LEAST SQUARES OPTIMIZATION OF LINEAR DYNAMIC SYSTEMS USING MAJORIZATION, MULTIPLIER AND QUASI-NEWTON METHODS

('3 opties.prm' dd 14/4/92)

p r e l i m i n a r y   v e r s i o n   p l e a s e   d o n ' t   q u o t e

Frits D. Bijleveld*, Catrien C.J.H. Bijleveld§,

Jan de Leeuw¶, Siem Oppe*

* SWOV Institute for Road Safety Research, Leidschendam, the Netherlands
§ Dept. of Psychometrics and Research Methodology, Faculty of Social Sciences, Leiden University, the Netherlands
¶ Depts. of Psychology and Mathematics, University of California at Los Angeles

Abstract

We discuss three numerical optimization methods for fitting linear dynamic systems. The first is the existing DYNAMALS algorithm, that uses Alternating Least Squares and a majorization substep. The second method uses Alternating Least Squares together with multiplier penalty functions or sequential augmented Lagrangian methods. In the third method, estimates of the unknowns are obtained in one least squares step that uses the BFGS algorithm after direct implementation. The last method is recommended for most cases; for unstable systems, and for errors-in-variables solutions, which includes the conventional approach to the analysis of linear dynamic systems, the DYNAMALS algorithm is recommended. For very long chains, the method using multiplier penalty functions is most appropriate. The methods are illustrated with a real life example.

Key words: Linear dynamic regression, state space models, optimization methods, road safety.

Introduction

Recently, Bijleveld and De Leeuw proposed an algorithm for fitting the longitudinal reduced rank regression or state space model (Bijleveld & De Leeuw, 1991). One of the drawbacks of their algorithm was the cumbersome majorisation procedure, and the numerical inevitability of an errors-in-variables solution for the dynamic regression problem. In the following we will propose an extension to their procedure that uses multiplier penalty functions, as well as a somewhat different method that directly estimates a number of unknowns using a quasi-Newton type of optimization procedure. Both methods produce solutions for the latent variables that are completely in the space of the predictor variables. Moreover, the proposed algorithms appear to be simple, efficient and reliable. The practical applicability of the methods will be illustrated through an example from traffic safety research.

The state space model

In the following vectors will be denoted by bold lower case characters, matrices by bold upper case characters. Suppose we have observed k input variables **x** and m output variables **y** at T consecutive occasions. We suppose that there is a time dependency in the measurements, which is modeled by supposing that the **x** influence the **y** through an unobserved or latent variable **z**. The **z** accommodate the time dependency in the measurements by following a Markov type of dependency; the latent states **z** thereby serve as the memory of the system. Usually, the dimensionality of the **z** is lower than that of the **x**, and in that sense the **z** also filter the dependence of the output on the input. A geometric representation of this model is in Figure 1.

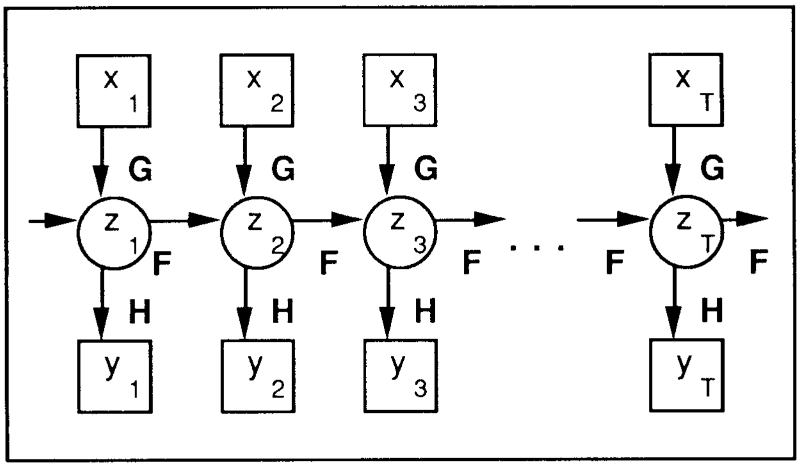

Figure 1. Geometric representation of the state space model

In formula this model can be written as follows:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{G}\,\mathbf{x}_t \qquad \text{(system equation)} \quad (1)$$

$$\mathbf{y}_t = \mathbf{H}\,\mathbf{z}_t, \qquad \text{(measurement equation)} \quad (2)$$

where ' '$_t$ is an observation at timepoint t, **F** is the $p \times p$ state transition matrix, **G** is the $p \times k$ control matrix, **H** is the $m \times p$ measurement matrix, and **z**, **x** and **y** are vectors of dimensionalities $p \times 1$, $k \times 1$ and $m \times 1$ respectively. Model (1) (2) is known under various names: linear dynamic system, state space model, longitudinal reduced rank regression model, linear dynamic model. For a discussion of the state space model see O'Connell (1984). In matrix notation the model becomes:

