431

FAMILIES OF STATE SPACE MODELS

JAN DE LEEUW, CATRIEN BIJLEVELD, AND FRITS BIJLEVELD

March 16, 1994

1. Introduction

Suppose that, for N subjects, we have measured at T time points, k input variables x, and we have stacked our observations in a data box X of size $(N \times k \times T)$. A set of output variables, measured at the same time points, is named y, and it is stacked in a data matrix Y of size $(N \times m \times T)$. Suppose moreover that the x are understood to influence the y variables, and that the time dependence of the measurements itself is of interest. Thus, we have two data boxes X and Y, where the rows represent the subjects, the columns represent the variables, and where the time points are stacked in the third dimension, viz. Figure ??. In Figure ??, three methods for slicing the data box can be distinguished. For instance for data box X, a first method is by slicing the data box horizontally: we then obtain a $k \times T$ datamatrix for one subject, which is the usual data matrix for time series analysis. When we slice the front of the data box, we get an $N \times T$ data matrix, that is T observations on one variable for N subjects, a not very common type of data matrix to analyze. When we slice the side of the data matrix, we get an $N \times k$ data matrix, which is a cross-sectional data matrix of N subjects measured over k variables.

We will consider models where the impact of the input on the output is supposed to be mediated by an unobserved factor, the so-called latent state, stacked in a data box Z, of dimensionality $(N \times p \times T)$. The input influences the output in the sense that the input variables influence the state, and the state in turn influences the output variables. At the same time, the state embodies the time-dependence in the measurements, transferring information from each time point to the next. In system analytic terms, the input variables X, latent state variables Z, output variables Y, and the relations between these, together constitute a system. The state at any time point t, comprises all information from the former time points that is relevant for the future. As such, the system is Markovian, and the state functions as the memory of the system. The dimensionality of the state, which we refer to as p, is usually lower than that of the input and/or the output. As such, the state also serves as a filter between input and output. Thus, the state is of crucial importance in a system.

In the following, we assume that our models are discrete, that is, we assume that both the latent state and the observed variables are categorical. Our expressions however also apply to continuous versions of the variables if probabilities are replaced by densities.

2. The Causal Structure of State Space Models

There is much recent research dealing with modeling the dependence structure of variables by using graphs. The interesting part of this research is to do the modeling in such a way that the structural properties of graphs, such as connectivity and separation, are isomorphic with the dependence structure of the variables. This is often called "causal modeling", a description which is unnecessarily controversial. Models are filters, or smoothers, which can be used to bring data in a form which is more interesting, more easy to communicate, or more easy to relate to existing theory. We separate the signal from the noise by using prior theory whenever it is available, and by using inductive techniques otherwise. We can talk about our results in causal language, as long as we realize that this language is simply another (verbal) model with which we overlay the statistical analysis. Verbal models are vague, and lead to many possible misconceptions.

The so-called causal models have also been used, mainly in the social and behavioural sciences, to dress up weak data. Using causal terminology suggest an invariance which simply is not there in these cases, because the outcomes depend largely on accidental properties of the data and arbitrary choices of the researcher. This has given causal models a bad name, although obviously the problem is not with the model but with the data and the way the model is applied to the data (and perhaps the way the techniques have been sold commercially).

Thus there are many reasons to distrust highly specific path models in which some of the arrows between variables are present, and some of the other comparable arrows are absent, and there are no clear reasons for either presence or absence. We prefer full models, in which the dependence that is modeled depends on some global and fairly uncontroversial choices. Exploratory factor analysis is one example, multiple regression is another. All the arrows are there between the predictors and the criterium, or between the factors and the indicators. Such full models are much more descriptive than the models whose fine structure suggests much more prior knowledge than we actually have. They can be used as data reduction techniques, and in fact in most cases they are not far from saturated models. Multiple regression and complete recursive path models are saturated, factor analysis and the state space models we discuss here can be made saturated by introducing sufficiently many latent variables.

