UCLA Statistics Series #93 November 1991

Fitting Categorical Time Series

Stef van Buren
Institute of Preventive Health Care, Leiden

Jan de Leeuw
Depts. of Mathematics and Psychology, UCLA

UCLA Classification: Theory and Methods. Social Sciences.

FITTING CATEGORICAL TIME SERIES

STEF VAN BUUREN

INSTITUTE OF PREVENTIVE HEALTH CARE, LEIDEN

JAN DE LEEUW

UNIVERSITY OF CALIFORNIA, LOS ANGELES

This paper outlines a method, called TRANSLAG, for fitting categorical time series by means of optimal scaling. The method paves the way to the application of autoregressive models, predictable components analysis, smoothing filters and other techniques to any mix of nominal, ordinal and numerical data. The main technical problem is to ensure that different lags of the same variable obtain identical data transformations. Two special cases of TRANSLAG are treated in detail: the autoregressive model and predictable components analysis. Other useful models are briefly indicated.

Keywords: categorical time series, least squares, majorization, predictable components analysis, autoregressive models, Box-Tiao transform, optimal scaling, smoothing, intervention analysis, spatial analysis

1. Introduction

A categorical time series consists of a succession of observations that sort each time point into a finite number of categories. Examples of categorical time series in psychology are: sequential behavioral observations in a laboratory setting, presence and absence of assumed therapeutic factors, measures of sleeping stadia, moods and emotions, histories of life events, diary data and so on. Applications and examples of (categorical) time series in the social sciences can be found in Anderson (1963), Jones, Crowell and Kapuniai (1969), Glass, Wilson and Gottman (1975), Hibbs (1977), Kratochwill (1978), Revenstorf, Hahlweg and Schindler (1978), Landis and Koch (1979), Cook and Campbell (1979), MacCleary and Hay (1980), Gottman (1981), Bohrer and Porges (1982), Visser (1982), Kazdin (1982), Gorsuch (1983), Gregson (1983), Kroonenberg (1983), Levenson and Gottman (1983), Barlow and Hersen (1984), Molenaar (1985, 1987), Immink (1986), Brunsdon and Skinner (1987), Larsen (1987), Bijleveld (1989), Oud, van der Bercken and Essers (1990), Von Eye (1990), Bijleveld and De Leeuw (1991) and Skinner (1991).

Good introductions into classical techniques for analyzing time series are Shumway (1988) and Chatfield (1989). The vast majority of these techniques is based on linear models that apply to numerical data only. This paper studies an extension of some of these models to categorical series. More specifically, we deal with autoregressive models and with predictable components analysis. These models apply to series with at least 50 time points, assume equally spaced time intervals and only use sample information. The method, called TRANSLAG is based on the OVERALS framework for polyset canonical correlation analysis of categorical data (cf. Van der Burg, De Leeuw and Verdegaal, 1988; Gifi, 1990). The insertion of lagged series makes it possible to formulate many time series models as special cases of OVERALS. The main technical problem we solve is the minimization of the loss under the additional restriction that different lags of the same series are subject to the same data transformation.

Section 2 introduces some notation, the TRANSLAG optimization problem and its solution. Section 3 and 4 discuss two special cases in detail: Section 3 describes the autoregressive model and Section 4 deals with predictable components analysis. Section 5 suggests additional ways to use TRANSLAG and indicates some areas for future research.

2. Method

Suppose that m categorical series on n time points are coded into indicator matrices G_j $(j = 1, \ldots, m)$. Let k_j be the number of categories of series j, let $D_j = G_j{}'G_j$ be a diagonal matrix of marginal frequencies and let $s = \sum_j k_j$. Categories are quantified by postmultiplying G_j by an initially unknown vector of weights y_j, so $x_j = G_j y_j$ produces a quantified series x_j. For ordinal and interval measurement levels, it is common practice to restrict the sequence of y_j values to be weakly monotonely increasing or to increment with fixed steps. See Gifi (1990) for details. A concise form to write the $n \times m$ matrix $X = [x_1, \ldots, x_m]$ in terms of G_j and y_j can be obtained by defining $G = [G_1, \ldots, G_m]$ and Y as a $s \times m$ matrix containing $y_1, \ldots, y_m$ as its diagonal blocks. It follows that $X = GY$. Subscript $t = 1, \ldots, n$ is used to index time and subscript $j = 1, \ldots, m$ is used to index series. In this way, x_t represents an m-vector of observations at time t and x_j denotes the entire series j of length n. Both x_t and x_j are column vectors.

Time is introduced by means of the *backshift matrix* B, also used by Bijleveld and de Leeuw (1991). The backshift matrix is functionally equivalent to the backshift operator found in many books on time series and it is defined as the $n \times n$ matrix

$$B = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Premultiplying a series x results in the *lagged series* Bx. An l-lagged series $B_l x$ contains element x_{t-l} in its t-th row . Note how B handles end effects: the first observation of Bx becomes zero and the final observation at time $t = n$ is shifted out. This treatment of end effects is consistent with the standard methods to compute autocovariances, autocorrelations and so on. Multiplying B by itself yields higher order backshift matrices. For example $B_2 = BB$ defines a second order backshift matrix and can be used to generate observations at $t-2$. The zero order backshift B_0 is defined as the $n \times n$ identity matrix and B' is the forward shifting matrix. It is sometimes of interest to obtain the difference $\Delta x_t = x_t - x_{t-1}$. The first difference can be written as $\Delta x = (I-B)x$, and the d-th difference is equal to $\Delta^d x = (I-B)^d x$.

