Sliced Inverse Regression for High-dimensional Time Series

Claudia Becker and Roland Fried

Fachbereich Statistik,
Universität Dortmund, D-44221 Dortmund, Germany

Abstract: Methods of dimension reduction are very helpful and almost a necessity if we want to analyze high-dimensional time series since otherwise modelling affords many parameters because of interactions at various time-lags. We use a dynamic version of Sliced Inverse Regression (SIR; Li (1991)), which was developed to reduce the dimension of the regressor in regression problems, as an exploratory tool for analyzing multivariate time series. Analyzing each variable individually, we search for those directions, i.e., linear combinations of past and present observations of the other variables which explain most of the variability of the variable considered. This can also provide information on possible nonlinearities. We apply a dynamic version of SIR to multivariate physiological time series observed in intensive care.

Key words: Time series analysis, Nonlinearities, Dimension Reduction

1 Introduction

Modern technical possibilities allow for simultaneous recording of many variables at high sampling frequencies. Possibly there are interactions between the observed variables at various time lags and we have to treat the data as multivariate time series. Often, the data contain strong dynamic structures which are unknown in the beginning.

In intensive care medicine for instance, physicians have to deal with a flood of high-dimensional data describing the patient's state. Vital signs of critically ill patients are recorded in the course of time by clinical information systems (CIS). Altogether, up to 2000 variables are reported. Thus, methods for compressing the information provided by the CIS in a suitable way would be very helpful.

A model-based approach to dimension reduction which has been applied successfully for physiological variables (Gather et al. (2001)) is the application of dynamic factor models. However, specification of a factor model needs a lot of assumptions. Methods such as graphical models for multivariate time series (Dahlhaus (2000)) and a dynamic version of Sliced Inverse Regression (SIR; Li (1991)) can be used at an early stage to explore the dynamic structure between the observed variables.

After briefly reviewing multivariate time series analysis in Section 2, we describe in Section 3 how SIR can be applied to dynamic process data. In Section 4, we give an application to multivariate time series of vital signs measured in intensive care and compare the results with the results obtained by using graphical models.

2 Multivariate Time Series Analysis

Consider a multivariate time series $\{\boldsymbol{x}(t) = (x_1(t), \ldots, x_m(t))' : t = 1, \ldots, T\}$, generated by a stochastic process $\{\boldsymbol{X}(t) : t \in \mathbb{Z}\}$, where $\boldsymbol{X}(t) \in \mathbb{R}^m$, $t \in \mathbb{Z}$. In classical time series analysis, the (time-lagged) relations between the variables are usually assumed to be linear because of computational convenience and since the theory of linear time series models is well established. Vector Autoregressive Moving Average VARMA(p, q) models probably form the most common model class. These models formally resemble multiple regressions. In a VARMA(p, q) model for a stochastic process $\{\boldsymbol{X}(t) : t \in \mathbb{Z}\}$ the influence of past observations and past shocks on the current observation is modelled using weight matrices $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_q \in \mathbb{R}^{m \times m}$:

$$\begin{aligned}\boldsymbol{X}(t) &= \boldsymbol{\Phi}_1 \boldsymbol{X}(t-1) + \ldots + \boldsymbol{\Phi}_p \boldsymbol{X}(t-p) \\ &\quad + \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}(t-1) + \ldots + \boldsymbol{\Theta}_q \boldsymbol{\varepsilon}(t-q) + \boldsymbol{\varepsilon}(t), \quad t \in \mathbb{Z}.\end{aligned}$$

The unobservable shocks $\boldsymbol{\varepsilon}(t), t \in \mathbb{Z}$, are assumed to form a white noise process. In most cases $\boldsymbol{\varepsilon}_t$ is assumed to be normally distributed. Several seemingly distinct VARMA models can describe the same stochastic dependencies, i.e., the VARMA representation of a stochastic process is not necessarily unique (see Reinsel (1997) for details). In practice, often VAR(p) models (with $\boldsymbol{\Theta}_i = \boldsymbol{0}$, $i = 1, \ldots, q$) are used since we get tractable formulae for estimation and predictions and we can guarantee uniqueness of the model in this way.

A basic interest in time series analysis is the prediction of future outcomes. In intensive care, the primary task is the online detection of interesting patterns of change in the data. The large number of unknown parameters needed for modelling multivariate time series can cause problems since the estimates have to be based on a large number of observations. Therefore, methods for dimension reduction could be very helpful. However, for suitable dimension reduction we have to consider the possibly time-lagged relations between the variables. Understanding these relations will help to fit an appropriate model in some lower-dimensional space.