$$\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}' \qquad (3)$$

$$\mathbf{Y} = \mathbf{ZH}' \qquad (4)$$

with **Z** the $T \times p$ matrix containing the latent states from time 1 until time T, **X** the $T \times k$ matrix containing the input variables from time 1 until time T, **Y** the $T \times m$ matrix containing the output variables from time 1 until time T, and **B** the $T \times T$ shift matrix such that $\mathbf{B}(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T) = (\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_{T-1})$. In fitting this model the following loss function was proposed by Bijleveld and De Leeuw (1991):

normalized fit should here be written as (m-SSQ(**Y**-**ZH**'))/m (see Bijleveld, 1989, p 115). The correlation between input and state was -.148; the correlation between state and output equalled .244. The inadequateness of this summary of the development of this type of accident in relation to the intervention is further illustrated by Figure 5, which shows how the strongly cyclical development of the casualties is characterized by the almost straight line of the latent state space scores that have a teeny-weeny little dip at the time of intervention.

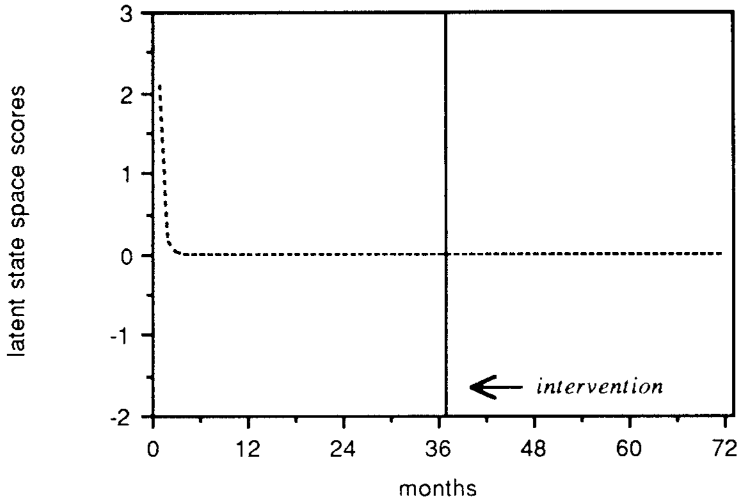

Figure 5. Latent state space scores from one-dimensional DYNAMALS solution with quasi-Newton option

The data were again submitted to DYNAMALS; this time we modelled two dimensions for the latent state. The matrix **G** now equalled:

-.0232
-.0520

The analysis results are in Table 3.

Table 3. DYNAMALS analysis with quasi Newton option

		correlations of the latent state with		
fit	.810	input	-.127	.049
|F|	1.267	output	.142	.766
# iterations	40			

Thus, the first latent state dimension contains the effect of the intervention measure, and the second latent state dimension reflects the autonomous behaviour of this type of road accident only. The scores on the two latent state space dimensions are plotted in Figures 6a and 6b. In Figure 6a, the highs and lows of the sinusoid have for purposes of clarity been connected. Figure 6a clearly illustrates the effect of the intervention measure; both the highs and the lows of the sinusoid are lowered by the intervention and, as it appears, the lows even more so than the highs.

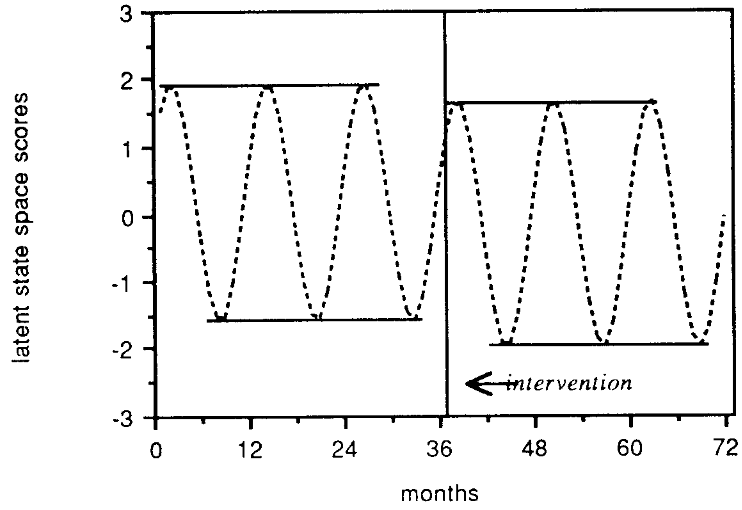

Figure 6a. Scores on the first latent state dimension from two dimensional DYNAMALS solution with the quasi-Newton option

