the unknown parameters in the system and measurement equations of the model. In an example with cross-sectional data, it is shown how the results conform closely to results from eigenanalysis. Optimal scaling of nominal and ordinal variables is added in a third substep, and illustrated with two examples involving cross-sectional and longitudinal data.

Key words: reduced-rank regression, state space analysis, optimal scaling.

Introduction

The linear model and the reduced-rank regression model are generalized to situations where a dependence exists between observations on different occasions. Thus, the techniques developed are intended for describing dynamic or longitudinal situations, in contrast to purely cross-sectional ones. The general framework also includes spatial multivariate data, in which the input of a given region influences not only the output of the region, but also the output of adjacent regions, or regions in the immediate neighborhood. In general, our models are appropriate for observations ordered in some clearly defined way (time and space are merely the most obvious examples), and when there is reason to suppose that close observations influence each other.

Consider the following empirical situation. At a number of occasions $t = 1, \ldots, T$, we observe two vector variables x_t and y_t. When y_t is influenced by x_t, (i.e., x_t is the cause of y_t), x_t can be thought of as an input variable, and y_t as an output variable. In econometrics, x_t is called exogenous, and y_t endogenous. In psychometrics, and in various other areas of applied statistics, x_t is called the independent variable, and y_t dependent. Thus, we have two sets of variables, and the two sets play a somewhat different, asymmetric, role in our thinking.

In multivariate analysis the occasions are often replications of the same basic structure, and the index t denotes individuals or objects, considered as a sample from some well-defined population. It is assumed that there is no causal connection between variables with different indices. Thus x_1 influences y_1, x_2 influences y_2, and so on, but there is no influence of x_1 on x_2 or on y_2. This is called the independence assumption. Another important aspect of this type of model is stationarity, where the influence of x_1

Financial support by the Institute for Traffic Safety Research (SWOV) in Leidschendam, The Netherlands is gratefully acknowledged.

Requests for reprints should be sent to C. Bijleveld, Department of Psychometrics, Faculty of Social Sciences, P. O. Box 9555, 2300 RB Leiden, THE NETHERLANDS.

0033-3123/91/0900-88004$00.75/0

on y_1 is supposed to be the same as that of x_2 on y_2, and so on. Such models are at the basis of regression analysis, and of linear models generally.

There is also a slightly more complicated class of independent and stationary models, which goes under various names: reduced-rank regression models, growth curve models, MIMIC models, or errors-in-variables models. Here, the influence of x on y is mediated by an unobserved latent vector variable z, with x determining z, and z determining y. In general, the number of latent variables in z is smaller than the number of variables in x or y, and in this sense, z filters the relationships between the two sets of variables. We call the space of the variables in z the latent or state space, and we use p for its dimensionality. For various versions and applications of reduced-rank regression, we refer to Anderson (1951, 1984) and Jöreskog and Goldberger (1975). Alternating least squares algorithms for fitting reduced-rank regression models have been discussed by de Leeuw and Bijleveld (1987).

If the independence assumption is dropped in the reduced-rank regression models, the dynamic generalization we discuss is obtained, which as pointed out below is identical to the state space models studied in mathematical systems theory (Kalman, Falb, & Arbib, 1969). State space models have been discussed recently in the context of covariance structure models by MacCallum and Ashby (1986) and Otter (1986). Because we have social and behavioral science applications in mind, however, a quite different algorithm is developed in this paper that does not rely on the assumption of multinormal errors and allows for optimal scaling of the input and output variables.

State Space Models

To simplify the discussion, we shall use several concepts borrowed from factor analysis. Explicitly, in factor analysis, m variables in the vector $\mathbf{y} = (y_1, \ldots, y_m)'$ are observed that are correlated. It is assumed that there exist p unobserved variables or factors in the vector $\mathbf{z} = (z_1, \ldots, z_p)'$ that "explain" the association between the observed variables, in the sense that the observed variables are independent given the factors. In the reduced-rank regression model, the dependence of the output $\mathbf{y}$ on the k variables in the input vector $\mathbf{x} = (x_1, \ldots, x_k)'$ is decomposed into dependence of the output $\mathbf{y}$ on the latent factor $\mathbf{z}$, and dependence of the latent factor $\mathbf{z}$ on the input $\mathbf{x}$. In the dynamic case, there is the unobserved state variable $\mathbf{z}$ to mediate the influence of the input $\mathbf{x}$ on the time-dependent $\mathbf{y}$. This dependence of the output variables is accommodated by assuming that all influence of the past on the present is mediated by the present state variables. This first and basic assumption renders the model Markovian.

The state space model can be written as follows:

$$\mathbf{z}_t = \mathbf{F}\mathbf{z}_{t-1} + \mathbf{G}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{z}_t + \boldsymbol{\delta}_t, \tag{2}$$

with $\mathbf{F}$ the p by p matrix containing parameters specifying the influence of the past p states $\mathbf{z}_{t-1}$ on the present p states $\mathbf{z}_t$, $\mathbf{G}$ the p by k matrix with parameters specifying the influence of the k input variables $\mathbf{x}$ on the p state variables $\mathbf{z}$, and $\mathbf{H}$ the m by p matrix with parameters specifying the influence of the p state variables $\mathbf{z}$ on the m output variables $\mathbf{y}$. The errors terms $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$ are needed since we do not expect a perfect fit to real data.

In the multivariate normal situation the random variables $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$ each have independent centered multivariate normal distributions, where $\boldsymbol{\varepsilon}_t$ has covariance matrix $\mathbf{Q}$ and $\boldsymbol{\delta}_t$ has covariance matrix $\mathbf{W}$. The maximum likelihood method can be used to estimate the structural parameters of the system (e.g., Hannan & Deistler, 1988; and

Ljung, 1987), which can be viewed as an attempt to approximate the distribution of the series as closely as possible. In contrast, our emphasis is on approximating the actual matrix of observations, and the fitting of the structural parameters in the matrices (**F**, **G**, **H**) combined with optimal scaling of the variables. It is possible, under fairly restrictive assumptions, to combine optimal scaling with maximum likelihood (de Leeuw, 1988, 1989), but the algorithms for (alternating) least squares estimation are simpler, and developed in much more detail.

Fitting (1) and (2) to data implies finding **Z** and (**F**, **G**, **H**) such that

$$\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}', \tag{3}$$

$$\mathbf{Y} = \mathbf{ZH}', \tag{4}$$

with **Z** the T by p matrix of latent states from time 1 up to time T, **BZ** the T by p matrix of latent states from time 0 up to time $T - 1$, **X** the T by k matrix of input variables, and **Y** the T by m matrix of output variables, both from time 1 up to T. Thus, **B** is the T by T shift matrix, also familiar from the ARMA approach to time series analysis. The shift matrix **B** constructs the state at time 0, $\mathbf{z}_0$, as a weighted sum of the states from time 1 to time T, $\{\mathbf{z}_1, \ldots, \mathbf{z}_T\}$. In the following we always choose **B** such that $\mathbf{z}_0 = \mathbf{z}_1$; choices such as $\mathbf{z}_0 = 0$, or $\mathbf{z}_0 = \text{mean}\,(\mathbf{z}_1, \ldots, \mathbf{z}_T)$ are possible too, and have no effect on the algorithm itself (see Bijleveld, 1989, pp. 83–85).