Suppose we define $K \geq 2$ sets of p linear combinations $Z_1, \ldots, Z_K$ as

$$Z_1 = \sum_{l=0}^{L_1} B_l X A_{1l},$$

$$\vdots$$

$$Z_K = \sum_{l=0}^{L_K} B_l X A_{Kl},$$

where $L_1, \ldots, L_K \geq 0$ are known lag numbers and where $Z_1, \ldots, Z_K$ are of order $n \times p$. The techniques discussed in this paper aim to find the minimum of

$$\sigma(Z; X; A_{11}, \ldots, A_{KL_k}) = \sum_{k=1}^{K} \text{SSQ}\,(Z - Z_k) = \sum_{k=1}^{K} \text{SSQ}\Big(Z - \sum_{l=0}^{L_k} B_l X A_{kl}\Big),$$

over Z, X and $A_{11}, \ldots, A_{KL_k}$. We call this the TRANSLAG problem and we abbreviate the loss function as $\sigma(\cdot)$. For particular choices of K and L_k the function reduces to a number of interesting special cases. Additional flexibility can be obtained by systematically restricting A_{kl}. For example, by setting the entire j-th row of A_{kl} to zero we exclude the l-th lag of the j-th series from the k-th set. For the moment, we defer a more detailed discussion on how these choices should be made and concentrate on the minimization of $\sigma(\cdot)$.

In order to prevent the trivial solution $\sigma(\cdot) = 0$ with $Z = 0$ we normalize the solution with $1'Z = 0$ and $Z'Z = I_p$. We also require $1'X = 0$ and $\text{dg}\, X'X = I_m$, where $\text{dg}(A)$ stands for the diagonal matrix of the diagonal elements of A. Postmultiplying Z and Z_k by an arbitrary orthonormal matrix does not change the loss, so $\sigma(\cdot)$ is invariant under orthogonal rotation. In the sequel we assume that Z is oriented towards its principal axes. The first dimension then corresponds to the best possible fit.

We minimize $\sigma(\cdot)$ by an iterative algorithm that alternates over the three parameter sets. Each of the steps lowers $\sigma(\cdot)$ so alternating the steps converges to a minimum. The steps are

a. minimization over Z for fixed Y and A_{kl} by least squares
b. minimization over A_{kl} for fixed Z and Y by least squares
c. minimization over Y for fixed Z and A_{kl} by majorization

The unconstrained minimum of $\sigma(\cdot)$ over Z is found by averaging, i.e.,

$$Z = \frac{1}{K}\sum_{k=1}^{K}\sum_{l=0}^{L_k} B_l X A_{kl}.$$

Solving the two-sided Procrustes problem of finding the best fitting orthonormal Z results in the desired constrained solution. A cheaper alternative (Gifi, 1990, p. 99) is Gram-Schmidt orthogonalization during the iterations, followed by a final principal axes rotation after final convergence.

The minimum over $A_{11}, \ldots, A_{KL_k}$ can be obtained as follows: Let U_k denote a matrix that contains all included series of set k and let U_k^+ be the Moore-Penrose inverse of U_k. The solution that minimizes $\sigma(\cdot)$ over $A_{k1}, \ldots, A_{kL_k}$ is U_k^+Z. This procedure can be executed for each set $k = 1, \ldots, K$ separately.

Minimizing $\sigma(\cdot)$ over $X = GY$ with G known, under side conditions $1'X = 0$ and $\operatorname{dg} X'X = I$ is more complicated. The central problem is to find the proper block-diagonal Y that holds the m vectors y_j of category weights. Vectors y_j should satisfy the appropriate measurement level constraints as well as the normalizations $1'D_jy_j = 0$ and $y_j'D_jy_j = 1$ for all $j = 1, \ldots, m$. If the latter conditions are met then $1'X = 0$ and $\operatorname{dg} X'X = I$ will also be true. Unfortunately, it is not possible to split the problem over the separate series if we have lagged series in a single set. Also, we can not split the problem over the separate sets since the same series may appear in more than one set at the same time. The consequence of this is that alternating least squares cannot be applied, and therefore we consider majorization. See Kiers (1990) for references.