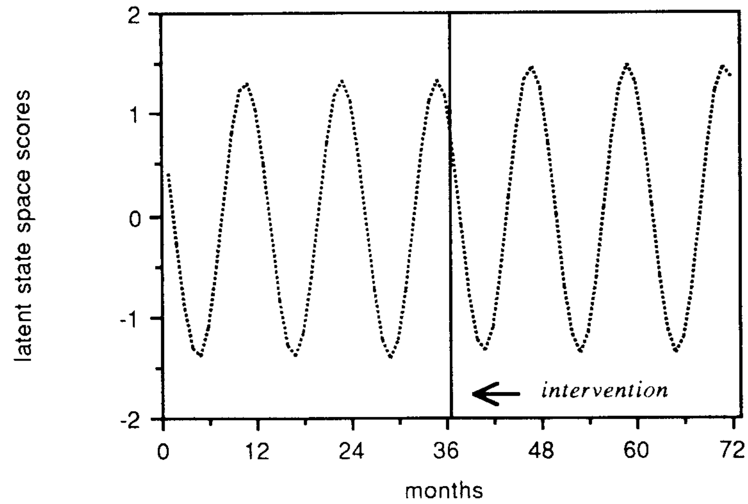

Figure 6b. Scores on the second latent state dimension from two dimensional DYNAMALS solution with the quasi-Newton option

To check the interpretation, an alternative input model could be a continious decrease in this type of accidents. Figure 6b shows the cyclical behaviour of the casualties combined with a (small) increasing trend. This component in obviously more related to the output than to the input.

On the basis of the analyses performed here, it appears as if there is an empirically demonstrable effect of the intervention. Because of the cyclical behaviour of the series, this effect would have been hard to demonstrate if one had restricted the analysis to just one dimension. The two-dimensional DYNAMALS analyses showed how the casualties could be related to the intervention measure in a meaningful way.

Discussion

All three methods can in a fairly simple way be extended with optimal scaling options; for general information see Gifi (1990), for specific information on optimal scaling in DYNAMALS, see Bijleveld and De Leeuw (1991).

In the direct implementation method without optimal scaling it is practically ensured that the algorithm does not get stuck in a saddle point or point of inflexion; of course, there is no way to avoid local minima in either of the three methods, as in fact in all existing algorithms. If the norm of **F** is much larger than 1, and T is also large, the direct method may become very slow due to the necessity of decreasing the step sizes, alternating least squares using majorization might be a better option in this case. In case of extremely long chains, the sequential augmented Lagrangian method might be better, if there are not too many constraints. For fitting a regression model, the direct method is probably superior to the other two, unless one clearly prefers orthogonal latent states. For investigating latent states that are supposed to follow something that resembles an exponential curve, the direct method is probably best suited.

From practical experience, it appears that all three methods converge at acceptable speed. The numerical performance of the three methods will be compared more systematically in subsequent publications.

References

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

Bijleveld, C.C.J.H. and De Leeuw, J. (1991). Fitting longitudinal reduced rank regression models by Alternating Least Squares. *Psychometrika, 56*, 433-447.

Blokpoel, A. (1987). Zijreflectie bij fietsen. De ontwikkeling van de aanwezigheid van zijreflectie; overwegingen en aanbevelingen voor maatregelen. [*Side-reflection for bicycles. The development of the installatioon of side-reflection; considerations and recommendations for intervention*] Leidschendam: SWOV, R 87-24.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33-42.

Fletcher, R. (1981a). *Practical Methods of optimization. Vol I. Unconstrained optimization*. Chichester: Wiley.

Fletcher, R. (1981b). *Practical Methods of optimization. Vol II. Constrained optimization*. Chichester: Wiley.

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

O'Connell, P.E. (1984). Kalman filtering. In : Ledermann, W. (Ed.) *Handbook of Applicable Mathematics Vol VI Statistics B*. Chichester: Wiley.

Powell, M.J.D. (1969). A method for nonlinear constraints in minimization problems. In: R. Fletcher, R. (Ed.) *Optimization*. London: Academic Press.

[1] The DYNAMALS program is currently being extended with options for the sequential augmented Lagrangian method and the direct implementation/quasi-Newton method. Prototypes of both algorithms, written as Macintosh or Dos-PC applications, are available from the first author.

$$\sigma_{\omega} = \omega^2 \, SSQ(\mathbf{Z} - \mathbf{BZF}' - \mathbf{XG}') + \mathbf{SSQ}\,(\mathbf{Y} - \mathbf{ZH}'), \qquad (5)$$

where 'SSQ' stands for the sum of squares over all the arguments. The weight ω defines the relative importance of the input and output. If ω=1, input and output play an equally balanced role; the model then is the linear dynamic or state space model. As ω becomes smaller, the output becomes more and more important, which has the consequence that the latent states **Z** are more and more in the space of the output variables; the bordering case with ω=0 yields the principal components analysis solution for **Z** and **H**. As ω becomes greater than 1, the input becomes more and more important, and in the limiting case where $\omega \rightarrow \infty$, the latent states are situated in the space of the input variables. This is obviously the most attractive choice, as it is the dynamic analogue of the ordinary least squares regression case.