It is a key result in system theory that a smallest value of p for which (3) and (4) are solvable can be found by algebraic means, and that the solution corresponding with this value of p (the so-called minimal realization of the system) can also be computed exactly (Kalman, Falb, & Arbib, 1969, especially chapter 10, pp. 288–308). There are many algorithms that compute the minimum realization, but in social science situations with high error levels, these algorithms will yield spuriously high estimates of the dimensionality (comparable to finding the number of common factors needed for an exact fit). Thus, we are not interested in computing the minimum dimensionality needed for an exact solution, but in computing an approximate solution in a given dimensionality. The considerations here are the same as in ordinary factor and component analysis.

There is a special case of (3) and (4) which occurs quite often. If there is no input, the state space model is written as

$$\mathbf{Z} = \mathbf{BZF}', \tag{5}$$

$$\mathbf{Y} = \mathbf{ZH}'. \tag{6}$$

Models without measured input are sometimes called dynamic factor analysis models (Immink, 1986; Molenaar, 1981). Because we do not explicitly model errors, and consequently, do not distinguish common and unique factors, it is more appropriate to call (5) and (6) a dynamic component model. Also, the special case of (3) and (4) with $\mathbf{F} = 0$ gives the (cross-sectional) reduced-rank regression models studied earlier with similar techniques by de Leeuw and Bijleveld (1987).

Defining the Loss Function

The techniques presented in this paper choose the unknowns **Z** and (**F**, **G**, **H**) in such a way that the sum of the squares of the prediction errors is as small as possible. In later sections we shall also consider the case in which the input **X** and the output **Y** are partially unknown (for instance, known only up to monotone transformations). The computational problem we consequently discuss is the minimization of

$$\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \omega^2 \text{SSQ}(\mathbf{Z} - \mathbf{BZF}' - \mathbf{XG}') + \text{SSQ}(\mathbf{Y} - \mathbf{ZH}'), \tag{7}$$

where SSQ(·) stands for the sum of squares, over all its arguments.

The weight ω can be used to adjust for the relative importance of predicting the output. If $\omega = 0$, the first term in (7) becomes irrelevant, and minimizing (7) degenerates to the principal component analysis of the output. The limiting case with $\omega \rightarrow \infty$ is more interesting. To study it properly, observe that the first part of the loss function can always be made equal to zero (even if $\mathbf{F}$ and $\mathbf{G}$ are fixed at known values). We merely need to choose $\mathbf{z}_0$ arbitrarily, and then recursively compute $\mathbf{z}_t = \mathbf{F}\mathbf{z}_{t-1} + \mathbf{G}\mathbf{x}_t$. Thus $\mathbf{z}_1 = \mathbf{F}\mathbf{z}_0 + \mathbf{G}\mathbf{x}_1$, $\mathbf{z}_2 = \mathbf{F}^2\mathbf{z}_0 + \mathbf{FG}\mathbf{x}_1 + \mathbf{G}\mathbf{x}_2$, and so on. If $\mathbf{z}_0$ is fixed to make things simple, $\mathbf{Z}$ is a function of $\mathbf{F}$ and $\mathbf{G}$, which can be written as $\mathbf{Z}(\mathbf{F}, \mathbf{G})$. Define

$$\sigma_\infty(\mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{SSQ}(\mathbf{Y} - \mathbf{Z}(\mathbf{F}, \mathbf{G})\mathbf{H}'), \tag{8}$$

and $\sigma_\infty(*, *, *)$ as the minimum of (8). Write $\mathbf{F}_\infty$, $\mathbf{G}_\infty$, $\mathbf{H}_\infty$ for the minimizers. Thus, minimizing (8) amounts to a principal component analysis of the output, with restrictions $\mathbf{z}_t = \mathbf{F}\mathbf{z}_{t-1} + \mathbf{G}\mathbf{x}_t$ on the component scores. Invoking the general theory of penalty functions (Zangwill, 1969, pp. 254–261) immediately gives the following result, where $\sigma(\omega)$ is the minimum of σ_ω over $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{Z}$, and $\mathbf{F}(\omega)$, $\mathbf{G}(\omega)$, $\mathbf{H}(\omega)$, and $\mathbf{Z}(\omega)$ are the minimizers of (7).

Theorem 1. If $\omega \rightarrow \infty$, then $\sigma(\omega) \rightarrow \sigma_\infty(*, *, *)$, $\mathbf{F}(\omega) \rightarrow \mathbf{F}_\infty$, $\mathbf{G}(\omega) \rightarrow \mathbf{G}_\infty$, $\mathbf{H}(\omega) \rightarrow \mathbf{H}_\infty$, and $\mathbf{Z}(\omega) \rightarrow \mathbf{Z}(\mathbf{F}_\infty, \mathbf{G}_\infty)$.

If $0 < \omega < \infty$, the situation becomes a bit more complicated, since unconstrained minimization of (7) over $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{Z}$ is not useful, as Theorem 2 indicates. We first discuss an auxiliary result. Defining $\sigma^* = \min \text{SSQ}(\mathbf{Y} - \mathbf{ZH}')$, σ^* can be found from the singular value decomposition of the output $\mathbf{Y}$.

Theorem 2. Inf $\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \sigma^*$, and the infimum over $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{Z}$ is only attained in very special cases.

Proof. It is clear that $\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \geq \sigma^*$. Now, take $\mathbf{F}_0$ and $\mathbf{G}_0$ arbitrary, $\mathbf{Z}_0$ and $\mathbf{H}_0$ from the singular value decomposition of $\mathbf{Y}$, and choose $(\mathbf{G}, \mathbf{H}, \mathbf{Z}) = (\alpha\mathbf{G}_0, \alpha^{-1}\mathbf{H}_0, \alpha\mathbf{Z}_0)$. Then, $\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \alpha^2 \text{SSQ}(\mathbf{Z}_0 - \mathbf{BZ}_0\mathbf{F}_0' - \mathbf{XG}_0') + \sigma^*$, and letting $\alpha \rightarrow 0$ makes $\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \rightarrow \sigma^*$. The minimum is attained if and only if $\mathbf{F}$ and $\mathbf{G}$ are chosen such that $\text{SSQ}(\mathbf{Z}_0 - \mathbf{BZ}_0\mathbf{F}' - \mathbf{XG}') = 0$, which is possible if and only if $\mathbf{Z}_0$ is in the space spanned by the columns of $\mathbf{BZ}_0$ and $\mathbf{X}$. □

Thus, unrestricted minimization of (7) is not a good idea, because iterative procedures will produce a trivial solution with a very large $\mathbf{H}$ proportional to $\mathbf{H}_0$, a very small $\mathbf{Z}$ proportional to $\mathbf{Z}_0$, and an arbitrary, but also very small, value of $\mathbf{G}$. Therefore, we impose the normalization restrictions $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$. This is not merely a choice of identification conditions, it is a significant restriction on the scores. It is chosen because of the similarity to the restrictions used in factor and component analysis, and in other forms of nonlinear multivariate analysis (Gifi, 1990). It also corresponds with an intuitive idea that we should look for independent factors. Observe that in (3) and (4) it causes no loss of generality to require $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$, because orthogonalization of $\mathbf{Z}$ can always be compensated by suitable modification of $(\mathbf{F}, \mathbf{G}, \mathbf{H})$.