For instance, in Gather et al. (2001) a factor model with four latent factors is constructed for a ten-dimensional time series of vital signs. This factor model is based on an approach suggested by Peña and Box (1987). The observed multivariate time series is assumed to be a linear transform of a lower-dimensional latent factor process which progresses according to a stationary time series model. More exactly, the analysis of Gather et al. assumes the factor process to follow a VAR(2) model with independent components. Another approach to factor analysis is suggested by Molenaar (1985). He assumes the observed time series to be a moving average of a lower dimensional factor process which is modelled using a state-space model. This is a simple example where a-priori information would be useful to specify a suitable model form.

Assuming all interactions to be linear means a severe restriction which certainly is not always appropriate. Linear time series models are not sufficient to describe all real-world phenomena. For this reason a couple of nonlinear models has been suggested (Tong (1990)). A multivariate nonlinear autoregressive model of order p can be written as

$$\boldsymbol{X}(t) = h(\boldsymbol{X}(t-1)', \boldsymbol{X}(t-2)', \ldots, \boldsymbol{X}(t-p)', \boldsymbol{\varepsilon}(t)'), \quad t \in \mathbb{Z}.$$

Here, the transfer function $h : \mathbb{R}^{(p+1)m} \to \mathbb{R}^m$ links the past and the present of the process. It specifies the nonlinear influence of past observations on the present observation. The specification of h is difficult. For $p = 3$ and a 10-dimensional time series we get a 30-dimensional regressor space even if simultaneous interactions are not considered. Hence, finding a suitable dimension-reducing subspace would be very helpful.

In the following we select one variable, denoted by Y, and concentrate on the influence of the other variables, denoted by $\boldsymbol{X}$, on this variable. Then we get a (possibly nonlinear) transfer function model

$$Y(t) = g(\boldsymbol{X}(t), \ldots, \boldsymbol{X}(t-p), Y(t-1), \ldots, Y(t-p), \varepsilon(t)) \tag{1}$$

where the innovations $\varepsilon(t)$ form a white noise process. In this model usually it is assumed that feedback is not present in the system, i.e. influence of Y on the explanatory variables $\boldsymbol{X}$ is not assumed to exist and $\varepsilon(t)$ is assumed to be uncorrelated with $\boldsymbol{X}(t-j), j \geq 0$. Even in case of a single explanatory variable, i.e., univariate $X(t)$, finding a suitable transfer function and estimation of its parameters is difficult (Karlsen et al. (2000)). Hence, appropriate dimension reduction is crucial to get an impression on the type of g. For example, a small number of linear combinations ("directions") of time lagged observations of the influential variables may be sufficient to explain most of the variability in $\{Y(t) : t \in \mathbb{Z}\}$.

3 Sliced Inverse Regression for Dynamic Data

Li (1991) proposes sliced inverse regression (SIR) as a tool for dimension reduction in non-dynamic regression problems. The basic idea is that instead of taking $Y = g(\boldsymbol{X}, \varepsilon)$, $\boldsymbol{X} = (X_1, \ldots, X_d)'$, one assumes that it is sufficient to consider a lower-dimensional space – the so-called central dimension reduction (dr) subspace (Cook (1996)) $\mathcal{B}$ of dimension $r < d$ – such that there exists $f : \mathbb{R}^{r+1} \mapsto \mathbb{R}$ with

$$Y = f(\boldsymbol{\beta}_1'\boldsymbol{X}, \ldots, \boldsymbol{\beta}_r'\boldsymbol{X}, \varepsilon), \text{ where} \tag{2}$$

$$\mathcal{B} = \text{ span}[\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r],$$

and $\mathcal{B}$ is a space of smallest possible dimension r such that (2) is satisfied. Hence, a reduction of the regressor space from d to r dimensions is supposed to be possible. SIR estimates $\mathcal{B}$ using information contained in an estimate of the inverse regression curve $E(\boldsymbol{X}|Y)$.

For given data $(y_i, \boldsymbol{x}_i')'$, $i = 1, \ldots, n$, $\boldsymbol{x}_i \in \mathbb{R}^d$,$y_i \in \mathbb{R}$, SIR is performed in five steps:

1. Standardization:

$$\begin{aligned} \boldsymbol{z}_i &= \widehat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \quad i = 1, \ldots, n, \text{ where} \\ \widehat{\boldsymbol{\Sigma}} &= \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'/n\,, \\ \overline{\boldsymbol{x}} &= \sum_{i=1}^{n} \boldsymbol{x}_i/n\,. \end{aligned}$$

2. Slicing: distribute $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ into H slices S_h, $h = 1, \ldots, H$, according to the order of the respective values of $y_1, \ldots, y_n$; let n_h denote the number of observations in S_h.

