Modelling High Dimensional Time Series by Generalized Factor Models

Manfred Deistler[1*],
joint work with
B.D.O. Anderson[2**], A. Filler[3*], W. Chen[4**]

***Abstract*— We discuss and analyze generalized linear dynamic factor models. These models have been developed recently and they are used to model high dimensional time series in order to overcome the "curse of dimensionality". The basic idea in factor models is to seperate "comovement" between the variables (caused by a relatively small number of factors) from individual (idiosyncratic) variation. Here factor analysis is considered in a time series context, where concentration of information is performed in the cross-sectional and in the time dimension. The models considered are linear dynamic in nature and stationarity of the processes is assumed. As opposed to the classical case, in the generalized case considered here, a certain form of weak dependence of the noise components is permitted. In the core part of the paper, we are concerned with structure theory, namely with realizing the singular rational spectral density of the latent variables by a linear system. Special emphasis ! is laid on the autoregressive case, which is generic in our setting. These autoregressions may have a singular innovation variance, which may cause multiple solutions for the Yule Walker equations. Finally, identification procedures, using a suitable denoising procedure and estimators suggested by our structure theory, are discussed.**

I. Introduction

The main reasons for (joint or conditional) modelling of multivariate time series are:

- To model the relations between time series
- To extract information common to all single time series
- Forecasting using imformation from the past contained in several time series

As is well known, traditional multivariate time series analysis is plagued by the so called "curse of dimensionality", i.e. by the fact that the dimension of the parameter-space may be large in relation to sample size, T say. For instance in multivariate "unstructured" (i.e. no extra restrictions being imposed) AR modelling the dimension of the parameter space is $N^2p+N(N+1)/2$ (where the first part corresponds to the system parameters and the second part to the noise parameters), with N the cross-sectional dimension (i.e. the number of single time series) and p the maximum lag. Thus the dimension of the parameter space grows with N^2, whereas the number of data points, NT, is linear in N. In a certain sense, this is not surprising in a situation where the relation of every single time series with all other time series has to be modelled. On the other hand, for high dimensional time series and relatively small sample size, alternative modelling approaches ! are needed. As an example think of a macro-economic model of the EU-15 countries, which has to be estimated from quarterly data. In such alternative approaches, the necessary dimension-reduction of parameter-spaces is performed by a-priori information or, data driven, by model selection. Examples for such approaches are:

- "structural modelling", or to use a term from systems engineering, "grey box modelling", where economic or physical a-priori information, e.g. in form of zero-restrictions on parameters, is used.
- Cointegration, where for non-stationary time series the (static) long run equilibrium relation is extracted.
- Graphical time series modelling, where e.g. sparsity of the inverse of the covariance matrix or of the spectral density is used.

Here we consider the approach via factor models. Factor models have been introduced more than a hundred years ago in psychometrics, in order to extract intelligence factors from measured test items ([6], [26]). The basic idea behind factor models is to separate comovement between the variables from individual ("idiosyncratic") fluctuations: Let y_t^N denote the vector of observations, then

$$y_t^N = \hat{y}_t^N + u_t^N \tag{1}$$

where $\hat{y}_t^N$ are the so called *latent variables* and u_t^N is the *idiosyncratic noise*. The comovement in the latent variables can be expressed by modelling these variables as a function of lower dimensional factors.

Factor analysis has been applied to many different fields like psychology, economics, chemistry or signal processing and has undergone a substantial development during the last hundred years, leading to many different kinds of factor models.

The original factor models were not used in a time series context, time series factor models have been proposed during the last 50 years ([5], [11], [15], [22], [23]). Whereas, in a certain sense, in original factor models, the aim was to condense information in cross-section, in time series factor models information in concentrated in two dimensions,

*Research Group Econometrics and System Theory, TU Wien
**Research School of Information Sciences and Engineering, Australian National University
[1]deistler@tuwien.ac.at
[2]brian.anderson@anu.edu.au
[3]alexander.filler@tuwien.ac.at
[4]weitian.chen@anu.edu.au
Support by the FWF (Austrian Science Fund under contracts P17065/N12, P17378 and P20833/N18), the ARC (Australian Research Council under Discovery Project Grant DP0664427) and National ICT Australia Ltd. is gratefully acknowledged.

namely in cross-section and in time dimension.