Minimization of (7) will be carried out using alternating least squares. Thus, we alternate the solution to two types of problems: first (7) is minimized with respect to $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ for given fixed $\mathbf{Z}$, and then over $\mathbf{Z}$ for fixed current $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, under the

restriction that $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$. The first type of problem is then reconsidered, and so on. The general theory of alternating least squares shows that this process is convergent. It is clear that the subproblem of the first type, solving for $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ for given $\mathbf{Z}$, is a linear problem that is fairly easy to solve. The subproblem of the second type is much more complicated, however.

Algorithm

Ordinary Least Squares

Consider the problem of minimizing (7) over $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$, for given current $\mathbf{Z}$. The solution for $\mathbf{H}$ is straightforward, since from (4), $\hat{\mathbf{H}}' = \mathbf{Z}'\mathbf{Y}$ is the solution for which $\text{SSQ}(\mathbf{Y} - \mathbf{Z}\mathbf{H}')$ is minimal.

Write $\mathbf{R}$ for the partitioned matrix:

$$\mathbf{R} = \begin{bmatrix} \mathbf{F}' \\ \mathbf{G}' \end{bmatrix}.$$

From (5) it can be seen that this matrix may be written as:

$$\mathbf{R} = (\mathbf{B}\mathbf{Z} \| \mathbf{X})^{+}\mathbf{Z},$$

where "$\|$" stands for horizontal concatenation, and "$(\mathbf{B}\mathbf{Z}\|\mathbf{X})^{+}$" stands for the generalized inverse $((\mathbf{B}\mathbf{Z}\|\mathbf{X})'(\mathbf{B}\mathbf{Z}\|\mathbf{X}))^{-1}(\mathbf{B}\mathbf{Z}\|\mathbf{X})'$. Thus, for $\mathbf{F}$ estimated as the transpose of the first p rows of $\mathbf{R}$, and $\mathbf{G}$ estimated as the transpose of the last k rows of $\mathbf{R}$, $\text{SSQ}(\mathbf{Z} - \mathbf{B}\mathbf{Z}\mathbf{F}' - \mathbf{X}\mathbf{G}')$ is minimal.

Majorization

Consider the problem of minimizing (7) over $\mathbf{Z}$, with $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$, and with the parameters $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ (temporarily) regarded as known constants. Write $\mathbf{Z} = \mathbf{Z}_{\text{old}} + \mathbf{\Delta}$, with $\mathbf{Z}_{\text{old}}$ the current best solution, and define $\mathbf{\Delta} = \mathbf{Z} - \mathbf{Z}_{\text{old}}$. Now, $\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H})$ equals

$$\omega^2\text{SSQ}\{(\mathbf{Z}_{\text{old}} - \mathbf{B}\mathbf{Z}_{\text{old}}\mathbf{F}' - \mathbf{X}\mathbf{G}') + (\mathbf{\Delta} - \mathbf{B}\mathbf{\Delta}\mathbf{F}')\} + \text{SSQ}\{(\mathbf{Y} - \mathbf{Z}_{\text{old}}\mathbf{H}') - \mathbf{\Delta}\mathbf{H}'\}.$$

If $\mathbf{P}_1 = \mathbf{Z}_{\text{old}} - \mathbf{B}\mathbf{Z}_{\text{old}}\mathbf{F}' - \mathbf{X}\mathbf{G}'$ and $\mathbf{P}_2 = \mathbf{Y} - \mathbf{Z}_{\text{old}}\mathbf{H}'$ are the two matrices of residuals for the previous solution, then

$$\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \sigma_\omega(\mathbf{Z}_{\text{old}}, \mathbf{F}, \mathbf{G}, \mathbf{H}) - 2\omega^2 \operatorname{tr} \mathbf{\Delta}'(\mathbf{B}'\mathbf{P}_1\mathbf{F} - \mathbf{P}_1) - 2 \operatorname{tr} \mathbf{\Delta}'\mathbf{P}_2\mathbf{H} + \omega^2\text{SSQ}(\mathbf{\Delta} - \mathbf{B}\mathbf{\Delta}\mathbf{F}') + \text{SSQ}(\mathbf{\Delta}\mathbf{H}'). \quad (9)$$

Now suppose a bound can be found of the form

$$\omega^2\text{SSQ}(\mathbf{\Delta} - \mathbf{B}\mathbf{\Delta}\mathbf{F}') + \text{SSQ}(\mathbf{\Delta}\mathbf{H}') \leq \gamma\text{SSQ}(\mathbf{\Delta}), \quad (10)$$

where γ depends on $\mathbf{B}$, $\mathbf{F}$, and $\mathbf{H}$, and define

$$\mathbf{S} = \gamma^{-1}(\omega^2\mathbf{B}'\mathbf{P}_1\mathbf{F} + \mathbf{P}_2\mathbf{H} - \omega^2\mathbf{P}_1). \quad (11)$$

Then

$$\sigma_\omega(\mathbf{Z}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \leq \sigma_\omega(\mathbf{Z}_{\text{old}}, \mathbf{F}, \mathbf{G}, \mathbf{H}) + \gamma\text{SSQ}(\mathbf{\Delta} - \mathbf{S}) - \gamma\text{SSQ}(\mathbf{S}). \quad (12)$$

But $\text{SSQ}(\mathbf{\Delta} - \mathbf{S}) = \text{SSQ}(\mathbf{Z} - (\mathbf{Z}_{\text{old}} + \mathbf{S}))$. An iteration step of this algorithm consists of minimizing $\text{SSQ}(\mathbf{Z} - (\mathbf{Z}_{\text{old}} + \mathbf{S}))$ over $\mathbf{Z}$ satisfying $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$. This is a simple Procrustes problem (Cliff, 1966), whose solution is well-known. If $\mathbf{Z}_{\text{old}} + \mathbf{S} = \mathbf{K}\mathbf{\Lambda}\mathbf{L}'$ is a

singular value decomposition, then $\mathbf{Z}_{\text{new}} = \mathbf{KL}'$ is the solution of the minimization problem. If $\mathbf{Z}_{\text{new}} = \mathbf{Z}_{\text{old}}$, the algorithm can be stopped. After computing $\mathbf{Z}_{\text{new}}$, set $\mathbf{Z}_{\text{old}} = \mathbf{Z}_{\text{new}}$, and repeat the computations. Thus, instead of minimizing the complicated loss function (9) itself, (9) is minimized indirectly through a majorization algorithm, in which the simpler loss function at the right hand side of (12) is minimized, of which it is known that its values are always higher than or equal to those of (9).