The choices we have to make to draw our path models or graphs or arrow diagrams are really simple. In multiple regression we only have to choose which variable is the criterion, in factor analysis we only have to choose the number of factors, in MIMIC type models we have to distinguish input and output variables and choose the number of factors, and in linear dynamic systems we have to order the blocks of input, state, and output variables in time. These are simple global choices, with which few people will disagree. The filtering done by the model is entirely in the dimensionality of the state space or factor space, and we easily see the effect of this by looking at different dimensionalities.

3. State Space Models

The basic model we are interested in is drawn in Figure 1.

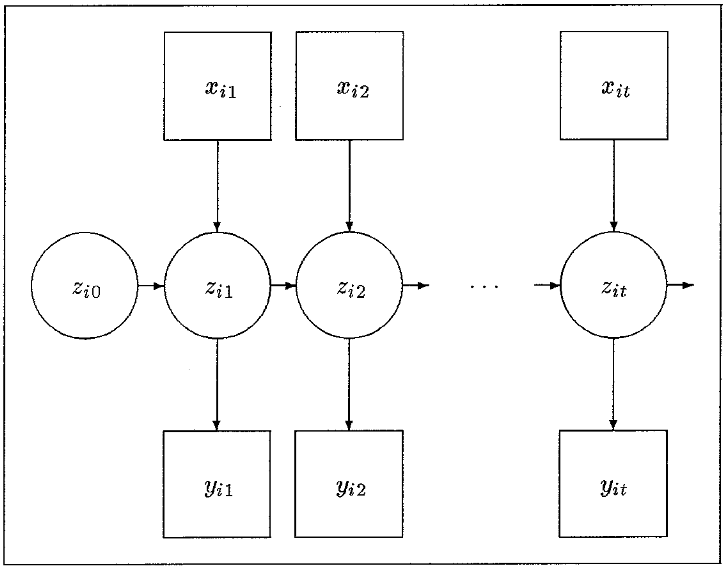

Figure 1. *State Space Model for Individual i.*

Actually there are n such models, one for each individual. We write prob$[(\wedge_{i=1}^{n} \wedge_{t=1}^{T} y_{it})(\wedge_{i=1}^{n} \wedge_{t=0}^{T} z_{it})(\wedge_{i=1}^{n} \wedge_{t=1}^{T} x_{it})]$ for the probability of observing the data-boxes $X, Y,$ and Z. Our basic task in this section is to derive a general expression for this probability, taking the properties of the model in Figure 1 into account. The key result used to translate directed acyclic graphs into statements about joint distributions is a simple one. We suppose that, given z_{it}, y_{it} is independent of all other variables in the system. Also, given $z_{i,t-1}$ and x_it, z_{it} is independent of all other variables in the system.

We first assume individuals are independent. This means

$$\begin{aligned}&\text{prob}[(\wedge_{i=1}^{n}\wedge_{t=1}^{T} y_{it})(\wedge_{i=1}^{n}\wedge_{t=0}^{T} z_{it})(\wedge_{i=1}^{n}\wedge_{t=1}^{T} x_{it})]\\ &=\prod_{i=1}^{n}\text{prob}[(\wedge_{t=1}^{T}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})]\end{aligned} \tag{1}$$

Theorem 1.

$$\begin{aligned}&\text{prob}[(\wedge_{t=1}^{T}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})] =\\ &\text{prob}[\wedge_{t=1}^{T}x_{it}\mid z_{i0}]\text{prob}[z_{i0}]\prod_{t=1}^{T}\text{prob}[y_{it}\mid z_{it}]\text{prob}[z_{it}\mid z_{i,t-1}\wedge x_{i,t}]\end{aligned}$$

Proof. The proof is by induction over T. The result is trivially true for $T = 1$. Assume it is true for $T - 1$. Start with a simple application of the definition of conditional probability.

$$\begin{aligned}\text{prob}[(\wedge_{t=1}^{T}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})] =&\text{prob}[y_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})]\star\\ &\text{prob}[z_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T}x_{it})]\star\\ &\text{prob}[x_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T-1}x_{it})]\star\\ &\text{prob}[(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T-1}x_{it})]\end{aligned} \tag{2}$$