Another important development was the introduction of generalized factor models, first in a static context by [7] and [8]. There the concept of uncorrelated noise components (strict idiosyncracy) was generalized by allowing for a certain form of weak dependence (weak idiosyncracy).

Combining linear dynamic factor models with the concept of weak dependence lead to the development of *generalized dynamic factor models* (GDFM's), which are the topic of this semiplenary lecture. The theory of GDFM's has been developed during the last decade, in particular in the following papers: [12], [14], [13], [24], [25]. GDFM's are now used, in particular in econometrics, for forecasting and analysis. A number of reserve banks, for instance the European Central Bank, base forecasts on such models.

The rest of the talk is organized as follows: In section 2 the model class of GDFM's is described, section 3 deals with structure theory, in particular with the realization of the singular rational spectral density of the latent variables by linear systems. In section 4 we treat the special case of autoregressive systems, which are generic in our context. Yule Walker equations are analyzed. In section 5 we discuss the use of structure theory for actual data dricen modelling.

II. The Model Class

We commence from equation (1), where $t \in \mathbb{Z}$ denotes discrete time and assume

- $(\hat{y}_t^N)$ and (u_t^N) are wide sense stationary with absolutely summable covariances
- $\mathbb{E}\hat{y}_t^N = \mathbb{E}u_t^N = 0$
- $\mathbb{E}\hat{y}_t^N u_s^{N\prime} = 0 \qquad \forall s, t$

Thus the spectral densities exist, and, using an evident notation, we get:

$$f_y^N(\lambda) = f_{\hat{y}}^N(\lambda) + f_u^N(\lambda), \quad \lambda \in [-\pi, \pi] \tag{2}$$

Motivated by the fact, that we consider cases, where both T and N are large, the asymptotic analysis for assessing the quality of estimation and inference procedures, is performed for $T \to \infty$ and $N \to \infty$. Then the underlying stochastic process is doubly indexed

$$(y_{it} | i \in \mathbb{N}, t \in \mathbb{Z}) \tag{3}$$

where y_{it} is the i-th element of the observation vector y_t^N for $i \leq N$.

Now we list the "core assumptions" defining GDFM's:

Assumption 1: There is an N_0 such that for all $N \geq N_0$, $f_{\hat{y}}^N$ is a rational spectral density with constant rank $q < N$ for all $\lambda \in [-\pi, \pi]$.

Assumption 2: The double indexed sequence (3) corresponds to a nested sequence of models in the sense that (using an evident notation) $\hat{y}_{it}$ and u_{it} do not depend on N for $i \leq N$.

Assumption 3: The rank q of $f_{\hat{y}}^N$ is independent of N ($N \geq$ some N_0).

Assumption 4: The dimension, n say, of a minimal state space realization of a stable and miniphase spectral factor of $f_{\hat{y}}^N$ is independent of N ($N \geq$ some N_0).

Let $\omega_{u,r}^N$ denote the r-th largest eigenvalue of f_u^N.

Assumption 5: (weak dependence): $\omega_{u,1}^N$ is uniformly bounded in λ and N.

Assumption 6: (strong dependence): The first q (i.e. the q largest) eigenvalues of $f_{\hat{y}}^N$ diverge to infinity, for all frequencies, as $N \to \infty$.

From now on, unless the contrary is stated explicitly, we will drop the superindex N.

Let us make a few comments on the core assumptions above. By assumption 1, we can write

$$w f_y = 0 \tag{4}$$

where w is a $(N-q) \times N$ polynomial matrix whose rows form a basis for the left kernel of f_y. Equation (4) is equivalent to

$$w(z)\hat{y}_t = 0 \tag{5}$$

where in this paper we use z for both, for a complex variable and for the backward shift (on the integers $\mathbb{Z}$). After eventually reordering the elements in $\hat{y}_t$, such that in $w = (w_1, w_2)$ the square matrix w_1 is non-singular, we may write (5) as:

$$w_1 \hat{y}_{0,t} = -w_2 \hat{y}_{I,t} \tag{6}$$

where $\hat{y}_t = (\hat{y}_{0,t}', \hat{y}_{I,t}')'$, (if $\det(w_1) \neq 0$, $|z| = 1$) $\hat{y}_{0,t}$ and $\hat{y}_{I,t}$ may be interpreted as outputs and inputs respectively. Equations (6) and (1) constitute an errors-in-variables model (see [18] and [23]), because here also the inputs may be subject to noise.