Fitting the state space model using alternating least squares with majorization

In Bijleveld & De Leeuw's algorithm, every iteration used two substeps for updating the least squares estimates of the transition matrices and the latent states. Estimation in the latent states substep proved to be rather complicated, since the restriction **Z'Z=I** was needed to obtain meaningful solutions. First **Z** = SVD(**Y**).

Ordinary least squares for **F**, **G**, and **H**

Consider the problem of minimizing (5) over **F**, **G** and **H**, for given current **Z**. The solution for **H** is straightforward, since from (4), **H'** = **Z' Y** is the solution for which SSQ(**Y**-**ZH'**) is minimal.

Write **R** for the partitioned matrix: **F'**//**G'**, where '//' stands for vertical concatenation. From (3) 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 '$(\mathbf{BZ}||\mathbf{X})^{+}$' stands for the generalized inverse $((\mathbf{BZ}||\mathbf{X})'(\mathbf{BZ}||\mathbf{X}))^{-1}(\mathbf{BZ}||\mathbf{X})'$. Thus, for **F** estimated as the transpose of the first *p* rows of **R**, and **G** estimated as the transpose of the last *k* rows of **R**, SSQ(**Z** - **BZF'** - **XG'**) is minimal.

Majorization

Consider the problem of minimizing (5) over **Z**, with **Z'Z=I**, and with the parameters in the matrices **F**, **G**, and **H** (temporarily) regarded as known constants. Write $\mathbf{Z}=\mathbf{Z}_{old}+\Delta$, with $\mathbf{Z}_{old}$ the current best solution. Now, $\sigma_{\omega}(\mathbf{Z},\mathbf{F},\mathbf{G},\mathbf{H})$ equals

$$\omega^2 \; SSQ\{(\mathbf{Z}_{old} - \mathbf{BZ}_{old}\mathbf{F}' - \mathbf{XG}') + (\Delta - \mathbf{B}\Delta\mathbf{F}')\} + SSQ\{(\mathbf{Y} - \mathbf{Z}_{old}\mathbf{H}') - \Delta\mathbf{H}'\}.$$

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

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

$$2\omega^2 \,\mathrm{tr}\, \Delta'(\mathbf{B}'\mathbf{P}_1\mathbf{F} - \mathbf{P}_1) - 2\,\mathrm{tr}\, \Delta'\mathbf{P}_2\mathbf{H} + \omega^2 SSQ(\Delta - \mathbf{B}\Delta\mathbf{F}') + SSQ(\Delta\mathbf{H}'). \qquad (6)$$

Now suppose a bound can be found of the form

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

where γ is estimated as the square of the largest singular value of the partitioned matrix $\omega(\mathbf{I}\text{-}\mathbf{B}\otimes\mathbf{F})//\mathbf{I}\otimes\mathbf{H}$, and define

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

Then

$$\sigma_{\omega}(\mathbf{Z,F,G,H}) \leq \sigma_{\omega}(\mathbf{Z}_{\text{old}},\mathbf{F,G,H}) + \gamma SSQ(\Delta - \mathbf{S}) - \gamma SSQ(\mathbf{S}). \tag{7}$$

But $SSQ(\Delta\text{-}\mathbf{S}) = SSQ(\mathbf{Z}\text{-}(\mathbf{Z}_{\text{old}}\text{+}\mathbf{S}))$. An iteration step of this algorithm consists of minimizing $SSQ(\mathbf{Z}\text{-}(\mathbf{Z}_{\text{old}}\text{+}\mathbf{S}))$ over $\mathbf{Z}$ satisfying $\mathbf{Z'Z} = \mathbf{I}$. This is a simple Procrustes problem (Cliff, 1966), whose solution is well-known. If $\mathbf{Z}_{\text{old}}+\mathbf{S} = \mathbf{K\Lambda L}'$ 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 (6) itself, (6) is minimized indirectly through a majorization algorithm, in which the simpler loss function at the right hand side of (7) is minimized, of which it is known that its values are always higher than or equal to those of (6).

Thus, the Alternating Least Squares method for estimating **F**, **G**, **H**, and **Z** goes as follows. Start with initial values for **Z**. Compute optimal values for **F**, **G**, and **H** using ordinary least squares; given these estimates, compute optimal values for **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 **Z** to estimate new optimal estimates of **F**, **G**, and **H**, compute new estimates of **Z**, and so forth. Instead of using the iterative majorization procedure to estimate **Z**, it might be better (in terms of overall speed of convergence) to alternate a non-iterative single $\mathbf{Z}_{\text{old}} \rightarrow \mathbf{Z}_{\text{new}}$ substep with the (**F**,**G**,**H**) substep. A normalized fit measure can be computed as $(\omega^2 p+m-\sigma_{\omega}(\mathbf{Z,F,G,H}))/(\omega^2 p+m)$. Details are in Bijleveld (1989) and in Bijleveld and De Leeuw (1991).