Theorem 3. The algorithm $\mathbf{Z}_{\text{new}} = \mathbf{KL}'$, with $(\mathbf{Z}_{\text{old}} + \mathbf{S}) = \mathbf{K\Lambda L}'$ and $\mathbf{S}$ given by (11), converges to a stationary point (i.e., to a point satisfying $\mathbf{Z}_{\text{new}} = \mathbf{Z}_{\text{old}}$).

Proof. The convergence proof of the procedure is based on the chain

$$\sigma_\omega(\mathbf{Z}_{\text{new}}, \mathbf{F}, \mathbf{G}, \mathbf{H}) =$$

$$\min\{\sigma_\omega(\mathbf{Z}_{\text{old}}, \mathbf{F}, \mathbf{G}, \mathbf{H}) + \gamma \text{SSQ}(\mathbf{Z} - (\mathbf{Z}_{\text{old}} + \mathbf{S})) - \gamma \text{SSQ}(\mathbf{S}) | \mathbf{Z}'\mathbf{Z} = \mathbf{I}\} <$$

$$\sigma_\omega(\mathbf{Z}_{\text{old}}, \mathbf{F}, \mathbf{G}, \mathbf{H}) + \gamma \text{SSQ}(\mathbf{Z}_{\text{old}} - (\mathbf{Z}_{\text{old}} + \mathbf{S}) - \gamma \text{SSQ}(\mathbf{S}) = \sigma_\omega(\mathbf{Z}_{\text{old}}, \mathbf{F}, \mathbf{G}, \mathbf{H}). \quad (13)$$

Thus, the transformation $\mathbf{Z}_{\text{old}} \rightarrow \mathbf{Z}_{\text{new}}$ decreases the loss function (strict inequality actually exists in (13) because the process is stopped if $\mathbf{Z}_{\text{old}} = \mathbf{Z}_{\text{new}}$). Because the transformation is generally continuous (excluding the degenerate case of zero singular values) it follows from Zangwill (1969, chap. 4, pp. 89–94) that convergence occurs to at least a stationary point. □

Thus, the method for estimating $\mathbf{F}$, $\mathbf{G}$, $\mathbf{H}$, and $\mathbf{Z}$ goes as follows. Start with initial values for, for instance, $\mathbf{Z}$. Compute optimal values for $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ using ordinary least squares; given these estimates, compute optimal values for $\mathbf{Z}$ using the iterative majorization procedure outlined above; compute the loss. This constitutes one step of the algorithm. Start the second step using the latest estimates of $\mathbf{Z}$ to estimate new optimal estimates of $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, compute new estimates of $\mathbf{Z}$, and so forth. Instead of using the iterative majorization procedure to estimate $\mathbf{Z}$, it might be better (in terms of overall speed of convergence) to alternate a noniterative single $\mathbf{Z}_{\text{old}} \rightarrow \mathbf{Z}_{\text{new}}$ substep with the $(\mathbf{F}, \mathbf{G}, \mathbf{H})$ substep.

Now consider what happens if not $\mathbf{Z}_{\text{new}}$ is used, but rather $\mathbf{Z}_{\text{new}}\mathbf{M}$, with $\mathbf{M}$ an arbitrary rotation matrix. Denoting by stars arguments over which minimization occurs, it follows that $\sigma(\mathbf{Z}_{\text{new}}, *, *, *) = \sigma(\mathbf{Z}_{\text{new}}\mathbf{M}, *, *, *)$. Thus, the decrease of the loss function as a result of the two substeps taken together will be the same, and is independent of $\mathbf{M}$. It follows that an update can be computed (much more cheaply) by setting $\mathbf{Z}_{\text{new}} = \text{GRAM}(\mathbf{Z}_{\text{old}} + \mathbf{S})$, with $\text{GRAM}(\cdot)$ the Gram-Schmidt orthogonalization (see Gifi, 1990).

There is one step in the actual implementation of the algorithm that is still unclear. This is the choice of γ in (10). Write $\lambda_{\max}(\cdot)$ for the largest singular value of a matrix, and define $\mathbf{A}$ as the partitioned matrix of order $T(p + m)$ by Tp:

$$\mathbf{A} = \begin{bmatrix} \omega(\mathbf{I} - \mathbf{B} \otimes \mathbf{F}) \\ \mathbf{I} \otimes \mathbf{H} \end{bmatrix}.$$

Theorem 4. If $\gamma \geq \lambda^2_{\max}(\mathbf{A})$, then (10) is true.

Proof. Define $\boldsymbol{\delta} = \text{vec}(\boldsymbol{\Delta})$. Then the left-hand side of (10) can be written as $\boldsymbol{\delta}'\mathbf{A}'\mathbf{A}\boldsymbol{\delta}$, from which $\boldsymbol{\delta}'\mathbf{A}'\mathbf{A}\boldsymbol{\delta} \leq \boldsymbol{\delta}'\boldsymbol{\delta}\lambda^2_{\max}(\mathbf{A})$.

By using the results on Theorem 4, in combination with the earlier results, a monotonically convergent algorithm is obtained for minimizing (7) over $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, and all $\mathbf{Z}$ such that $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$. This does not guarantee, of course, that convergence is fast enough for practical purposes, and certainly not that the solutions found by the algorithm will be satisfactory. This will have to be studied by extensive numerical studies, and by the analysis of practical examples.

Comparison of Eigenanalysis and ALS Results

Several analyses were performed to show the effect of the choice of ω, and to compare results obtained through the alternative method of eigenanalysis with results obtained with the proposed method. For this purpose we consider data on the fifty states of the USA. These data are cross-sectional for which it can be assumed that there is no dependence between subsequent measurements. Using cross-sectional data is, in a sense, not really what we are interested in, but their use simplifies some of the limiting results derived above, and similar results can be expected in the time-dependent case.

We have used a version of these data taken from Meulman (1986, pp. 48–54), in which there is a total of twelve variables. The first seven variables are to be considered as input. They are, respectively, percentage of blacks (BLACK), percentage of hispanics (HISPA), ratio of urban to rural (URBAN), per capita income in dollars (INCOM), life expectancy in years (LIFE), homicide rate (HOMIC), and unemployment rate (UNEMP). The last five variables are output variables, having to do with intellectual and educational achievement in the fifty states. They are: percentage high school graduates (HIGHS), percentage public school enrollment (PUBLI), pupil to teacher ratio (PUPIL), illiteracy rate (ILLIT), and failure rate on the selective service mental ability test (FAILU).