Let $C_l = B_lG$ be a l-th order lagged version of G and let

$$Y = Y^o + (Y - Y^o) = Y^o + \Delta$$

where Y^o is some old solution satisfying all appropriate constraints. Writing $\sigma(Y)$ for the loss as a function of Y only, and substituting $Y = Y^o + \Delta$ into it, the loss can be written as

$$\begin{aligned}\sigma(Y) &= \textstyle\sum_k \mathrm{SSQ}\,(Z - \sum_l C_l(Y^o + \Delta)A_{kl}) \\ &= \textstyle\sum_k \mathrm{SSQ}\Big((Z - \sum_l C_l Y^o A_{kl}) - \sum_l C_l \Delta A_{kl}\Big)\end{aligned}$$

Define

$$P_k = Z - \sum_{l=0}^{L_k} C_l Y^o A_{kl}$$

as the matrix of least squares residuals of the k-th set for Y^o and let $\delta = \mathrm{vec}\ \Delta$. Now

$$\begin{aligned}\sigma(Y) &= \sigma(Y^o) - 2\textstyle\sum_k \mathrm{tr}\ P_k'(\sum_l C_l \Delta A_{kl}) \\ &\qquad + \textstyle\sum_k \mathrm{tr}\ (\sum_l C_l \Delta A_{kl})'(\sum_l C_l \Delta A_{kl}) \\ &= \sigma(Y^o) - 2\ \mathrm{tr}\ \Delta'(\textstyle\sum_k \sum_l C_l' P_k A_{kl}') \\ &\qquad + \delta'\,(\textstyle\sum_k (\sum_l A_{kl}' \otimes C_l)'(\sum_l A_{kl}' \otimes C_l))\,\delta \\ &= \sigma(Y^o) - 2\delta' u + \delta' W \delta\end{aligned}$$

where

$$u = \mathrm{vec}\ \sum_{k=1}^{K}\sum_{l=0}^{L_k} C_l' P_k A_{kl}'$$

and

$$W = \sum_{k=1}^{K}\Big(\sum_{l=0}^{L_k} A_{kl}' \otimes C_l\Big)'\Big(\sum_{l=0}^{L_k} A_{kl}' \otimes C_l\Big).$$

Because $\delta' W \delta \leq \alpha \delta' \delta$ for any symmetric W if $\alpha \geq \lambda_1(W)$ the maximum eigenvalue of W the loss is majorized by

$$\sigma(Y) \leq \sigma(Y^o) - 2\delta' u + \alpha \delta' \delta.$$

If we substitute for δ we find

$$\delta = \mathrm{vec}\ (Y - Y^o) = \mathrm{vec}\ Y - \mathrm{vec}\ Y^o = y - y^o.$$

Let the update vector y^u be given by $y^u = u/\alpha$. The problem of minimizing the quantity $-2\delta' u + \alpha \delta' \delta$ over δ becomes equivalent to minimizing

$$(y - (y^o + y^u))'(y - (y^o + y^u))$$

over y. This problem has the simple solution $y = y^o + y^u$ for unrestricted y, but if y is subject to constraints, which is generally the case, the solution is also easy to find.

Since Y is block-diagonal with $s(m-1)$ elements equal to zero, the optimal y will also contain $s(m-1)$ zeroes. It follows that computing the complete update vector y^u is inefficient because only s values are actually needed to find y. The remaining elements of y^u are redundant and need not be computed. Thus, instead of $\big(y - (y^o + y^u)\big)'\big(y - (y^o + y^u)\big)$ we can minimize

$$\big(y_j - (y_j^o + y_j^u)\big)'\big(y_j - (y_j^o + y_j^u)\big)$$

over y_j for each variable $j = 1, \ldots, m$ separately. The update vector y_j^u is given by

$$y_j^u = \frac{1}{\alpha} G_j' \Big(\sum_{k=1}^{K} \sum_{l=0}^{L_k} B_l' P_k a_{klj}' \Big),$$

where a_{klj}' is the j-th row of A_{kl} written as a column. The redundant elements are ignored by this update. Solving the m subproblems under the appropriate constraints will decrease the total loss $\sigma(\cdot)$ over Y.

Since steps a., b. and c. all decrease the loss, iteration also decreases the loss. We stop the algorithm of the loss difference becomes less than a treshold of 0.0005. In many cases, this yields convergence in about 50–100 iterations.

3. Autoregressive models

The autoregressive model is very popular in time series analysis. The model prescribes that the current score x_t depends on a linear combination of previous observations $x_{t-1}, x_{t-2}, \ldots, x_{t-L}$ plus an error which incorporates everything new in the series at time t that is not explained by the past values. If the predictor lags x_{t-l} are consecutive we obtain the well-known Box-Jenkins AR(L) model (cf. Box and Jenkins, 1976). If only a particular lag serves as a predictor, say x_{t-4} for quarterly series or x_{t-12} for monthly series, we arrive at a seasonal autoregressive model that can be used to portray periodic phenomena.

The autoregressive problem is to predict a time series x_t by a linear combination of one or more lags x_{t-l} where $l = 1, \ldots, L$, i.e.

$$\begin{aligned} x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_L x_{t-L} + e_t \\ &= \sum_{l=1}^{L} \phi_l x_{t-l} + e_t. \end{aligned}$$

TABLE 1
Swedish Harvest Index (1749 - 1850). Source MacCleary and Hay (1980).

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

where we assume that $E\Big((\sum_l \phi_l x_{t-l})' e_t\Big) = 0$. The model can be written in matrix notation as

$$x = \sum_{l=1}^{L} B_l x \phi_l + e.$$

Seasonal models can be specified by restricting some of the ϕ's to zero so that the corresponding lags are excluded from the analysis. For categorical data, the problem is to maximize the multiple correlation between x and the predictor series $B_l x$ over the weights and over the category quantifications y in $x = Gy$. Formulated in TRANSLAG terms, we minimize the least squares loss

$$\sigma(z; x; a_0, a_1, \ldots, a_L) = \mathrm{SSQ}(z, xa_0) + \mathrm{SSQ}(z, \sum_{l=1}^{L} B_l x a_l)$$

over z, $x = Gy$ and $a_0, \ldots, a_L$ under normalizations $1'z = 1'x = 0$ and $z'z = x'x = 1$, and possibly restrictions $a_l = 0$ for some $l > 0$. After convergence, the regression weights ϕ_l can be found by applying standard projection techniques to the optimally scaled data.