As opposed to dynamic factor models with strictly idiosyncratic noise, GDFM's are identifiable only for $N \to \infty$, in the sense that the elements of $f_{\hat{y}}^N$ and $\hat{y}_t^N$ are uniquely determined from $\hat{y}_t^N$ for $N \to \infty$ ([8], [14]). For instance, the static factors may be obtained from the y_t^N by the following Hilbert space construction in the space of all square integrable random variables (Note: Once static factors are given, the latent variables are easily obtained in a unique way): Choose an arbitrary basis for the space of all limits of linear combinations of the form

$$\sum_{i=1}^{N} a_i^N y_{i,t} \qquad \text{where } \lim_{N\to\infty} \sum_{i=1}^{N} a_i^{N\,2} = 0$$

Dynamic factors are obtained analogously [14]. Such a procedure leads to "averageing out" the noise from the observations.

A particular way to approximate $f_{\hat{y}}$ is to use dynamic principal components analysis (PCA). Note however that PCA gives dimension reduction in cross section only and does not give a nested sequence of models. In addition

in dynamic PCA problems of non-causality may arise ([13]).

As has been shown in [14], GDFM's can be characterized by properties of f_y.

The following simple example shows the meaning and the consequences of weak and strong dependence respectively: Let

$$y_t = \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \varepsilon_t}_{\hat{y}_t} + \underbrace{\begin{pmatrix} u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix}}_{u_t} \tag{7}$$

where (ε_t) is one-dimensional white noise with $\mathbb{E}\varepsilon_t^2 = 2\pi$, and (u_t) is white noise with $\mathbb{E}u_t u_t' = 2\pi I_N$. Then (2) is of the form

$$f_y = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \dots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + I$$

Then

$$N^{-1}\sum_{i=1}^{N} y_{i,t} = \varepsilon_t + \underbrace{N^{-1}\sum_{i=1}^{N} u_{i,t}}_{\to 0 \text{ for } N\to\infty}$$

III. A STRUCTURE THEORY OF GDFM'S

Our final aim is to identify the linear system underlying the latent variables $\hat{y}_t$ from the observations $y_t,\ t = 1, \dots, T$. In a first step, at least, we are not interested in modelling the noise (u_t). In the structure theory described in this and the next section, an idealized setting is considered, as we commence from the population second moments of the of the latent varibles, rather than from the sample second moments of the observations. Nevertheless this structure theory provides valuable insights as well as an identifictaion algorithm.

Thus we here commence from the rational singular spectral density $f_{\hat{y}}$ to obtain a linear system. We distinguish two major steps, spectral factorization and realization of a tall spectral factor.

A. Spectral Factorization and Wold-Decomposition

The following result has been given in [21], [16].

Theorem 1: Every $N \times N$ rational spectral density $f_{\hat{y}}$ of constant rank q for all $\lambda \in [-\pi, \pi]$ can be factorized as

$$f_{\hat{y}}(\lambda) = (2\pi)^{-1} w(e^{-i\lambda}) w(e^{-i\lambda})^* \tag{8}$$

where $w(z), z \in \mathbb{C}$ is an $N \times q$ real rational matrix of full column rank which has no poles and zeros for $|z| \leq 1$ and where * denotes conjugate transpose.

In addition, it is easy to show that $w(z)$ is unique up to postmultiplication by constant orthogonal matrices.

The spectral factors

$$w(z) = \sum_{j=0}^{\infty} w_j z^j, w_j \in \mathbb{R}^{N\times q} \tag{9}$$

correspond to a causal linear finite dimensional system

$$\hat{y}_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j} \tag{10}$$

where the inputs (ε_t) are white noise with $\mathbb{E}\varepsilon_t\varepsilon_t' = 2\pi I_q$. By (10), $(\hat{y}_t)$ is obtained by a linear dynamic transformation from (ε_t) where $N > q$. For this reason, (ε_t) is called a *dynamic factor*. In addition (ε_t) is a *minimal* dynamic factor, since there is no other dynamic factor, expressing $(\hat{y}_t)$ by a linear dynamic system, of dimension less than q. Clearly the minimal dynamic factors are not unique and they need not to be white noise.