Observe that for cross-sectional data, $\mathbf{F} = 0$. Substituting in (7), and minimizing over $\mathbf{G}$ and $\mathbf{H}$, shows that $\sigma_\omega(\mathbf{Z}, 0, *, *) = \omega^2 \text{ tr } \mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z} + \text{tr } \mathbf{Y}'(\mathbf{I} - \mathbf{Z}\mathbf{Z}')\mathbf{Y}$, implying that the minimization problem is equivalent to the eigenvalue problem of maximizing the quadratic form $\text{tr } \mathbf{Z}'\{\mathbf{Y}\mathbf{Y}' - \omega^2(\mathbf{I} - \mathbf{X}\mathbf{X}')\}\mathbf{Z}$ over $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$, where $\mathbf{X} = \text{GRAM}(\mathbf{X})$. If ω is large, then the penalty term forces the $\mathbf{Z}$ corresponding with the largest eigenvalues to be in the column space of $\mathbf{X}$. Thus, the seven largest eigenvalues converge to those of $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$, and the last five eigenvalues converge to the largest eigenvalues of $\mathbf{X}_-'\mathbf{Y}\mathbf{Y}'\mathbf{X}_-$, with $\mathbf{X}_-$ a basis for the orthogonal complement of the column space of $\mathbf{X}$ (de Leeuw & Bijleveld, 1987). We computed the solution for ω equal to 0, 1, and 10. Table 1 gives the ordered eigenvalues, when ω is equal to 0 , 1, and 10, with ω^2 subtracted for the seven input variables. Convergence in fact was reached fairly rapidly.

Next, the correlations were computed between the eigenvectors corresponding to the five largest eigenvalues, and the input and output variables $\mathbf{X}$ and $\mathbf{Y}$. For $\omega = 0$, these eigenvectors are the principal components of the output; if ω increases, they become related more and more to the input, and for large ω they are in the space of the input variables.

The same data were analyzed with the alternating least squares algorithm described above, for which we had written a program named DYNAMALS (linear DYNAMic systems analysis by Alternating Least Squares). Again, the analyses were performed for $\omega = 0$, 1 and 10. The normalized fit of the respective DYNAMALS-solutions was: 1, .856 and .996. In each case, the results conformed closely to the eigen solutions. Also, the correlations of the input and output variables with the DYNAMALS-estimated state variables $\mathbf{z}_1$ to $\mathbf{z}_5$ were approximately the same as those computed by the eigenanalysis: for the first dimension of the state $\mathbf{z}_1$ no differences were found between the correlations from the eigen- and DYNAMALS-solutions, from $\mathbf{z}_2$

TABLE 1

Eigenvalues for Various Values of ω

	ω		
eigenvalue	0	1	10
1	2.674	2.543	2.245
2	1.331	.685	.267
3	.526	.223	.166
4	.323	.076	.050
5	.146	.036	.028
6	.000	.000	.000
7	.000	.000	.000
8	.000	.703	1.166
9	.000	.385	.528
10	.000	.182	.269
11	.000	.108	.209
12	.000	.058	.071

towards higher and less important dimensions of the state differences appeared, with the largest absolute difference found equaling .005.

To summarize the influence of the weight ω, Figure 1 provides the development of the correlations of the input and output variables with the states for ω equal to 0, 1 and 10. In the figure the dots are the correlations for $\omega = 10$; from these dots lines go to the solutions for $\omega = 1$ and subsequently $\omega = 0$.

As expected, Figure 1 shows that the correlations of the seven input variables with the state increase for increasing ω; the correlations of the five output variables with the state decrease. The correlations of the variables with the second dimension of the state change most; apparently this second dimension is less stable than the first.

Optimal Scaling

The alternating least squares techniques discussed in this paper can be combined easily with optimal scaling of the variables. This is illustrated, for example, in de Leeuw (1988). Instead of two substeps in a main iteration, one for updating **F**, **G**, and **H** for given **Z**, and one for updating **Z** for given **F**, **G**, and **H**, there are now three substeps: in the third substep the scaling of the variables in **X** and **Y** is updated, for given **Z**, **F**, **G**, and **H**.

For the loss function (7) and for given **Z**, **F**, **G**, and **H**, the only part that depends on variable $\mathbf{y}_j$ is of the form $\text{ssq}(\mathbf{y}_j - \tilde{\mathbf{y}}_j)$, where $\tilde{\mathbf{y}}_j = \mathbf{Z}\mathbf{h}_j$. It follows that the update of