The question how many and which lags to include in the analysis can be handled in more or less the same way as the iterative Box-Jenkins strategy based on autocorrelations and partial autocorrelations (ACF's and PACF's). There are two complications however. First, TRANSLAG does not require the included lags to be contiguous as in the Box-Jenkins model. This has consequences for the computation of the PACF, since the series that are partialled out are not necessarily all lower order lags, so the standard way of computing the PACF's must be adapted. The second problem is more serious and has to do with the effect of optimal scaling on the autocorrelation function. Because ordinal and nominal transformations of the series will change the autocorrelations of the series, the ACF's may not be comparable across different models that use different optimal transformations, and so the value of using these ACF's as an identification tool is questionable. Below we demonstrate that the Box-Jenkins identification technique can still be used though, as long as the most influential predictors are preserved.

TABLE 2
Results of four ordinal autoregressive models.

	lag	AR(1)	AR(2)	AR(3)	AR(4)
autocorrelations	1	.5038	.5016	.5022	.5015
	2	.1661	.1509	.1532	.1486
	3	.0050	-.0128	-.0114	-.0176
	4	-.0112	-.0133	-.0148	-.0127
	5	.0192	.0072	.0073	.0070
regression weights	1	.5041	.5691	.5630	.5665
	2		-.1344	-.1077	-.1060
	3			-.0433	-.0735
	4				.0499
σ		.4960	.4851	.4835	.4813
multiple r		.5040	.5150	.5164	.5187
modified Box-Pierce		24.89	20.67	21.91	21.22

As an example we use the Swedish Harvest Index series listed in MacCleary and Hay (1980). The series consists of 102 time points and records the annual Swedish grain harvest between 1749–1850 on a ten point scale. MacCleary and Hay argue that the data are measured on an ordinal scale (see pp. 21 and 124), so standard Box-Jenkins techniques do not apply anymore. The data are listed in Table 1. It is slightly different from the series listed in MacCleary and Hay since we took the integer fraction of some entries that, for some unclear reason, were not whole numbers.

We fitted four models to the series, the first-order through the fourth-order autoregressive model under optimal monotone transformations of the data. The results are summarized in Table 2. The columns of this table correspond to each analysis. The first five autocorrelations of optimally scaled series are very similar across all four analyses. This indicates that all models show up with basically the same data transformation, which, as a matter of fact, gives rise to a large tie-block in the middle categories. The regression weight for the first-order term is largest in all solutions so the first order term is the most important in predicting the series. Also, the loss value and the multiple correlation do not decrease very much after lag 1 has been included. The modified Box-Pierce statistic measures the amount of serial dependency in the residual as expressed by the first 25 autocorrelations of the residual. For numerical series, this statistic approximates a χ^2 distribution with df $= 25 - l$ under the nulmodel of no autodependence. None of the models leaves any substantial autocorrelation in the residuals, so we conclude that a simple AR(1) model is adequate here.

Now suppose that we remove all previous lags from the analysis so that each model holds exactly one predictor lag. Table 2 contains the results of the analyses for lags 1 to 4. The solution for lags 2, 3 and 4 do not fit very well.

TABLE 3
Results of four ordinal exclusive lag models

	lag	lag 1	lag 2	lag 3	lag 4
autocorrelations	1	.5038	.4870	.3012	.2276
	2	.1661	.1877	-.0796	-.0445
	3	.0050	.0376	-.2113	-.0617
	4	-.0112	-.0176	-.1007	-.2574
	5	.0192	.0341	-.0617	-.1529
regression weights	1	.5041			
	2		.1882		
	3			-.2127	
	4				-.2575
σ		.4960	.8121	.7881	.7426
multiple r		.5040	.1880	-.2120	-.2574
modified Box-Pierce		24.89	51.54	31.15	41.07

The autocorrelations for the lag 1 and 2 solutions are more or less similar, but those for lags 3 and 4 are entirely different. This occurs because different criteria are being optimized: the lag 2 analysis maximizes the second-order autocorrelation, the third maximizes the third-order autocorrelation, and so on. Thus, optimal transformations are not only data dependent, as is the standard autocorrelation, but they are also model dependent.

The above analyses suggest that a good way to proceed is to work in a stepwise way. First, we select the most important lags, possibly based on the ACF and PACF of the data treated at an interval level and fit a model on them. Then by inspecting the autocorrelations of the residual less informative lags may be dropped and more informative lags may be included, and so on. Working this way, it is unlikely that abrupt changes will appear in the transformation function since we preserve 'good' predictors.