The Smith-McMillan form of $w(z)$ is given by

$$w = udv \tag{11}$$

where u and v are unimodular (i.e. polynomial with constant nonzero determinant) and d is an $N\times q$ rational matrix whose top $q\times q$ block is diagonal with diagonal elements $\frac{n_i}{d_i}$, where d_i and n_i are coprime, monic polynomials and d_{i+1} divides d_i and n_i divides n_{i+1}. All other elements of d are zero. The matrix d is unique for given w and the (finite) zeros of w are the finite zeros of the n_i and the poles of w are the zeros of the d_i. Note that $w(z)$ has no poles and no zeros for $|z| \leq 1$.

For $N > q$, w has no unique left inverse, not even a unique causal left inverse. We define a particular left inverse by

$$w^- = v^{-1}(d'd)^{-1}d'u^{-1} \tag{12}$$

As is easily seen, w^- has no poles and no zeros for $|z| \leq 1$. As is also easily seen, for given w, the input ε_t in (10) is uniquely determined from $\hat{y}_t, \hat{y}_{t-1}, ...$ independently of the particular choice of the causal left inverse by

$$\varepsilon_t = w^-(z)\hat{y}_t \tag{13}$$

Thus (10) corresponds to Wold-Decomposition (see e.g. [17]), and the spaces spanned by $\{\hat{y}_{is} | i = 1, ..., N, s \leq t\}$ and $\{\varepsilon_{is} | i = 1, ..., q, s \leq t\}$ respectively, in the Hilbert space of square integrable random variables, coincide. In addition (ε_t) can be chosen independently of N, from a certain N_0 onwards.

B. State Space Realization

Every rational causal transfer function w can be realized by a state space system, by an ARMA system (or a left matrix fraction description) or by a right matrix fraction description (MFD). Let us start with state space realizations of the form

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{14}$$

$$\hat{y}_t = Hx_t \tag{15}$$

where x_t is the n-dimensional state and $F \in \mathbb{R}^{n\times n}, G \in \mathbb{R}^{n\times q}, H \in \mathbb{R}^{N\times n}$. We assume that the system is minimal, stable, i.e.

$$|\lambda_{max}(F)| < 1 \tag{16}$$

(where $\lambda_{max}(F)$ denotes an eigenvalue of maximum modulus) and mini-phase, i.e. the right side of (18) below has no zeros for $|z| \leq 1$. The transfer function for (14)–(15) is given by

$$w(z) = H(I - Fz)^{-1}G = HG + \sum_{j=1}^{\infty} HF^j Gz^j. \tag{17}$$

As $w(z)$ (which has rank q almost everywhere) has no poles or zeros in $|z| \leq 1$, $w(0) = HG$ has rank q also. Since G has q columns this means $rkG = q$. If (F, G, H) is minimal, the poles of $w(z)$ are the reciprocals of the eigenvalues of F. Also, the transfer function w has a zero for some finite z_0 if and only if the matrix

$$M(z) = \begin{pmatrix} I - Fz & -G \\ H & 0 \end{pmatrix} \tag{18}$$

has rank less than $n+q$ at z_0. Starting with the power series expansion (9), the form (14)–(15) can be obtained by the "Akaike-Kalman procedure" [1] from the equation

$$\underbrace{\begin{pmatrix} \hat{y}_t \\ \hat{y}_{t+1|t} \\ \hat{y}_{t+2|t} \\ \vdots \end{pmatrix}}_{\hat{Y}_t} = \underbrace{\begin{pmatrix} w_0 & w_1 & \cdots \\ w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{\mathcal{H}} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \end{pmatrix} \tag{19}$$