Although solutions converge quickly to a solution which is close to the desired solution, obviously, relatively large values of ω lead to numerical complications.

Fitting the state space model using alternating least squares with sequential augmented Lagrangian methods

We will now propose to minimize the loss function (5) under the constraint $\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}'$. This amounts to setting $\omega\rightarrow\infty$, see above. The algorithm we will use is one of a family of methods known as sequential augmented Lagrangian methods, multiplier methods or multiplier penalty functions. We use the model given by Fletcher (1981b, p.130):

$$\phi\ (\mathbf{x},\theta,\sigma) = f\ (\mathbf{x}) + .5\ (\mathbf{c}(\mathbf{x}) - \theta)^T\mathbf{S}\ (\mathbf{c}(\mathbf{x}) - \theta), \tag{8}$$

with $\mathbf{x}$ a vector containing the data (i.e. **X**, **Y**, **Z**, **F**, **G**, and **H**), θ and σ vectors containing the algorithm-specific parameters, **S** a diagonal matrix containing the elements of σ, $\mathbf{c}(\mathbf{x})$ the constraint, in this case $\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}'$, and $f(\mathbf{x})$ a function of the data, in this case $SSQ(\mathbf{Y} - \mathbf{ZH}')$. Note that by setting θ=0 and $\mathbf{S}=2\omega^2$, (8) reduces to (5).

Now choose $\lambda_i = \theta_i \sigma_i$, with i=1,2, ..., m, with m the number of constraints, in this case the number of observations multiplied by the dimensionality of the statespace. Model (8) can equivalently be written as

$$\phi\ (\mathbf{x},\lambda,\sigma) = f\ (\mathbf{x}) - \lambda^T\ \mathbf{c}(\mathbf{x})^T\mathbf{S}\ \mathbf{c}(\mathbf{x}). \qquad (9)$$

For minimizing (9), we use Powell's algorithm (1969), which is summarized as follows:

(i) Initially set $\lambda = \lambda^{(1)}=\mathbf{0}$, $\sigma = \sigma^{(1)}=\mathbf{1}$, $k = 0$, $\|\mathbf{c}^{(0)}\|_\infty=\infty$.
(ii) Find the minimizer $\mathbf{x}(\lambda,\sigma)$ of $\phi\ (\mathbf{x},\lambda,\sigma)$ and denote $\mathbf{c} = \mathbf{c}(\mathbf{x}(\lambda,\sigma))$.
(iii) If $\|\mathbf{c}\|_\infty > 1/4\ \|\mathbf{c}^{(k-1)}\|_\infty$ set $\sigma_i=10\sigma_i$ $\forall\ i : |\ c_i\ | > 1/4\ \|\mathbf{c}^{(k-1)}\|_\infty$ and go to (ii).
(iv) Set $k = k+1$, $\lambda^{(k)}= \lambda$, $\sigma^{(k)}= \sigma$, $\mathbf{c}^{(k)}= \mathbf{c}$.
(v) Set $\lambda = \lambda^{(k)}- \mathbf{S}^{(k)}\mathbf{c}^{(k)}$ and then (ii).

The proposed method has the advantage that the cumbersome majorization procedure is not needed anymore. Furthermore, the restriction $\mathbf{Z'Z=I}$ is also not required any more. We nevertheless orthonormalize $\mathbf{Z}$ after each substep, because the proposed algorithm approaches zero error, but in practice never actually reaches it.

A program and a problem

Implementation of this algorithm immediately causes a small problem. Generally, the flaws of algorithms with many constraints are that some constraints may contradict each other. In this case, the default estimation used in DYNAMALS of the starting state $\mathbf{z}_0$ as $\hat{\mathbf{z}}_0=\mathbf{z}_1$ is contradictory with the constraint requesting $\mathbf{z}_t = \mathbf{F}\ \mathbf{z}_{t-1} + \mathbf{G}\ \mathbf{x}_t$. This causes the algorithm to explode: σ approaches infinity and the fit decreases. We have circumvented this problem by estimating $\mathbf{z}_0$ alongside with $\mathbf{z}_1$, ..., $\mathbf{z}_T$ from the system and measurement equations. The algorithm can thus be summarized as follows:

(1) Run DYNAMALS without majorization until convergence.
(2) Estimate $\mathbf{z}_0$ from the system and measurement equations.
(3) Obtain new estimates of $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$ using the Lagrangian method.
(4) Orthonormalize $\mathbf{Z}$.
(5) Estimate $\mathbf{z}_0$ from the system and measurement equations.
(6) unless convergence (3)