4. Predictable components

Box and Tiao (1977) proposed a canonical analysis that extracts predictable components from multivariate time series. The first predictable component is a linear combination of the original series that forecasts itself as well as possible. Like PCA, the second component optimizes the same criterion, but under the condition that it is orthogonal to the first. An obvious use of the technique is to identify those components that can serve as smoothed indicators of overall growth in, for example, the stock market. Alternatively, the technique can be used as a dimension reduction device to bring out the major time dependent characteristics of a multivariate data set.

Let X of order $n \times m$ contain m quantitative measurements sampled at n points of time. Suppose that x_t can be modelled by the multiple autoregressive process $x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_L x_{t-L} + e_t$, where the Φ_l are $m \times m$ matrices and where e_t is an m-dimensional white noise process. Box and Tiao (1977) show that this multivariate autoregressive process can be reparametrized as a collection of m uncoupled univariate autoregressive processes on some new series u_t. The transforms works by finding linear combinations $u_j = Xa_j$ for $j = 1, \ldots, m$ that are contemporaneously independent, i.e. $E(u_j' u_{j'}) = 0$ for $j \neq j'$, and that are ordered according to their respective predictive powers. The predictability measure γ_j reflects how much the j-th component can predict itself by a univariate L-th order autoregressive model $u_{t,j} = f_1 u_{t-1,j} + f_2 u_{t-2,j} + \ldots + f_L u_{t-L,j} + \tilde{e}_{t,j} = \hat{u}_{t,j} + \tilde{e}_{t,j}$, where f_l are scalar autoregressive weights. Let $\hat{\sigma}^2 = E(\hat{u}_{t,j}^2)$ and let $\sigma_j^2 = E(u_{t,j}^2)$ then the predictability for u_j is equal to $\gamma_j = \hat{\sigma}_j^2 / \sigma_j^2$, i.e. the proportion of variance of u_j explained by the systematic part $\hat{u}_j$. For the first predictable component, the goal is to find the weight vector a_1 such that the linear combination $u_1 = Xa_1$ has maximum predictability γ_1. Next, a second predictable component, orthogonal to the first, can be identified, and so on. To see how this problem can be solved using the TRANSLAG model we write all components simulataneously as

$$U = \sum_{l=1}^{L} B_l U F_l + E.$$

Since $U = XA$ we can rephrase this in terms of the observed data as

$$XA = \sum_{l=1}^{L} B_l XAF_l + E = \sum_{l=1}^{L} B_l XA_l + E,$$

where $A_l = AF_l$. It is now easy to see that the problem of determining maximum predictability is equivalent to finding the largest canonical correlations between a set of the observed series X and a set of lagged series $[BX, B_2X, \ldots, B_LX]$. The relationship with CCA has been studied before by Parzen and Newton (1980) and Velu, Reinsel and Wichern (1986). The latter authors found that γ_j is equal to the squared canonical correlation. For categorical data, the problem is now to minimize

$$\sigma(Z; X; A_0, \ldots, A_L) = \mathrm{SSQ}(Z, XA_0) + \mathrm{SSQ}(Z, \sum_{l=1}^{L} B_l XA_l),$$

over $Z, X = GY$ and $A_0, \ldots, A_L$ under constraints $1'Z = 0$, $1'X = 0$, $Z'Z = I$ and diag $X'X = I$. The predictable components are then equal to $U = XA_0$. Since U approaches the orthogonal matrix Z the components are nearly orthogonal.

TABLE 4
Diary data (n = 131)