3. Slice means: $\widehat{\boldsymbol{m}}_h = \sum_{S_h} \boldsymbol{z}_i/n_h$, $h = 1, \ldots, H$.

4. PCA of slice means: $\widehat{SIR} = \sum_{h=1}^{H} n_h \widehat{\boldsymbol{m}}_h \widehat{\boldsymbol{m}}_h'/n$ with eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_d$ and respective normalized eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_d$.

5. Estimate dr directions: $\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\Sigma}}^{-1/2} \widehat{\boldsymbol{\eta}}_k$, $k = 1, \ldots, r$.

The number H of slices has to be chosen in advance. The reduced dimension r of $\mathcal{B}$ can be estimated simultaneously, cf. Li (1991).

If we want to apply this procedure to multivariate time series, it has to be developed further to take the dynamic structure of the data into account. Let $(Y(t), \boldsymbol{X}(t)')'$ denote the observation of $(Y, \boldsymbol{X}')'$ at time t. In the following, a transfer function model for the measurements of Y as given in (1) is assumed. Becker et al. (2001) suggest to search for a dr subspace of the regressor space using a modified version of equation (2):

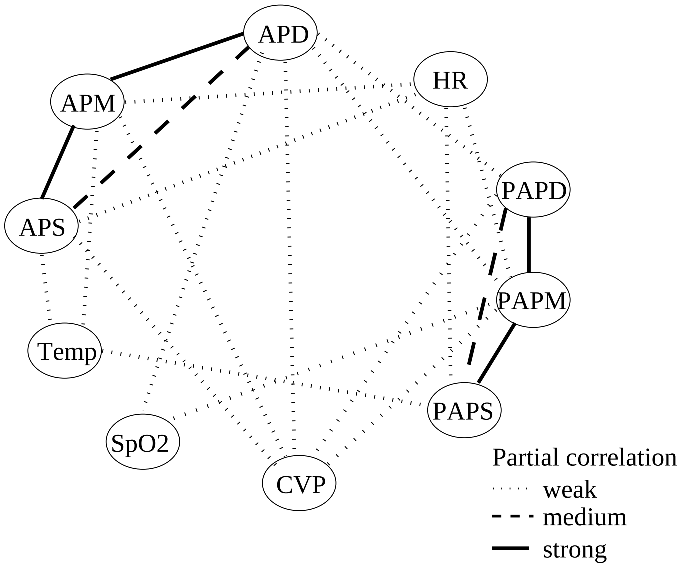
$$Y(t) = f(\boldsymbol{\beta}_1' \widetilde{\boldsymbol{X}}(t), \ldots, \boldsymbol{\beta}_r' \widetilde{\boldsymbol{X}}(t), \varepsilon)\ \forall t,$$

where $\widetilde{\boldsymbol{X}}(t) = (\boldsymbol{X}(t)', Y(t-1), \boldsymbol{X}(t-1)', \ldots, Y(t-p), \boldsymbol{X}(t-p)')'$ if we take p time lags into account. The idea is to bind various lagged measurements of the variables together, forming higher dimensional observations, and then to apply the original SIR method to $(Y(t), \widetilde{\boldsymbol{X}}(t)')'$. The number p of time lags has to be chosen appropriately using preliminary information. Experience shows that for online monitoring data observed in intensive care, $p = 2$ is sufficient (Gather et al. (2000)).

4 Application

In a given data situation, where we do not know beforehand about the association structure between the variables but expect that there is some structure, we can apply dynamic SIR as an explanatory tool. In Becker et al. (2001) a simulation study is

Figure 1: Graphical model for the hemodynamic system of one patient.



performed for several linear and nonlinear VAR and dynamic regression models. It turns out that dr subspaces can be discovered by applying dynamic SIR. To get reliable results, for linear models time series with about 250 observations seem to be sufficient, while for nonlinear models 3000 or even more observations are needed. For a retrospective analysis of multivariate time series observed in intensive care this does not pose a severe restriction since usually several thousands of observations are available.

In the online-monitoring application in intensive care we are concerned with time series describing the hemodynamic system, consisting of arterial and pulmonary artery blood pressures (diastolic, systolic, mean), denoted by APD, APS, APM, PAPD, PAPS, PAPM, central venous pressure (CVP), heart rate (HR), blood temperature (TEMP), and pulsoximetry (SP02). In Gather et al. (2001), graphical models for multivariate time series are used to find relations between the variables. However, such undirected graphical models based on graphical partial correlation analysis accumulate the relations for all time lags and do not mirror causalities for instance.