Figure 1 now tells us that

$$\text{prob}[y_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})] = \text{prob}[y_{iT}\mid z_{iT}] \tag{3}$$

$$\text{prob}[z_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T}x_{it})] = \text{prob}[z_{iT}\mid z_{i,T-1}\wedge x_{i,T}] \tag{4}$$

$$\text{prob}[x_{iT}\mid(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T-1}x_{it})] = \text{prob}[x_{iT}\mid\wedge_{t=1}^{T-1}x_{it}\wedge z_{i0}] \tag{5}$$

But this means that we have proved the recursion

$$\begin{aligned}&\text{prob}[(\wedge_{t=1}^{T}y_{it})(\wedge_{t=0}^{T}z_{it})(\wedge_{t=1}^{T}x_{it})] =\\ &\text{prob}[y_{iT}\mid z_{iT}]\text{prob}[z_{iT}\mid z_{i,T-1}\wedge x_{i,T}]\text{prob}[x_{iT}\mid\wedge_{t=1}^{T-1}x_{it}\wedge z_{i0}]\\ &\text{prob}[(\wedge_{t=1}^{T-1}y_{it})(\wedge_{t=0}^{T-1}z_{it})(\wedge_{t=1}^{T-1}x_{it})].\end{aligned} \tag{6}$$

By the induction hypothesis this means the result is true for T. □

Corollary 1. If

$$\text{prob}[\wedge_{t=1}^{T}x_{it}\mid z_{i0}] = \text{prob}[\wedge_{t=1}^{T}x_{it}]$$

then

$$\text{prob}[\wedge_{t=1}^{T} y_{it} \mid \wedge_{t=1}^{T} x_{it}] =$$
$$\int \cdots \int_{z_{i0},\ldots,z_{iT}} \text{prob}[z_{i0}] \prod_{t=1}^{T} \text{prob}[y_{it} \mid z_{it}]\text{prob}[z_{it} \mid z_{i,t-1} \wedge x_{i,t}] dz_{iT} \ldots dz_{i0}.$$

Proof. Start with the result in Theorem 1. We remove the marginal distribution of the input variables by conditioning, and then integrate out the state variables. □

4. Specific Submodels

There are a number of useful distinctions that can be drawn in discussing this class of models. In the first place there are models with and without input. There are models in which the state variables are discrete and models in which they are continuous. There are models in which the input and/or output variables are discrete or continuous. There are models which are cross-sectional, in the sense that $T = 1$, and models which are time-series, in the sense that $N = 1$. Discussing the models in these terms shows that they do indeed cover a lot of the latent variable models discussed in psychometrics and other disciplines. We shall discuss a number of these special cases more detail. This is a simple and straightforward widening of the framework introduced by Lazarsfeld [?] and Guttman [?] in the forties, and then extended by Anderson [1], McDonald [3], and Bartholomew [2] for cross-sectional models, and of the framework discussed for example by Metz [?] for time series.

In the class of cross-sectional models without input we find factor analysis (continuous state, continuous output), latent class analysis (discrete state, discrete output), latent trait analysis (continuous state, discrete output), and of course various combinations of these techniques.

5. Consequences: Moments

We use formula (7) to compute the conditional expected value and variance of the output given the input. The natural assumptions in this case are linearity of regression and homoscedasticity. We look at the equations for a fixed value of i. More precisely, we assume that the conditional expectation of $\underline{y}_t$ given $\wedge_{s=1}^{t}\underline{z}_s$ and $\wedge_{s=1}^{t}\underline{x}_s$ only depends linearly on $\underline{z}_t$. Moreover, the conditional expectation of $\underline{z}_t$ given $\wedge_{s=1}^{t-1}\underline{z}_s$ and $\wedge_{s=1}^{t}\underline{x}_s$ only depends on $\underline{z}_{t-1}$ and $\underline{x}_t$. Thus