where $\hat{y}_{t+r|t}$ denotes the (best linear least squares) predictor of $\hat{y}_{t+r}$ given the infinite past $\hat{y}_t, \hat{y}_{t-1}, \ldots$. The matrix $\mathcal{H}$ is called the (block) Hankel matrix of the transfer function. As is well known, every basis for the (finite dimensional) space spanned by the (one-dimensional) components of $\hat{Y}_t$ in the Hilbert space of all square integrable random variables, defines a minimal state. Let $S \in \mathbb{R}^{n\times\infty}$ denote the matrix selecting the first components from $\hat{Y}_t$ making up a basis. Note that although S is an infinite matrix, it has only a finite number (n in fact) of nonzero entries, where n is the basis dimension. Then the equations

$$x_t = S\hat{Y}_t \tag{20}$$

$$S \begin{pmatrix} w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = FS\mathcal{H} \tag{21}$$

$$G = S\,(w_0', w_1', \ldots)' \tag{22}$$

$$(w_0, w_1, \ldots) = HS\mathcal{H} \tag{23}$$

(compare [9]) define a (minimal) state space system (14)–(15) in *echelon form*. Every other minimal state is obtained by premultiplying the echelon state by a constant nonsingular matrix.

C. ARMA Representations and Right MFD's

We also consider ARMA systems

$$a(z)\hat{y}_t = b(z)\varepsilon_t \tag{24}$$

where a is $N \times N$ and b is $N \times q$ and where (a, b) are left coprime polynomial matrices. The set of observationally equivalent left coprime ARMA systems is obtained by left multiplication by unimodular matrices (see e.g. [17]). Note that for a coprime ARMA system, the zeros of w are the zeros of b and the poles of w are the zeros of $\det a$. Thus by our assumptions

$$\det a(z) \neq 0, |z| \leq 1 \tag{25}$$

and

$$b(z) \text{ has full rank } q, |z| \leq 1 \tag{26}$$

A right MFD has been used in [13].

D. Static Factors

A *static factor* is a process (z_t) of dimension r say, with $r < N$ such that $(\hat{y}_t)$ can be obtained by a linear static transformation

$$\hat{y}_t = Lz_t \tag{27}$$

Again, we are only interested in *minimal static factors*, i.e. in static factors of minimal dimension. As easily can be seen, r is equal to the rank of the zero lag covariance matrix $\mathbb{E}\hat{y}_t\hat{y}_t'$ and writing

$$\mathbb{E}\hat{y}_t\hat{y}_t' = MM', M \in \mathbb{R}^{N\times r}$$

we may define a minimal static factor with unit variance as

$$z_t = (M'M)^{-1}M'\hat{y}_t \tag{28}$$

Thus

$$\hat{y}_t = Mz_t \tag{29}$$

As is easily seen, minimal static factors are unique up to multiplication by constant nonsingular matrices.

From (15) we see that a minimal state x_t is a static factor. Note that, by controllability, the state covariance matrix $\mathbb{E}x_tx_t'$ has full rank n. However the matrix H may have rank less than n; then there exists a nonsingular matrix T say, such that

$$\hat{y}_t = HTT^{-1}x_t = (H_1, 0)T^{-1}x_t = H_1z_t \tag{30}$$

where H_1 has full column rank, r say, and z_t is a minimal static factor. Clearly, the state is a minimal static factor if and only if $rkH = n$ holds. In general we have $n \geq r \geq q$. A particular static factor can be obtained from the echelon form (20) by selecting the first r linearly independent components of $\hat{y}_t$

$$z_t = S_1\hat{y}_t$$

where

$$S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}, S_1 \in \mathbb{R}^{r\times\infty}, S_2 \in \mathbb{R}^{(n-r)\times\infty}$$

By assumption 4, also r does not depend on N, from a certain N_0 onwards. From (10) and (28) we obtain

$$z_t = (M'M)^{-1}M'w(z)\varepsilon_t = k(z)\varepsilon_t \tag{31}$$

for a transfer function $k(z)$, which has no poles and no zeros for $|z| \leq 1$. As is easy to show, a minimal state space realization for $k(z)$ may be obtained from (F, G, H) as (F, G, C) and vice-versa by choosing

$$C = (M'M)^{-1}M'H \tag{32}$$

$$H = MC \tag{33}$$

Thus, on the one hand, there is a simple linear static relation between the latent variables and (minimal) static factors and on the other hand, both have "essentially" the same dynamics. Since (z_t) has the advantage of being lower dimensional, with dimension r not depending on N, from now on, we emphasize modelling of the static factors.