Fitting the state space model using direct least squares implementation with quasi Newton methods

A simpler and generally more efficient method for our minimization problem is the following. Suppose that the Latent states are completely in the space of the predictor variables.Then, the subsequent latent state values $\mathbf{z}_0$ to $\mathbf{z}_T$ may per time point be written as:

$$\mathbf{z}_1= \mathbf{F}\ \mathbf{z}_0 + \mathbf{G}\mathbf{x}_1,$$

$$\mathbf{z}_2= \mathbf{F}\ (\mathbf{F}\mathbf{z}_0 + \mathbf{G}\mathbf{x}_1) + \mathbf{G}\mathbf{x}_2,$$

$$\mathbf{z}_3= \mathbf{F}\ ((\mathbf{F}\mathbf{z}_0 + \mathbf{G}\mathbf{x}_1) + \mathbf{G}\mathbf{x}_2) + \mathbf{G}\mathbf{x}_3, \text{ etc.}$$

This may be summarized as:

$$\mathbf{z}_t = \mathbf{F}^t\mathbf{z}_0 + \Sigma_{g=0}^{t-1} \mathbf{F}^g\mathbf{G}\mathbf{x}_{t-g}, \qquad (10)$$

so that $\mathbf{z}_t = f_t(\mathbf{F},\mathbf{G},\mathbf{z}_0)$. Loss function (5) may then be written as:

$\Sigma_{t=1}^{T} \Sigma_{j=1}^{m} (\mathbf{y}_{tj}-(\mathbf{H}\mathbf{z}_t)_j)^2$, or, keeping in mind (10), as:

$$ssq = \Sigma_{t=1}^{T}\Sigma_{j=1}^{m} (\mathbf{y}_{tj} - \Sigma_{h=1}^{p}\mathbf{H}_{jh} f(\mathbf{F},\mathbf{G},\mathbf{z}_0)_{th})^2, \qquad (11)$$

with h the index for the p dimensions of the latent state $\mathbf{z}$, and j the index for the m dimensions of the output $\mathbf{y}$. If we write ξ for any parameter of $\{\mathbf{F},\mathbf{G},\mathbf{z}_0\}$, the derivative of (11) equals:

$$\frac{\partial\,(ssq)}{\partial\,\xi} = -2\,\Sigma_{t=1}^{T}\Sigma_{j=1}^{m} (\mathbf{y}_{tj}-\Sigma_{h=1}^{p}\mathbf{H}_{jh} f(\mathbf{F},\mathbf{G},\mathbf{z}_0)_{th}).(\Sigma_{h=1}^{p}\mathbf{H}_{jh}\frac{\partial \mathbf{z}_{th}}{\partial\,\xi}) \quad \text{for } \xi \in \{\mathbf{F},\mathbf{G},\mathbf{z}_0\},$$

$$\frac{\partial\,(ssq)}{\partial\,\mathbf{H}_{jh}} = -2\,\Sigma_{t=1}^{T} (\mathbf{y}_{tj}-\Sigma_{h=1}^{p}\mathbf{H}_{jh} f(\mathbf{F},\mathbf{G},\mathbf{z}_0)_{th}).\mathbf{z}_{th} \qquad \text{otherwise.}$$

The values of the parameters for which the derivative is closest to zero are estimated using the BFGS or Broyden-Fletcher-Goldfarb-Shanno algorithm (Fletcher, 1981a, pp. 33-62). This quasi-Newton type algorithm uses the gradient and information on the values of the function to determine the search area for the minimum. It incorporates knowledge on the past iterations in its search, which makes it very efficient and stable.

If during a search step any $\mathbf{z}_t$ becomes very large, we divide the step size by a suitable number, thereby decreasing the step size, until a reasonably small value for $\mathbf{z}_t$ has been obtained. This local constraint has been introduced to avoid numerical problems, that might occur by chance, or when the norm of $\mathbf{F}$ is greater than 1. We have thus far set arbitrary criteria for the largeness of $\mathbf{z}_t$, as well as for the suitableness of the number it is subsequently divided by.

DYNAMALS II does not orthonormalize the latent states. This may be an advantage or a disadvantage, depending on one's paradigm or the structure in the data.[1]

The effect of intervention on casualties in cyclists' road accidents- a real life example

We analyze the number of fatal and non-fatal casualties in the Netherlands from night-time accidents, where bicyclists are hit sideways by any other traffic users. This example serves to illustrate the methods, and does not pretend to give an evaluation of the data. On January 1 1987, a new bicycle safety device became obligatory in the Netherlands; this device consisted of reflecting bands either on the bicycle tyres or in its spokes. It was assumed that this would make cyclists better visible for any traffic participants that carried light, and thus would reduce the number and seriousness of accidents after dusk, where bicyclists are hit sideways (Blokpoel, 1987).