$$\underline{y}_t = H_t\underline{z}_t + \underline{\delta}_t, \tag{7}$$

$$\underline{z}_t = F_t\underline{z}_{t-1} + G_t\underline{x}_t + \underline{\epsilon}_t, \tag{8}$$

where

$$\underline{\delta}_t \perp \wedge_{s=1}^{t} \underline{z}_s, \tag{9}$$

$$\underline{\delta}_t \perp \wedge_{s=1}^{t} \underline{x}_s, \tag{10}$$

$$\underline{\epsilon}_t \perp \wedge_{s=1}^{t-1} \underline{z}_s, \tag{11}$$

$$\underline{\epsilon}_t \perp \wedge_{s=1}^{t} \underline{x}_s, \tag{12}$$

and homoscedasticity means

$$\mathbf{V}\left(\underline{\delta}_t\right) = \Omega_t, \tag{13}$$

$$\mathbf{V}\left(\underline{\epsilon}_t\right) = \Theta_t. \tag{14}$$

Lemma 1.

$$\underline{\delta}_t \perp \wedge_{s=1}^{t} \underline{\epsilon}_s,$$

$$\underline{\epsilon}_t \perp \wedge_{s=1}^{t-1} \underline{\epsilon}_s. \tag{15}$$

Proof. Again we use recursion. □

We can solve the stochastic difference equation (7), and we obtain

$$\underline{z}_t = P_{t0}\underline{z}_0 + \sum_{s=1}^{t-1} P_{ts}G_s\underline{x}_s + \sum_{s=1}^{t} P_{ts}\underline{\epsilon}_s, \tag{16}$$

and thus

$$\underline{y}_t = H_tP_{t0}\underline{z}_0 + \sum_{s=1}^{t-1} H_tP_{ts}G_s\underline{x}_s + \sum_{s=1}^{t} H_tP_{ts}\underline{\epsilon}_s + \underline{\delta}_t. \tag{17}$$

where

$$P_{ts} = \prod_{k=s}^{t} F_k. \tag{18}$$

It follows directly that

$$\mathbf{E}\left(\underline{y}_t \mid \underline{z}_0 \wedge \wedge_{s=1}^{T}\underline{x}_s\right) = H_tP_{t0}\underline{z}_0 + \sum_{s=1}^{t-1} H_tP_{ts}G_s\underline{x}_s, \tag{19}$$

$$\mathbf{V}\left(\underline{y}_t \mid \underline{z}_0 \wedge \wedge_{s=1}^{T}\underline{x}_s\right) = \Theta_t + \sum_{s=1}^{t} H_tP_{ts}\Omega_sP_{ts}'H_t', \tag{20}$$

(21)

while

$$\mathbf{C}\,(\underline{y}_t, \underline{y}_v) = \sum_{s=1}^{\min t,v} H_t P_{ts} \Omega_s P'_{vs} H'_v. \tag{22}$$

References

1. T. W. Andersen, *Some scaling models and estimation procedures in the latent class model*, Probability and Statistics. The Harald Cramér Volume. (Stockholm, Sweden) (U. Grenander, ed.), Almqvist and Wicksell, Stockholm, Sweden, 1959.
2. D. J. Bartholomew, *Latent variable models and factor analysis*, Griffin, London, GB, 1987.
3. R. P. McDonald, *A note on the derivation of the general latent class model*, Psychometrika **27** (1962), 203–206.

Departments of Psychology and Mathematics UCLA, 405 Hilgard Avenue, Los Angeles, CA 90024-1555

E-mail address: deleeuw@laplace.stat.ucla.edu

Department of Psychometrics and Research Methodology, Faculty of Social Sciences, University of Leiden, P.O. Box 9555, 2300 RB Leiden, The Netherlands

E-mail address: byleveld@rulfsw.leidenuniv.nl

SWOV Institute for Road Safety Research, Leidschendam, The Netherlands

E-mail address: bijleveld@swov.nl