	E	P	X	I	S	F	A		E	P	X	I	S	F	A		E	P	X	I	S	F	A
1	1	2	1	1	1	3	1	45	2	1	1	1	1	5	1	89	2	2	1	1	2	4	2
2	3	2	1	1	1	1	1	46	2	1	1	1	1	5	1	90	3	2	1	1	3	4	3
3	3	2	1	1	1	2	1	47	2	1	1	1	1	3	1	91	2	2	1	1	1	3	3
4	3	2	1	1	1	3	1	48	2	1	1	1	1	5	1	92	2	2	3	1	1	1	2
5	3	2	1	1	1	2	1	49	3	1	1	1	1	3	2	93	3	2	1	1	2	2	2
6	4	2	2	1	1	1	1	50	2	1	1	2	1	1	2	94	3	2	2	1	2	5	2
7	4	2	1	1	1	5	1	51	3	2	2	2	1	2	1	95	2	2	1	1	1	1	3
8	3	1	2	1	1	5	1	52	3	2	1	2	1	4	1	96	3	2	1	1	3	4	3
9	4	2	1	1	1	1	1	53	3	2	1	2	2	5	2	97	2	2	1	1	3	4	2
10	4	1	1	1	1	2	1	54	2	2	1	2	1	2	3	98	2	2	1	1	1	5	3
11	1	2	2	1	1	2	1	55	2	2	1	2	1	5	1	99	2	2	1	1	1	3	2
12	1	2	1	1	1	4	1	56	2	2	1	1	1	3	1	100	2	2	1	1	3	4	3
13	2	2	1	1	1	1	2	57	2	1	1	1	2	1	2	101	2	2	1	1	3	1	3
14	2	2	1	1	1	2	2	58	2	2	1	1	1	3	1	102	3	2	1	1	3	2	3
15	2	2	2	1	1	4	2	59	2	2	1	1	1	1	1	103	3	2	1	1	3	2	3
16	2	2	1	1	1	2	2	60	2	2	1	1	1	2	3	104	2	2	1	1	3	2	3
17	3	2	1	1	1	2	2	61	2	2	1	1	1	2	1	105	3	2	1	1	3	2	3
18	3	2	1	1	1	2	2	62	1	2	1	1	1	2	2	106	2	2	1	1	3	5	3
19	3	2	1	1	2	3	2	63	1	2	2	1	1	1	2	107	3	2	2	2	3	5	3
20	2	2	2	1	1	1	2	64	2	2	2	1	1	4	3	108	3	2	2	2	1	3	1
21	2	2	1	1	1	3	1	65	2	2	1	1	1	1	1	109	3	2	1	2	1	3	1
22	2	2	1	2	1	1	1	66	3	2	2	1	1	5	1	110	3	2	1	2	1	5	2
23	3	1	1	2	1	2	2	67	1	2	1	1	1	1	1	111	4	2	1	2	3	1	3
24	2	1	1	2	1	2	2	68	2	2	1	1	1	5	1	112	1	2	1	2	1	4	1
25	3	1	1	2	1	2	2	69	2	2	1	1	3	4	3	113	1	2	1	1	1	1	2
26	1	1	1	2	1	4	2	70	1	2	2	1	1	2	3	114	2	2	2	1	1	4	1
27	1	1	1	1	2	1	2	71	1	2	1	1	3	1	2	115	1	2	1	1	1	5	2
28	2	1	3	1	1	1	3	72	2	2	1	1	1	2	2	116	2	2	1	1	2	5	3
29	2	1	1	1	1	1	3	73	2	2	2	1	1	2	2	117	2	1	1	1	2	4	2
30	3	2	2	1	2	5	2	74	2	2	1	1	1	4	3	118	2	2	1	1	3	4	3
31	4	2	1	1	2	2	1	75	4	2	2	1	1	5	3	119	3	2	3	1	3	1	3
32	2	2	1	1	1	2	2	76	2	2	1	1	3	1	2	120	2	2	1	1	1	1	1
33	2	2	1	1	1	2	3	77	2	2	1	1	1	3	2	121	2	1	1	1	2	1	2
34	2	2	1	1	1	2	1	78	3	2	2	2	1	1	3	122	2	1	1	1	2	3	2
35	2	2	1	1	1	3	1	79	2	2	1	2	2	2	2	123	1	2	1	1	1	3	3
36	2	2	2	1	1	3	3	80	3	2	1	2	2	2	2	124	1	1	2	1	1	5	2
37	2	2	2	1	2	2	2	81	3	2	1	2	3	3	3	125	2	1	1	1	3	3	3
38	2	2	1	1	2	2	3	82	3	2	1	2	2	2	3	126	2	1	1	1	1	2	1
39	1	2	1	1	2	2	2	83	3	2	2	2	2	4	2	127	2	2	1	1	1	1	3
40	2	2	1	1	1	2	1	84	3	2	1	1	3	2	3	128	2	2	2	1	2	2	2
41	2	2	1	1	1	1	1	85	3	2	1	1	2	3	3	129	2	2	1	1	3	5	3
42	2	2	2	1	1	1	2	86	4	2	3	1	2	1	3	130	3	2	1	1	2	5	2
43	1	2	1	1	1	1	1	87	2	2	1	1	2	1	2	131	1	1	2	1	2	3	2
44	2	1	3	1	2	5	3	88	2	2	2	1	3	1	2								

As an example, let us consider the diary series in Table 4. The series records a number of psychological and medical factors and some daily activities of a woman in her mid-twenties for 131 consecutive days. The seven series measure: E = Emotional State (1=down, 2=normal, 3=good, 4=active-hysteric), P = Physical State (1=ill, 2=healthy), X = Sexual Activity (1=no, 2=some, 3=much), I = Indisposed (1=no, 2=yes), S = Smoking (1=none or missing, 2=1–10 cig., 3=10+ cig.), F = Food (1=Italian, 2=Dutch, 3=Bread, 4=Snacks, 5=other) and A = Alcohol (1=none or missing, 2=1-3 beer/wine, 3=3+ beer/wine). The intent of the data collection was to examine what factors might influence the occurrence of eruptive fever. Our major interest concerns the problem whether anything can be predicted at all in these data. We assume that the influence of an observation wears out in about five days so we set $L = 5$.

The squared canonical correlations are respectively 0.74,0.58, 0.32, 0.26, 0.26, 0.22, and 0.16. The first two predictable components are plotted in Figure 1. In order to interpret the components it is useful to look at the

FIGURE 1
First and second predictable components of the Diary data

correlations between the predictable components and the series from which they were constructed. There is a total of 6(lags) × 7(series) × 2(dimensions) = 84 correlations to look for. Table 5 reports these correlations. In addition to the correlations, it can be helpful to study the scaling of the individual categories. A sensible way to do this is to plotYA_0, YA_1 and so on.

The first component has a very strong periodic tendency corresponding to a 28 day menstruation cycle. The correlation between the period series and the component is 0.99. A second series that 'loads' on this component is emotional state. The second component depends on a combination of physical state, smoking and alcohol consumption, with physical state being the most important contributor. It turns out that the dips in the second component exactly match the periods of illness, while the peaks correspond to periods with much drinking and smoking. The correlations for the higher lags diminish for nearly all seven series. This not only indicates that first-order time relationships dominate higher order influences in the series themselves, but also leads to the conclusion that there are few, if any, clear dynamic cross-variable relationships between the series.