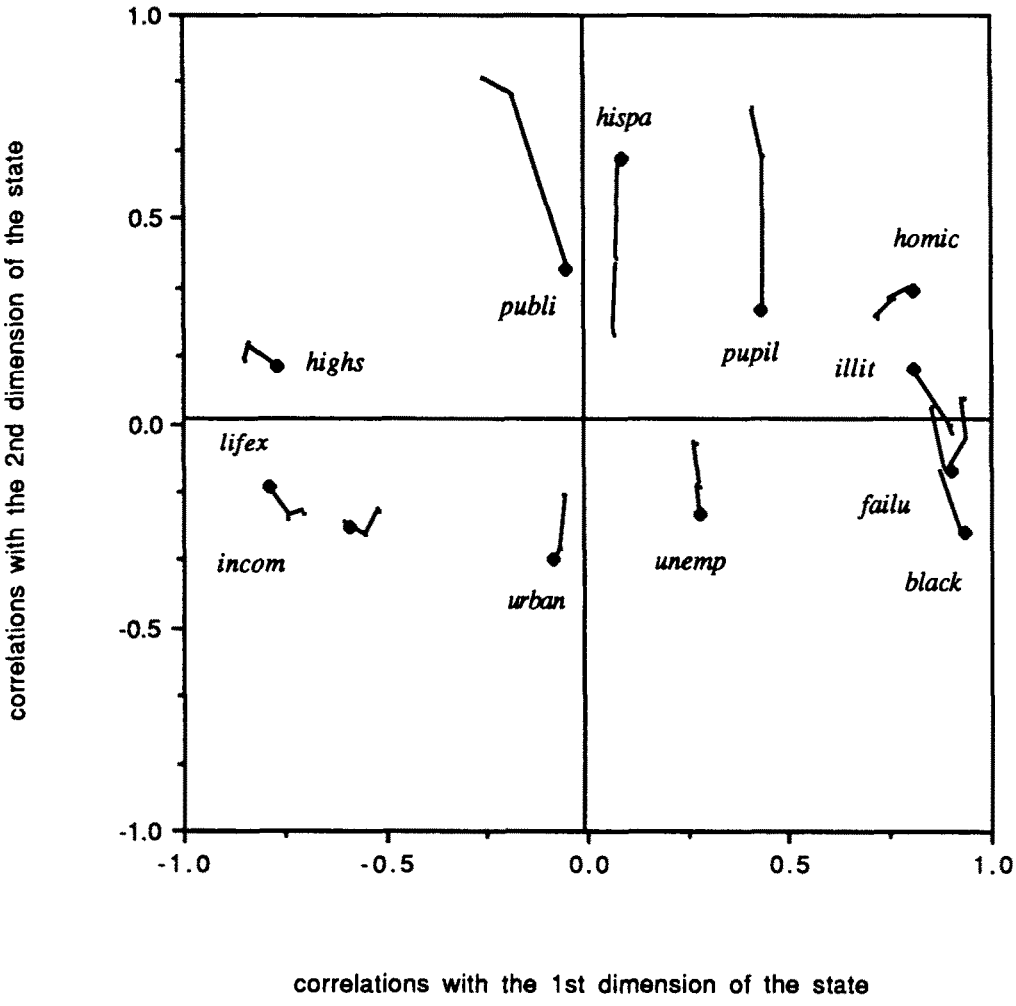

FIGURE 1
Correlations of the variables with the state for $\omega = 0, 1, 10$.

variable $\mathbf{y}_j$ is of the form $\mathbf{y}_j \leftarrow \text{norm}(\text{proj}(\tilde{\mathbf{y}}_j))$, with proj denoting the projection on the cone of admissible transformations and norm denoting the normalization to unit length. There is a norm in the update formula, because the transformed output variables are required to have mean zero and unit length for purposes of identification. This, again, is the usual practice in nonlinear multivariate analysis. We use ssq and norm in lower case, because they are now applied to vectors and not to matrices. The admissible transformations can be the cone of monotone transformations, the subspace of nominal transformations, a subspace of spline transformations, and so on. For details, we refer to the optimal scaling literature mentioned above.

For updating input variable $\mathbf{x}_i$, the situation is a bit more complicated. The relevant part of the loss function can be written as $\text{SSQ}((\mathbf{Z} - \mathbf{BZF}' - \mathbf{X}_i\mathbf{G}_i) - \mathbf{x}_i\mathbf{g}_i')$, with $\mathbf{g}_i$ the i-th column of $\mathbf{G}$. Here, $\mathbf{X}_i\mathbf{G}_i'$ contains the contributions of the input variables except $\mathbf{x}_i$. If $\tilde{\mathbf{x}}_i = (\mathbf{Z} - \mathbf{BZF}' - \mathbf{X}_i\mathbf{G}_i')\mathbf{g}_i/\text{ssq}(\mathbf{g}_i)$, $\text{ssq}(\mathbf{x}_i - \tilde{\mathbf{x}}_i)$ has to be minimized, giving $\mathbf{x}_i \leftarrow \text{norm}(\text{proj}(\tilde{\mathbf{x}}_i))$. Cycling over the variables, changing them one at a time, gives the third alternating least squares substep.

Many variations of this algorithm are of course possible. Cycling over the scaling of **X** and **Y** can be performed various times before the update of **Z** and **F**, and **G** and **H**. The updating of **Z** and **F**, and **G** and **H** can be iterated until convergence before computing a new scaling of the variables. The general experience so far is that small improvements in each substep lead to simple computations and overall convergence at an acceptable rate, but no formal proof for this general statement is available. It is mainly based on practical experience, and on comparing the various combinations in other situations.

Examples

Analysis of Cross-Sectional American States Data

The American states data, that were already analyzed above, were again analyzed with the DYNAMALS program; this time the seven input variables were treated ordinally. This analysis was carried out with five dimensions for the latent state; the weight ω was set at 1. Compared to the numerical solution with $\omega = 1$ that was illustrated earlier in Figure 1, the normalized fit improved to .948 with ordinal treatment of the input variables.

A picture of the American states' scores on the first two state variables, together with the correlations of the input and output variables with these state variables is in Figure 2. The correlations of the variables with the dimensions of the latent state are represented as vectors. The abbreviations used in the picture are NV (Nevada), UT (Utah), WA (Washington), CO (Colorado), CA (California), WV (West Virginia), GA (Georgia), OR (Oregon), WY (Wyoming), NC (North Carolina), AL (Alabama), MS (Missouri), SC (South Carolina), LA (Louisiana), KY (Kentucky), IA (Iowa), AR (Arizona), KS (Kansas), NE (Nebraska), NJ (New Jersey), NY (New York), AK (Arkansas), ND (North Dakota), and RI (Rhode Island). While all American states participated in the analysis, only those American states that had scores in the periphery were plotted in the Figure.

From the Figure, on the first dimension, the vectors of ILLIT, BLACK, FAILU, and HOMIC point in approximately the same direction; in the opposite direction point INCOME, LIFEX, and HIGHS. On the second dimension, PUPIL and PUBLI load positively. The correlations of HISPA and UNEMP with either dimension were low, so they will not be considered in the interpretation. The first dimension may be interpreted as a poverty dimension; states with high scores on this dimension have high percentages of blacks, illiteracy, failure on the Selective Service mental ability test, high homicide rate, small percentage of high school graduates, low life expectancy, and low income. The vectors for HOMIC and INCOM/LIFEX are at almost opposite angles. The second dimension may be interpreted as an education dimension; while PUBLI and PUPIL are the variables that load on this dimension, they are at an angle of approximately 50 degrees. States with low scores on this dimension like North Dakota, Nebraska, and Arkansas on the left-hand side, and Rhode Island, New Jersey, and New York on the right-hand side have low pupil to teacher ratio's and small public school enrollment. States with high scores on this dimension like Nevada and Utah are marked by high public school enrollment. The results may be summarized as follows. Southern states such as Missouri, South Carolina, Louisiana, Alabama, and Georgia, situated in the right-hand part of the picture, are poor states; rich states are Wyoming, Iowa, Washington, Nebraska, Kansas, Oregon, and Colorado. States with high educational achievements are Nevada, Utah, Washington, Colorado, and California; on the opposite end are Rhode Island, New York, North Dakota, and New Jersey.