Using dynamic SIR for multivariate time series gives an impression of the dynamics in the relations between the variables. As an illustrative example, we apply graphical models and the dynamic SIR procedure to a 10-dimensional time series of vital signs observed for a critically ill patient. A total of 360 observation times is included in the analysis. Figure 1 shows the graphical model for all vital signs derived from analyzing the partial correlations.

We concentrate on the relations of APD to the other vital signs further on. Applying dynamic SIR, we find a three-dimensional reduced subspace which is spanned by the directions d_t^1, d_t^2, d_t^3

$$\begin{aligned}
d_t^1 &\approx 1.31APM_t - 0.57APS_t + 0.47APD_{t-1} - 0.43APM_{t-1} \\
d_t^2 &\approx 0.53APM_t + 0.59SPO2_t + 0.45APM_{t-1} - 0.52APS_{t-1} \\
&\quad -0.62PAPD_{t-1} - 0.72TEMP_{t-1} - 0.52SPO2_{t-1} \\
&\quad -0.57PAPD_{t-2} + 0.60PAPM_{t-2} \\
d_t^3 &\approx -0.65PAPD_t + 0.73PAPM_t + 0.79APD_{t-1} - 1.04APM_{t-1} \\
&\quad -0.54PAPM_{t-1} + 0.68PAPS_{t-1}
\end{aligned}$$

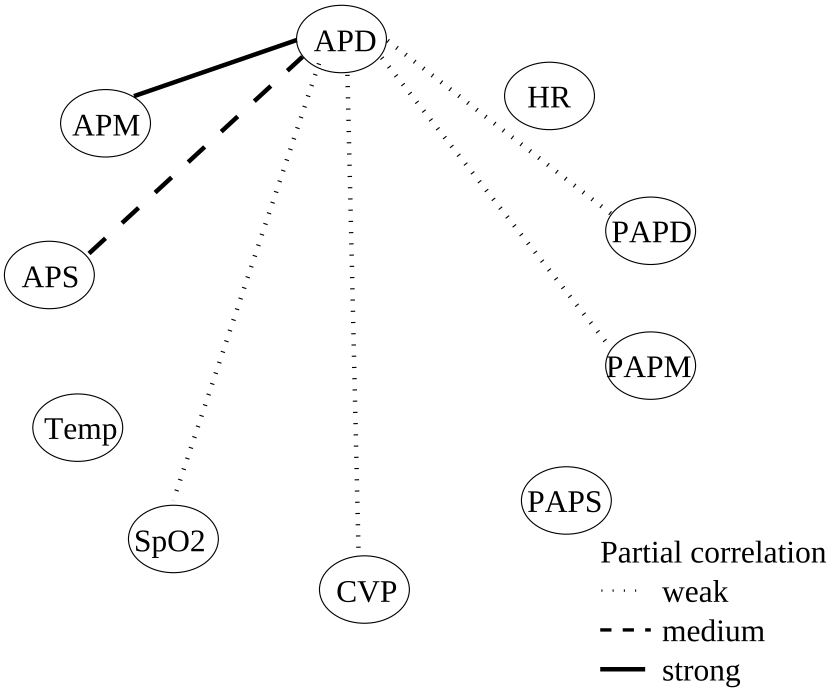
In Figure 2 we provide a graphical illustration of these findings. We connect APD to any of the other vital signs by a line if the simultaneous observation of the latter has a high weight on one of the directions, and draw an arrow from the vital sign towards APD if a past observation of the former has a large weight. The strength of the relation (weak, medium, strong) is interpreted according to the value of the weight with respect to the standard deviation of the respective variable and the number of the corresponding direction. This is due to the fact that the first dr direction corresponds to the largest amount of variability in the relation between response and explanatory variables. As we can see, the results are rather similar to the results obtained by graphical models. Both methods identify APM and APS as the most important variables when we want to explain APD. Using dynamic SIR we additionally get some information about the possible dynamics of the interactions since strong simultaneous relations and also somewhat weaker influences of past values seem to exist. Moreover, we could use dynamic SIR to get an impression of the functional form of the relations using plots of APD against each dr direction.