Clearly, (z_t) can be modelled by an ARMA system. In [27] the autoregression-regression approach [19], [20] has been applied for this purpose, but also other ARMA identification methods may be applied.

The simplest case occurs if $H = n$ holds, i.e. if the minimal state is a static factor and thus is described by the AR(1) system (14,15). This case has been described in detail in [24].

IV. Zeroless Transfer Functions and Autorgressive Systems

A transfer function is called zeroless if all numerator polynomials in the matrix d of its Smith McMillan form (11) are equal to one. Note that the transfer function w, as well as the transfer function k for $r > q$, are zeroless. As will be pointed out in this section, tall rational transfer functions, are relatively to a specified set - generically, i.e. on an open and dense subset - zeroless. A transfer function w (or k) is zeroless if and only if it can be realized by an AR system. The genericity mentioned above is the justification why we restrict ourselves to the AR case. As is well known for square transfer functions AR systems are highly non-generic. One of the main advantages of AR systems vs. ARMA systems is that simple least squares estimation methods like the Yule Walker equations give estimatiors that are both numerically fast on the one hand and asymptotically efficient on the other hand. For the next theorem see [3]

Theorem 2: Consider an $r \times q, r > q$ rational transfer function k with a minimal state space realization (F, G, C) with state dimension n. Then for given n, the transfer function k is zeroless for generic values of (F, G, C).

An intuitive understanding of Theorem 2 can be obtained from the fact, that for $r > q$, $k(z)$ has at least two $q \times q$ minors, which generically have no common zeros.

Note that by definition of a minimal static factor $\mathbb{E}z_t z_t'$ is nonsingular, whereas for $r > q$, its spectral density is singular.

The proof of the next theorem, which relates zeroless transfer functions and AR systems is given in [2], [10]

Theorem 3: Let (z_t) denote a minimal static factor, then the following statements are equivalent:

(i) The spectral factors k of the spectral density f_z of (z_t) satisfying the properties listed in Theorem 1 are zeroless.

(ii) There exists a polynomial left inverse, k^- say, of k, corresponding to (13) and thus the input ε_t in (10) is determined from a finite number of outputs $z_t, z_{t-1}, \ldots, z_{t-L}$ for some L.

(iii) (z_t) is the stationary solution of a stable AR system

$$z_t = e_1 z_{t-1} + \ldots + e_p z_{t-p} + \nu_t \ , \ e_i \in \mathbb{R}^{r \times r} \tag{34}$$

where

$$det(\underbrace{I - e_1 z - \ldots - e_p z^p}_{e(z)}) \neq 0 \text{ for } |z| \leq 1 \tag{35}$$

and ν_t are white noise innovations with $\mathbb{E}\nu_t\nu_t' = \Sigma_\nu$, $rk\ \Sigma_\nu = q$.

If $rk\ \Sigma_\nu = r$ holds, then we have an "ordinary" or regular AR system. For $rk\ \Sigma_\nu < r$, we call the AR system singular. As is well known, in the regular case the matrices

$$\Gamma_m = \begin{pmatrix} \gamma_0 & , & \cdots & , & \gamma_{m-1} \\ \cdots & , & \gamma_0 & , & \cdots \\ {\gamma_{m-1}}^T & , & \cdots & , & \gamma_0 \end{pmatrix}$$

where $\gamma_j = \mathbb{E}z_{t+j}z_t'$ are nonsingular for all $m \in \mathbb{N}$ and $e(z)$ is uniquely determined from the (popualtion) second moments of (z_t) [16]. For singular AR systems the situation is more subtle.

In a certain sense, singular AR systems share some properties with more general ARMA systems. Let $\Sigma_\nu = ff'$, $f \in \mathbb{N}^{r \times q}$ and $rk\ \Sigma_\nu = q$, then $\nu_t = f\varepsilon_t$, where (ε_t) is white noise with $\mathbb{E}\varepsilon_t\varepsilon_t' = 2\pi I_q$ and the system (34) can be written as

$$e(z)z_t = f\varepsilon_t \tag{36}$$

Then we have [2], [10], [4]

Theorem 4:

(i) Every singular AR system (34) satisfying (35) can be written such that $(e(z), f)$ are left coprime