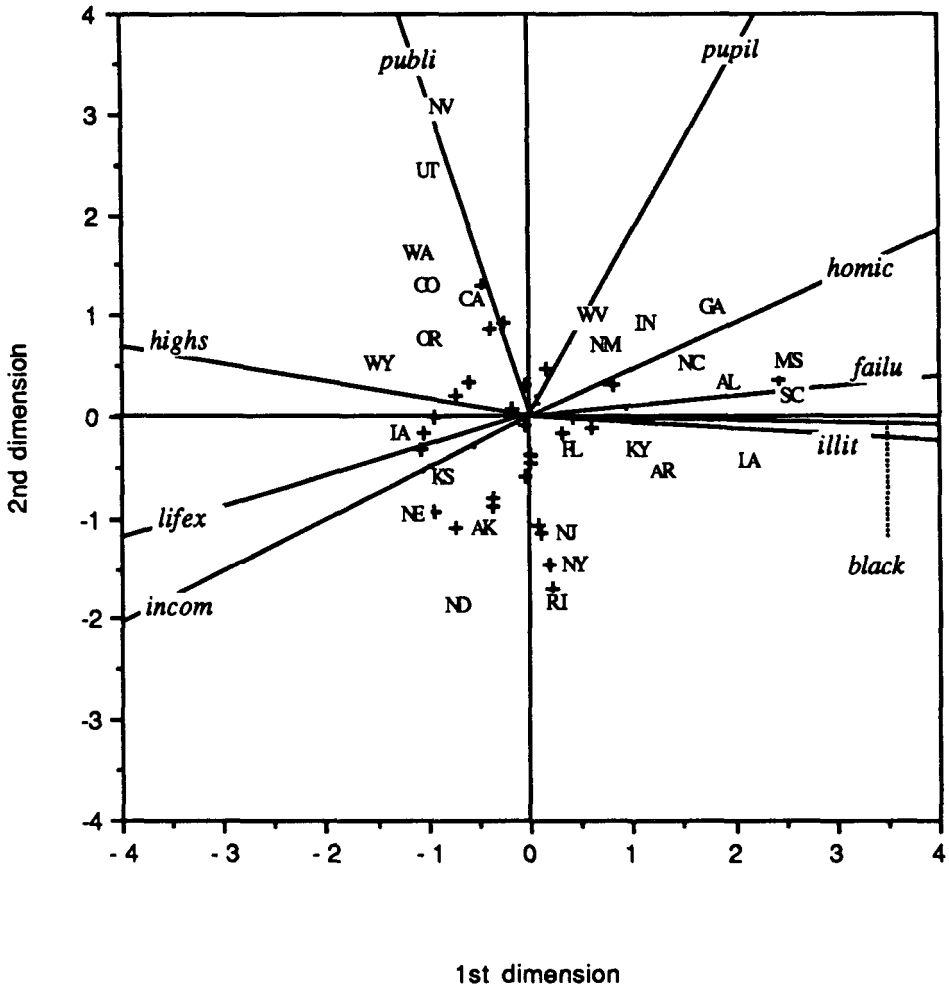

FIGURE 2
Correlations of the variables with the latent states and latent state scores of the American states.

Analysis of Time-Dependent Blood Pressure Data

We present an example of the application of our technique in the analysis of time-dependent data, analyzing the relation between medication and blood pressure from data obtained for a 57 year old white male under medical treatment for hypertension. For 113 days this patient recorded every morning and every evening, under more or less standardized circumstances, his diastolic and systolic blood pressure. Two other series of data were available. The first, called MEDICATION, registers the various medicines the patient took. As the patient was under medical treatment for hypertension, blood pressure can be expected to decrease under the influence of the medicines prescribed. For the first 52 days of the recording period, the patient took 400 milligrams (mgs) a day of metoprololtartraat, abbreviated as "meto. 400 mg"; the patient then switched to 240 mgs a day of sotalolhydrochloride, abbreviated as "sota. 240 mg". After 11 days, a daily diureticum was added, abbreviated as "sota. 240 mg + diureticum", and after another 16 days, the dosage of sotalolhydrochloride was lowered to 160 milligrams a day, abbreviated as "sota. 160 mg + diureticum". The other series, called WEEKDAY, records the day of the week on which blood pressure was measured. As blood pressure can be influenced by stress and other factors, blood pressure might be

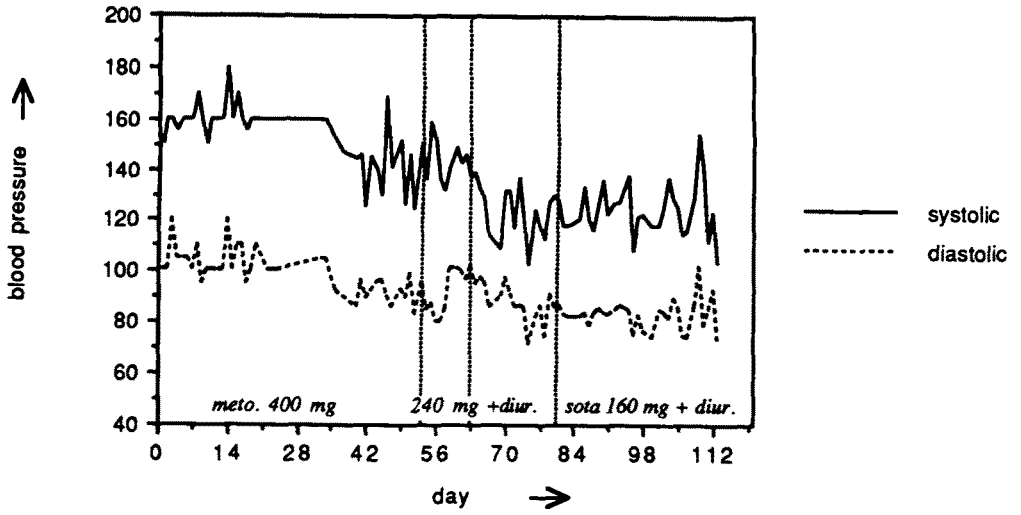

FIGURE 3
Diastolic and systolic blood pressure measurements from day 1 to day 113.

expected to be generally lower in the weekends, and higher during the working-week. The diastolic and systolic blood pressure data are in Figure 3.

Medication and weekday were thus the input variables; as no ordering of the categories is apparent for either of the two, they were treated as nominal variables. The diastolic and systolic blood pressures served as the output variables, and were treated numerically. Several analyses were carried out, for the morning and evening blood pressure measurements separately. The prescribed medication was not expected to take effect immediately, so for both morning and evening measurements, analyses were carried out for various lags for the medication variable. (For instance, for a lag of one day for the medication variable, the relation was analyzed between medication from day 1 to day 112, weekday from day 2 to day 113, and diastolic and systolic blood pressures also from day 2 to day 113.) In all instances, one latent state variable was modeled, and the weight ω set at 1.

The results from the analyses for morning and evening data with the various lags were fairly similar. The fits of the different solutions ranged from .854 to .879, and the interpretations were identical for all solutions. As an example, the solution for the evening measurements will be discussed, with a lag of one day for medication, which happened to be the solution with the best fit. The algorithm had converged in 17 iterations to a normalized fit of .879. The correlations of input variables and blood pressure variables with the one-dimensional state are in Table 2.

Weekday correlates barely with the states, but medication does. Systolic blood pressure correlates stronger with the state than diastolic blood pressure, which is somewhat contrary to expectation, as the diastolic blood pressure is always thought to be more reliable and important of the two; however, both correlations are fairly high. The transition matrix F (which, incidentally, is a 1 by 1 matrix here) equaled .781, pointing to a moderate to strong effect of the prior state on the present state. To evaluate the effects of the various medicines, and of the days of the week, the category quantifications of the categories of medication and weekday are presented in Table 3.