The SWOV Institute for Road Safety Research kindly provided us with monthly data on this type of accident for the period January 1 1984 until December 31 1989, derived from VOR data. The casualties variable is plotted in Figure 2. The legal intervention took place at month 37, which is exactly mid-way in our series of 72 observation points. The intervention was introduced to the model as a dummy variable with coding 0 for the months that the reflection devices were not obligatory yet, and coding 1 for the months where the devices had become obligatory.

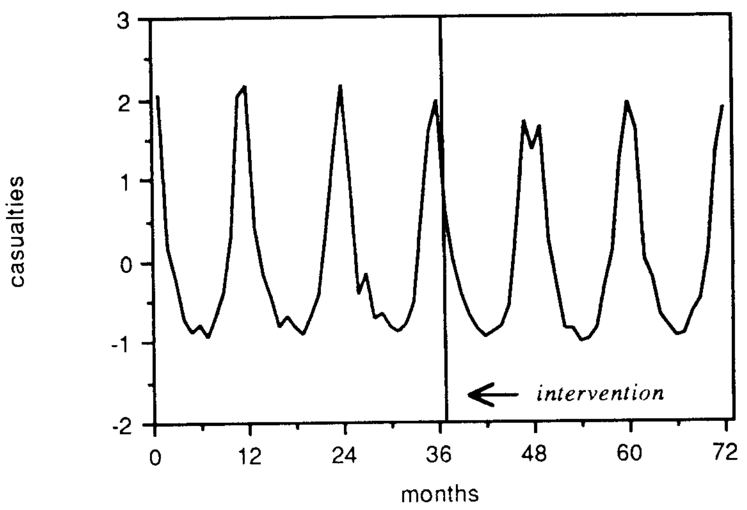

Figure 2. Monthly casualties from January 1984 to December 1989

A problem with the analysis of the effectiveness or the impact of such prohibitory legal measures as the bicycle reflector, or, for that matter, the safety belt, is that the onset of the impact of the legal measure occurs before the measure is properly enforced: traffic users install the device in their car or on their bicycle before the deadline, and usually much longer before that. In the Netherlands this can be most clearly seen from examples on safety belt introduction. A problem of a more technical nature is that the casualties data exhibit a strongly cyclical character: in winter, when days are shortest, more accidents of this type occur; in summer the reverse is true.

Because the method using augmented Lagrangian methods was thought to be less efficient for the analysis of our data, we will illustrate only DYNAMALS using alternating Least Squares with majorization, and DYNAMALS using the direct option.

DYNAMALS with the majorization option

The data were submitted to DYNAMALS. The dummy intervention variable served as input variable, the casualties variable served as output. The convergence criterion was set at 5.10^{-5}. We modelled one dimension of the latent state for increasing values of ω. This produces a trajectory of solutions (see Bijleveld & De Leeuw, 1991), where the latent states are situated more and more in the space of the input variables. In this case, the latent states, apart from differences due to normalization, very quickly become virtually identical to the input; this happens in fact for ω larger than 2. To help the algorithm along, we chose initial values for **Z** as a zero mean version of the input. For large ω this may have the dramatic consequence that convergence is reached in two iterations: the solution is immediately optimal. A summary of the results of the analysis with ω=1 is in Table 1. A plot of the latent state space scores against time is in Figure 3; for illustrative purposes we have also drawn in the development of the values of the latent state for ω=2.

Table 1. DYNAMALS analyses for $\omega=1$

		correlations of the latent state with	
fit	.829	input	-.074
IFI	.833	output	.977
# iterations	28		