(ii) Let $(e(z), f)$ be left coprime; then the class of all observationally equivalent $(\bar{e}(z), \bar{f})$ where the degree of $\bar{e}(z)$, $\delta(\bar{e}(z))$ say, is given by

$$\left(\bar{e}(z), \bar{f}\right) = u(z)(e(z), f) \tag{37}$$

where the polynomial matrix $u(z)$ satisfies

$$\det u(z) \neq 0, |z| \leq 1; u(0) = I \tag{38}$$

$$\delta(u(z)e(z)) \leq p; \delta(u(z)f) = 0 \tag{39}$$

In addition, $(\bar{e}(z), \bar{f})$ is left coprime if and only if $u(z)$ is unimodular.

(iii) Let $(e(z), f)$ be left coprime and $e(0) = I$; Then $(e(z), f)$ is unique if and only if $rk(e_p, f) = r$ holds.

As is easy to show, for singular AR systems the matrix Γ_{p+1} is singular and the matrix Γ_p may be singular.

The Yule Walker equations [16] are of the form

$$(e_1, \ldots, e_p)\Gamma_p = (\gamma_1, \ldots, \gamma_p) \tag{40}$$

$$\Sigma_\nu = \gamma_0 - (e_1, \ldots, e_p)(\gamma_1, \ldots, \gamma_p)' \tag{41}$$

Here (40) is used to determine $e_1, ..., e_p$ (and thus $e(z)$) from the (in this case population) second moments of the minimal static factors, and then (41) is used to obtain Σ_ν (and thus f). For (40) a solution $(e_1, ..., e_p)$ always exists and if Γ_p is nonsingular then this solution is unique. The variance matrix Σ_ν is uniquely determined from (41) in any case. If Γ_p is singular, then for every row in $(e_1, ..., e_p)$ the solution set is an affine subspace of $\mathbb{R}^{rp}$, namely a particular solution plus the left kernel of Γ_p.

In the set of all solutions $(e_1, ..., e_p)$ of (40) we define the minimum norm solution as follows:

$$(\bar{e}_1, ..., \bar{e}_p) = (\gamma_1, \ldots, \gamma_p)\Gamma_p^{\#} \tag{42}$$

where $\Gamma_p^{\#}$ is the Moore Penrose inverse of Γ_p. The solution set may contain elements corresponding to stable as well as to unstable elements. However we have the following result [10]

Theorem 5: The minimum norm solution (42) always corresponds to a stable system.

V. From Data to Model

In this section we discuss the use of the results obtained from structure theory for actual identification.

As has been stated already, our aim is to obtain a model for the latent variables from the observations. Generally speaking an identification procedure consists of two steps

- Model selection. Here integers, such as the dimension q of (minimal) dynamic factors, the dimension r of the static factors, the state dimension n or the maximum lag p for the AR case have to be estimated.
- Parameter estimation in a specified model. In our context this may be thought of as consisting of three steps:
  - Denoising, i.e. removing the noise from the observations in order to obtain estimates of the latent variables.
  - Estimation of the static factors.
  - Estimation of an AR model for the static factors from the estimators of the static factors.

There are several procedures available for denoising. The following procedure has been proposed in [24]. Commence from the sample variance matrix

$$\hat{\Gamma}_{y,T}^N = T^{-1} \sum_{t=1}^{T} y_t^N y_t^{N\prime}$$

and perform a static PCA. Then, as can be shown, under suitable assumptions, the r largest eigenvalues of $\hat{\Gamma}_{y,T}^N$ for $T, N \to \infty$, tend to infinity, whereas the other eigenvalues converge to finite values. In this way, r can be estimated. Let O_T^N denote the matrix of eigenvectors corresponding to these largest eigenvalues, then we obtain an estimator of z_t by

$$\hat{z}_t = O_T^{N\prime} y_t^N$$

Now $\hat{z}_t$ is used to estimate the AR order p (e.g. using information criteria), as well as the autoregressive coefficients from the Yule Walker equations, now using

$$\hat{\gamma}_j = T^{-1} \sum_{t=1}^{T-j} \hat{z}_{t+j} \hat{z}_t' \quad \text{for} j \geq 0$$

instead of γ_j in (40). The stability properties of a suitably regularized Yule Walker estimator are described in [10]. From (41) an estimator of Σ_ν, $\hat{\Sigma}_\nu$ say, is obtained. By a static PCA on $\hat{\Sigma}_\nu$, q and ε_t can be estimated. In a second denoising step we get a second estimator of z_t, using the estimated AR parameters and the estimator of ε_t.