As the blood pressures have negative correlations and medication has a positive correlation with the state, medicines with a negative quantification had a–relatively–negative influence on the blood pressures, that is, they increased or did not decrease blood pressure. Blood pressure was lowered by the medication category with positive

TABLE 2

Correlations of Input and Output Variables with the Latent State

MEDICATION	.861
WEEKDAY	-.051
DIASTOLIC BP	-.914
SYSTOLIC BP	-.943

quantification. Especially during the first period when metoprololtartraat was used, blood pressure was high. The change to sotalolhydrochloride introduced a substantial lowering of blood pressure, but the diureticum had the largest impact on blood pressure. When the change from 240 mg to 160 mg of sotalolhydrochloride a day was made, only a small further improvement of the blood pressures occurred. The quantifications of weekday show that generally there was a slight increase of blood pressure through the working week, from low on Mondays to high over the weekend, but as the correlation of weekday with the states was rather low, no conclusions should be drawn from this. Summarizing, it may be said that a contribution to the lowering of this patient's blood pressure was made by the medication administered. The diureticum substantiated the decrease in blood pressure started by sotalolhydrochloride, after which the patient's blood pressures could be maintained at an acceptable level by a lower dosage of the same medication. The fact that blood pressure was not substantively lowered by the introduction of sotalolhydrochloride itself, but only when this medicine was given in conjunction with a diureticum, is in accordance with the experiences from the medical practice.

TABLE 3

Category Quantifications of MEDICATION and WEEKDAY

meto 400 mg	-.095	Monday	-.145
sota 240 mg.	-.014	Tuesday	-.100
sota. 240 mg+ diureticum	.102	Wednesday	-.024
sota. 160 mg+ diureticum	.105	Thursday	.012
		Friday	.014
		Saturday	.095
		Sunday	.147

Discussion

The technique presented here is a very general one. The cross-sectional versions constitute various forms of errors-in-variables analysis, such as redundancy analysis or reduced-rank regression (de Leeuw & Bijleveld, 1987). The versions without input define various forms of dynamic principal component analysis. The possibility exists of weighting the importance of the input and the dynamics relative to weighting the output by choosing ω, and the option is available to use the various variable transformations allowed by the Gifi system of nonlinear multivariate analysis. This implies a considerable gain in generality compared with existing techniques and programs, but, of course, this comes at a price. The first price is that there is no testable criterion to choose the dimensionality of the state space; secondly, the proposed method does not provide stability information. In the context of dynamic systems analysis there are two competing techniques. The first is the algebraic minimum realization method discussed extensively in Kalman, Falb and Arbib (1969); the second is the maximum likelihood method discussed, for instance, in Ljung (1987). In the algebraic method, the problem of finding the dimensionality of the state space is solved in a satisfactory way, at least from a theoretical point of view, and the problem of stability does not arise. The method will be quite unsuitable for practical social and behavioral science problems, however, which have high levels of errors and uncertain interpretation of state space variables. The maximum likelihood method yields stability information by employing suitable versions of the central limit theorem and law of large numbers, but of course for this method to apply we must assume normally distributed disturbances, which is often unrealistic.

There are a number of interesting compromises between stability and realism. We do not suggest that we have seen them all, because, in particular, the engineering literature is so vast and difficult to translate into statistical terms. The most interesting alternative techniques, from our point of view, are the canonical analysis techniques of Akaike (1976) and Aoki (1987). It will be necessary in the future to compare the DYNAMALS method with those developed on the basis of their results. Also, of course, additional experimentation and theory development is needed to study the stability of DYNAMALS solutions, and various problems related with choice of dimensionality. Such studies will be set up in the same way as the comparable stability and cross-validatory studies for other linear and nonlinear multivariate analysis programs.

There is one problem that is, in a sense, specific to dynamic modeling. Solutions can be stable or unstable, depending on what is basically the size of $\mathbf{F}$. If $\lambda_{\max}(\mathbf{F}) < 1$, the values of the latent states would, with no influence from outside, converge to zero, a condition called stability. If one of the eigenvalues is larger than one, the values of the latent states would be ever-increasing, a condition of explosiveness. This is a different form of stability than the form studied in statistics, for instance by bootstrapping or by computing confidence intervals, but it is important for the interpretation to check if the solutions computed by the technique are stable or not. An unstable solution for $\mathbf{F}$ can be compared, in many respects, to a Heywood case in factor analysis, or a negative variance in variance component analysis. Further study is needed to assess the seriousness of this "improper solution" problem in the DYNAMALS technique.

References

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In R. Mehra & D. Lainiotis (Eds), *System identification: Advances and case studies* (pp. 27–96). New York: Academic Press.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics, 22*, 327–351.

Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics, 12*, 1–45.
Aoki, M. (1987). *State space modeling of time series*. Berlin: Springer-Verlag.
Bijleveld, C. C. J. H. (1989). *Exploratory linear dynamic systems analysis*. Leiden: DSWO Press.
Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33–42.
de Leeuw, J. (1988). Multivariate analysis with linearization of the regressions. *Psychometrika, 53*, 437–454.
de Leeuw, J. (1989). *Fitting reduced-rank regression models by alternating maximum likelihood* (UCLA Statistics Series #35). Los Angeles: UCLA.
de Leeuw, J. & Bijleveld, C. (1987). *Fitting reduced-rank regression models by alternating least squares* (Research Report 87-05). University of Leiden, Department of Data Theory.
Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.
Hannan, E. J., & Deistler, M. (1988). *The statistical theory of linear systems*. New York: Wiley.
Immink, W., (1986). *Parameter estimation in Markov models and dynamic factor analysis*. Unpublished doctoral dissertation, University of Utrecht.
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.
Kalman, R. E., Falb, P. L., & Arbib, M. A. (1969). *Topics in mathematical system theory*. New York: McGraw Hill.
Ljung, L. (1987). *System identification. Theory for the user*. Englewood Cliffs, NJ: Prentice Hall.
MacCallum, R., & Ashby, F. G. (1986). Relationships between linear systems theory and covariance structure modeling. *Journal of Mathematical Psychology, 30*, 1–27.
Meulman, J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
Molenaar, P. C. M. (1981). *Dynamic factor models*. Unpublished doctoral dissertation, University of Utrecht.
Otter, P. W. (1986). Dynamic structural systems under indirect observation: Identifiability and estimation aspects from a system theoretic perspective. *Psychometrika, 51*, 415–428.
Zangwill, W. I. (1969). *Nonlinear programming, a unified approach*. Englewood Cliffs, NJ: Prentice Hall.

Manuscript received 5/21/88
Final version received 8/7/90