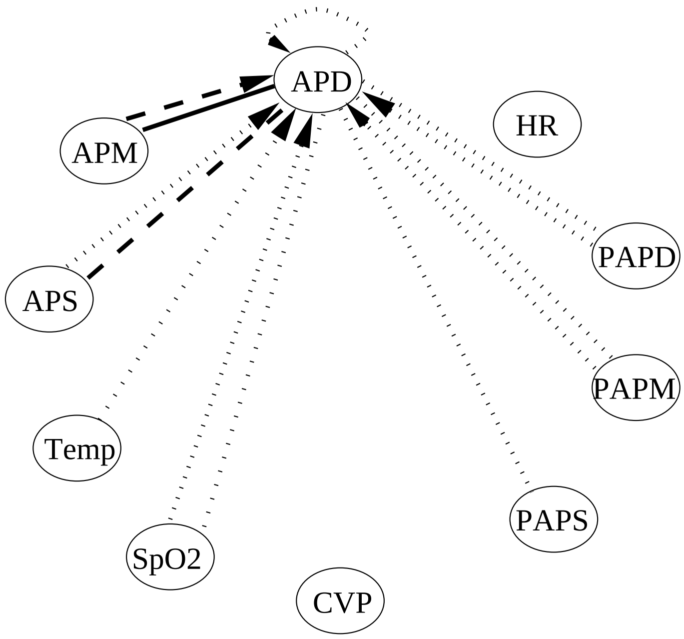
There are some weak relations of APD to other variables identified by graphical modelling but not by dynamic SIR and vice versa. This is due to some fundamental differences between both approaches: Undirected graphical models treat both variables symmetrically analyzing their partial correlations after subtracting the linear influences of the remaining variables, while dynamic SIR uses an asymmetric approach via a regression model under the assumption that one of the variables can be explained using a structural equation model. Wermuth and Lauritzen (1990) discuss some differences between these frameworks in the non-dynamic case.

Finding a three-dimensional reduced subspace points at some nonlinearities (cf. Becker et al. (2001)). In our application, such nonlinearities can be due to therapeutic interventions influencing all variables in different ways. A linear factor model as used in Gather et al. (2001) can therefore only provide a local description of the relations in the steady state. This can be totally satisfactory if detection of relevant changes like intervention effects is the main goal.

Altogether, we consider dynamic SIR-type methods to be promising for dimension reduction of long multivariate time series from dynamic systems. Yet, the theoretical properties of dynamic SIR as well as the interactions found within the physiological time series measured in intensive care should be further validated.

Figure 2: Relations of APD to other vital signs for one patient derived from graphical model (top) and from dynamic SIR (bottom).





Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475) is gratefully acknowledged.

References:

BECKER, C., FRIED, R. and GATHER, U. (2001): Applying Sliced Inverse Regression to Dynamical Data. Preprint, Department of Statistics, University of Dortmund, Germany.

COOK, R. D. (1996): Graphics for Regressions With a Binary Response, *J. Amer. Statist. Assoc., Vol. 91, 983–992.*

DAHLHAUS, R. (2000): Graphical Interaction Models for Multivariate Time Series. *Metrika, 51, 157-172.*

GATHER, U., FRIED, R. and IMHOFF, M. (2000): Online Classification of States in Intensive Care, in Gaul, Opitz, Schader (Eds.): Data Analysis. Scientific Modeling and Practical Application, Springer, Berlin et al., 413–428.

GATHER, U., FRIED, R., LANIUS, V. and IMHOFF, M. (2001): Online Monitoring of High Dimensional Physiological Time Series - A Case-Study. Technical Report 3/2001, SFB 475, Department of Statistics, University of Dortmund, Germany.

KARLSEN, H. A., MYKLEBUST, T. and TJOSTHEIM, D. (2000): Nonparametric Estimation in a Nonlinear Cointegration Type Model. Discussion Paper, SFB 373, Berlin, Germany.

LI, K.-C. (1991): Sliced Inverse Regression for Dimension Reduction (with discussion). *J. Amer. Statist. Assoc., Vol. 86, 316–342.*

MOLENAAR, P. C. M. (1985): A Dynamic Factor Model for the Analysis of Multivariate Time Series. *Psychometrica, Vol. 50, 181–202.*

PEÑA, D. and BOX, G. E. P. (1987): Identifying a Simplifying Structure in Time Series. *J. Americ. Stat. Assoc., Vol. 82, 836–843.*

REINSEL, G. C. (1997): Elements of Multivariate Time Series Analysis. Second Edition. Springer, New York et al.

TONG, H. (1990): Non-linear Time Series. A Dynamical System Approach. Clarendon Press, Oxford.

WERMUTH, N. and LAURITZEN, S. L. (1990): On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Chain Models. *J. R. Statist. Soc. B, Vol. 52, 21–50.*