5. Discussion

This section discusses the performance of the algorithm and suggest additional ways of using TRANSLAG.

Given the complexity of the computations the numerical stability of the algorithm is remarkably stable. Even extremely tight convergence criteria like $1E-7$ do not affect convergence, and so far we have not seen cases in which the algorithm diverges. Like OVERALS, not much is known about the

TABLE 5
Correlations between series and the predictable components 1 and 2.

	Component 1							Component 2						
lag	E	P	X	I	S	F	A	E	P	X	I	S	F	A
0	.39	-.04	-.10	.99	-.07	-.06	-.10	.16	.73	.03	.03	.71	-.12	.45
1	.35	-.05	-.09	.79	-.04	.04	-.01	.13	.48	-.10	.03	.37	-.11	.35
2	.28	-.06	-.08	.57	-.02	.08	-.08	.12	.42	.00	.01	.28	.00	.37
3	.21	-.04	-.09	.35	.00	.13	-.05	.11	.25	-.06	.00	.27	.03	.24
4	.11	-.05	-.10	.14	.02	.08	.02	.10	.17	.01	.03	.2	-.02	.28
5	.11	-.10	-.10	-.07	.02	.00	.09	-.04	.15	-.13	.03	.19	.12	.33

occurrence of local minima in some of the more complicated cases. In our limited experience, we got the impression that local minima do not constitute a serious problem as long as tight convergence criteria are maintained. As the present algorithm is rather slow more research will be needed to improve the computational speed of the method.

We also do not know how stable the results are. It is well-known that in the presence of autocorrelation least squares estimates remain unbiased, but not are minimum variance. Moreover, multicollinearity in higher-order autoregressive models may add to the variability of the estimates. Systematic applications of randomization methods such as the jackknife or bootstrap may shed light on the stability of the parameters.

TRANSLAG only handles complete data matrices. In principle, it is possible to allow for missing data according to the strategies outlined in Gifi, or by using imputation methods. This is also a point of further research.

Besides autoregression and predictable components, there seem to be at least three other applications of TRANSLAG. A first of these is smoothing of categorical series. Suppose that we want to smooth x by the *weighted running mean smoother* $y_t = \sum_{k=-K}^{K} x_{t-k}a_k$, where $a_{-K}, \ldots, a_K$ are given filtering weights that determine the properties of the smoother. Some well-known choices correspond to the running average filter, the Hanning filter, the Spencer 15-point filter and the Gaussian kernel. Recent overviews of such techniques are Goodall (1990) and Hastie and Tibshirani (1990). Using TRANSLAG, it is not difficult to apply these filters to categorical series. If we minimize

$$\sigma(z; x) = \sum_{k=-K}^{K} \mathrm{SSQ}(z - B_k x a_k)$$

over z and $x = Gy$ then

$$z = \frac{1}{K} \sum_{k=-K}^{K} B_k x a_k.$$

is the filtered series. This technique seems especially useful to quantify univariate series for with no additional information available, other than being smooth.

Another possibility is intervention analysis (see Glass, Willson and Gottman, 1975). The goal of intervention analysis is to infer whether a specific event has an effect on the level of the series. As an example, let x_1 denote a target series and let x_2 be the series that codes the presence and absence of the event. What we would like is to do is to use a common t-test conditional on the level of x_2. However, this procedure is questionable since x_1 is likely to be autocorrelated. Therefore, the t-test is only used on the residuals after an autoregressive model has been fitted on x_1. In terms of TRANSLAG we minimize

$$\sigma(z; x_1, x_2; a_0, \ldots, a_L, \beta) = \mathrm{SSQ}(z - x_1 a_0) + \mathrm{SSQ}(z - x_2\beta - \sum_{l=1}^{L} B_l x_1 a_l)$$

over the relevant parameters. The regression weights can be found—and subsequently tested—by projecting x_1 on the predictor space after optimal transforms have been obtained.

Finally, the most spectacular and far-reaching application is the analysis of spatial data. In time series analysis, observations are linked in the direction of time by means of the backshift matrix. Spatial dependency problems are more complicated since observations may influence each other in several directions simultaneously. Examples of spatial dependency occur in agriculture where experiments plots have common borders, in the analysis of social networks and in the analysis of regional data. It is possible to code dependencies among analysis units by means of an adjacency matrix, a more general form of the backshift matrix. Note that the fact that B is a backshift matrix is not used anywhere in Section 2 so the algorithm holds for any real $n \times n$ matrix B. A systematic study of this largely unexplored but vast area could easily fill up a thesis.

REFERENCES

Anderson, T.W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika, 28*. 1–25.

Barlow, D.H. and Hersen, M. (1984). *Single case experimental design: strategies for studying behavior change, 2nd ed.* Pergamon Press, Oxford.

Bijleveld, C.C.J.H. (1989). *Exploratory linear dynamic systems analysis.* DSWO Press, Leiden.