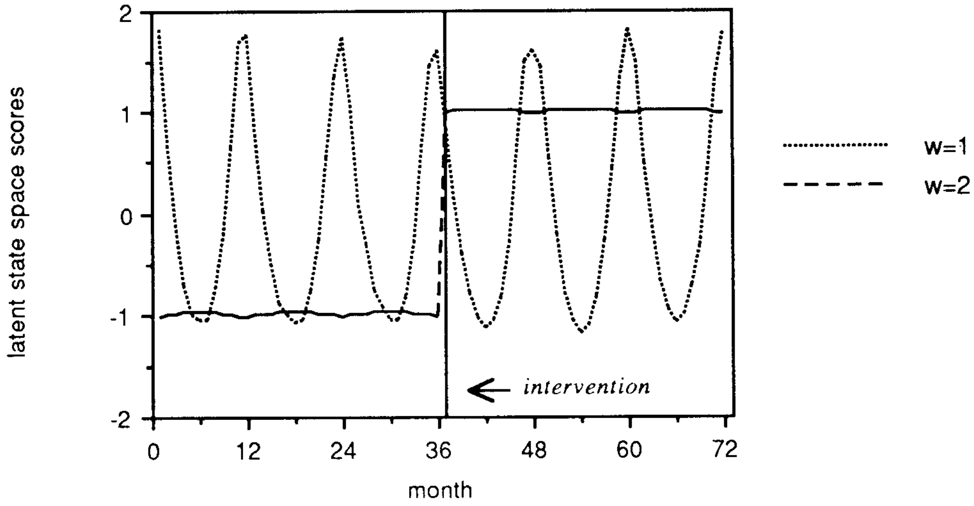

Figure 3. Latent state space scores against time for $\omega=1$ and $\omega=2$

From the solution, one can derive that there is a small, but positive impact of the intervention on the numbers of casualties. The solution is not entirely satisfactory, however. The development of the latent state space scores is not informative; for different values of ω the latent states follow something in between the cyclical curve of the casualties variable and the one-step function of the dummy intervention variable, that is, between the input and the output.

To further investigate the data, we ran the DYNAMALS analyses again, allowing for two dimensions of the latent state; we chose the same analysis options. The results of these endeavours are in Table 2. (We did not consider DYNAMALS analyses for one dimensional input with two-dimensional orthonormal latent states in the space of the input to be a useful effort.)

Table 2. DYNAMALS analyses with two dimensions of the latent state for $\omega=1$

		correlations of the latent states with		
fit	.981	input	-.110	.960
IFI	1.811	output	.960	.068
# iterations	31			

The first latent state dimension may be characterized as a casualties dimension; from the correlation of the input with this dimension, one can again see that there is a small but positive impact of the intervention on the numbers of casualties. The second dimension is the orthogonal complement of this dimension. It follows the intervention variable in that it starts low, increases and next remains high; it exhibits small peaks however, presumably due to the orthogonalisation constraint. Both curves are in Figures 4a and 4b.

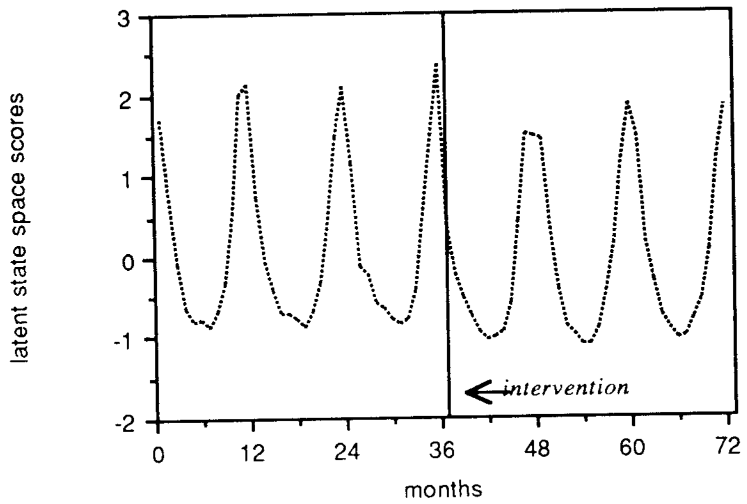

Figure 4a. First dimension of the latent state space scores from DYNAMALS solution with alternating least squares and majorization

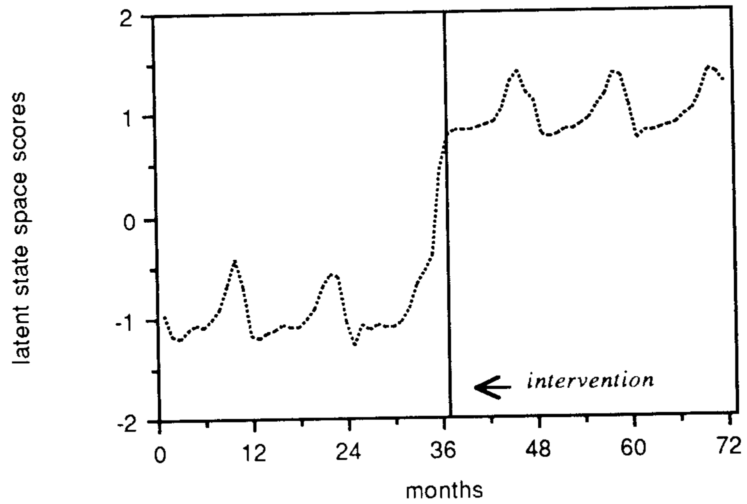

Figure 4b. Second dimension of the latent state space scores from DYNAMALS solution with alternating least squares and majorization

DYNAMALS with the quasi-Newton option

As in the alternating least squares with majorization DYNAMALS option, we started by modelling one dimension for the latent state. The algorithm converged fairly quickly, in 33 iterations, to a normalized fit of .073.; as the loss in the input part of the loss function (5) is here by definition equal to zero, the formula for the