Bijleveld, C.C.J.H. and de Leeuw, J. (1991). Fitting longitudinal reduced-rank regression models by alternating least squares. *Psychometrika*, to appear.

Bohrer, R.E. and Porges, S.W. (1982). The application of time series statistics to psychological research: an introduction. In: G. Keren (Ed.), *Statistical and methodological issues in psychology and social sciences research.* Erlbaum, Hillsdale, N.J.

Box, G.E.P. and Jenkins, G.W. (1976). *Time series analysis, forecasting and control (revised edition)*. Holden-Day, San Francisco.

Box, G.E.P. and Tiao, G.C. (1977). A canonical analysis of multiple time series. *Biometrika, 64*, 355–365.

Brunsdon, T.M. and Skinner, C.J. (1987). The analysis of dependencies between time series in psychological experiments. *British Journal of Mathematical and Statistical Psychology*, *40*, 125–139.

Chatfield, C. (1989). *Analysis of time series: An introduction. 4th Ed.* Chapman and Hall, London.

Cook, T.D. and Campbell, D.T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin, Boston.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley, Chichester.

Glass, G.V. , Willson, V.L. and Gottman, J.M. (1975). *Design and analysis of time series experiments*. Colorado Associated University Press, Boulder, Colorado.

Gorsuch, R.L. (1983). Three methods for analyzing limited time-series (*N* of 1) data. *Behavioral Assessment*, *5*, 141–154.

Gottman, J.M. (1981). *Time series analysis*. Cambridge University Press, Cambridge.

Goodall, C. (1990). A survey of smoothing techniques. In: J. Fox and J. Scott Long (Eds.), *Modern methods of data analysis*, 126–176. Sage, London.

Gregson, R.A.M. (1983). *Time series in psychology*. Erlbaum, Hillsdale, N.J.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized additive models*. Chapman and Hall, London.

Hibbs, D.A., jr. (1977). On analyzing the effects of policy interventions: Box-Jenkins and Box-Tiao versus structural equation models. In: D.R. Heise (Ed.), *Sociological Methodology 1977*. Jossey-Bass, San Francisco.

Immink, W. (1986). *Parameter estimation in Markov models and dynamic factor analysis*. Dissertation, University of Utrecht.

Jones, R.H., Crowell, D.H. and Kapuniai, L.E. (1969). Change detection model for serially correlated data. *Pychological Bulletin*, *71*, 352–358.

Kazdin, A.E. (1982). *Single case research designs: methods for clinical and applied settings*. Oxford University Press, New York.

Kiers, H.A.L. (1990). Majorization as a tool for optimizing a class of matrix functions. *Psychometrika*, *55*, 417–428.

Kratochwill, T.R. (Ed.) (1978). *Single subject research: strategies for evaluating change*. Academic Press, New York.

Kroonenberg, P.M. (1983). *Three-mode principal component analysis*. DSWO Press, Leiden.

Landis, J.R. and Koch, G.G. (1979). The analysis of categorical data in longitudinal studies of behavioral development. In: J.R. Nesselroade and P.B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*, 233–261. Academic Press, New York.

Larsen, R.J. (1987). The stability of mood variability: a spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology*, *52*, 1195–1204.

Levenson, R.W. and Gottman, J.M. (1983). Marital interaction: physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, *45*, 587–597.

MacCleary, R. and Hay, R.A. (1980). *Applied time series analysis for the social sciences*. Sage, Beverly Hills.

Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.

Molenaar, P.C.M. (1987). Dynamic factor analysis in the frequency domain: Causal modelling of multivariate psychophysiological time series. *Multivariate Behavioral Research, 22*, 329–353.

Oud, H. , Van den Bercken, J.H.L. and Essers, R.J. (1990). Longitudinal factor score estimation using the Kalman filter. *Applied Psychological Measurement, 14*, 395–418.

Parzen, E. and Newton, H.J. (1980). Multiple time series modelling II. In P.R. Krishnaiah (Ed.), *Multivariate analysis V*, 181–197. North Holland, Amsterdam.

Revenstorf, D., Hahlweg, K. and Schindler, L. (1978). Lead and lag in aspects of marital interaction. *Behavior Analysis and Modification, 2*, 174–184.

Shumway, R.H. (1988). *Applied statistical time series analysis.* Prentice-Hall, Englewood Cliffs.

Skinner, C.J. (1991). Time series. In: P. Lovie and A.D. Lovie (Eds.), *New developments in statistics for psychology and the social sciences. Vol. 2.* British Psychological Society and Routledge, London.

Van der Burg, E., de Leeuw, J. and Verdegaal, R. (1988). Homogeneity analysis with k sets of variables: an alternating least squares method with optimal scaling features. *Psychometrika, 53*, 177–197.

Velu, R.P. , Reinsel, G.C. and Wichern, D.W. (1986). Reduced rank regression models for multiple time series. *Biometrika, 73*, 105–118.

Visser, R.A. (1982). *On quantitative longitudinal data in psychological research.* Dissertation, University of Leiden. Published in 1985 as: Analysis of longitudinal data in behavioural and social research. DSWO Press, Leiden.

Von Eye, A. ,Ed. (1990). *Statistical methods in longitudinal research. Vol. 2: Time series and categorical longitudinal data.* Academic Press, New York.