VI. Conclusions

In this paper we analyzed generalized linear dynamic factor models (GDFM's) from a system theoretic point of view. In a certain sense, identification of GDFM's may be divided into two steps: Denoising of the observations, in order to obtain estimators of the latent variables and, as the second step, identifying a linear system driven by white noise, for the latent variables. The second step, again consists of two parts: First identifying the linear static transformation from the latent variables to the static factors and second estimating the linear system underlying the static factors. The emphasis of our paper is on the corresponding structure theory and on models for the static factors. It is pointed out that AR models are generic in this context, these AR models may have singular innovation variance. A key advantage of parameter estimation in AR models, compared to ARMA or state space models, is that asymptotically efficient estimators are obtained by solving the Yule Wa! lker equations, which are linear. In our case the solutions of these Yule Walker equations are not necessarely unique, which creates some extra complexity.

REFERENCES

[1] H. Akaike. Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control*, 19(6):667–674, 1974.

[2] B. D. O. Anderson and M. Deistler. Generalized linear dynamic factor models a structure theory. *Proceedings of IEEE CDC, Cancun*, 2008.

[3] B. D. O. Anderson and M. Deistler. Properties of zero-free transfer function matrices. *SICE*, 2008.

[4] B. D. O. Anderson, M. Deistler, W. Chen, and A. Filler. Ar models of singular spectral matrices. *Proceedings of IEEE CDC, Shanghai*, 2009.

[5] D. R. Brillinger. *Time Series: Data Analysis and Theory*. Holden Day, San Francisco, 1981.

[6] C. Burt. Experimental tests of general intelligence. *British Journal of Psychology*, 3:94–177, 1909.

[7] G. Chamberlain. Funds, factors and diversification in arbitrage pricing models. *Econometrica*, 51(5):1305–1324, 1983.

[8] G. Chamberlain and M. Rothschild. Arbitrage, factor structure and meanvariance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

[9] M. Deistler. Linear dynamic multiinput/multioutput systems. In D Pena, GC Tiao, and RS Tsay, editors, *A Course in Time Series Analysis*. Wiley, 2001.

[10] M. Deistler, B. D. O. Anderson, A. Filler, Ch. Zinner, and W. Chen. Generalized dynamic factor models-an approach via singular autoregressions. *European Journal of Control*, 2010.

[11] R. F. Engle and M. W. Watson. A one-factor multivariate time series model of metropolitan wage rates. *J. of the American Statistical Association*, 76:774–781, 1981.

[12] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, 65:453–473, 2000.

[13] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830840, 2005.

[14] M. Forni and M. Lippi. The generalized dynamic factor model: representation theory. *Econometric Theory*, 17:1113–1141, 2001.

[15] J. Geweke. The dynamic factor analysis of economic time series models. In D. J. Aigner et al., editor, *Latent Variables in Socio-Economic Models*, pages 365–383. North-Holland Amsterdam, 1977.

[16] E. J. Hannan. *Multiple Time Series*. Wiley, 1970.

[17] E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Wiley, New York, 1988.

[18] R. E. Kalman. System identification from noisy data. In A. R. Bednarek and L. Cesari, editors, *Dynamical Systems II*, pages 331–342. Academic Press, New York, 1982.

[19] D. Q. Mayne, K. J. Astrom, and J. M. C. Clark. A new algorithm for recursive estimation of parameters in controlled ARMA processes. *Automatica*, 20(6):751760, 1984.

[20] D. Poskitt. Identification of echelon canonical forms for vector linear processes using least squares. *The Annals of Statistics*, 20:195–215, 1992.

[21] Yu. A. Rozanov. *Stationary Random Processes*. Holden-Day, San Francisco, 1967.

[22] T. J. Sargent and C. A. Sims. Business cycle modelling without pretending to have too much a priori economic theory. In C. A. Sims, editor, *New Methods in Business Cycle Research*. Minneapolis: Federal Reserve Bank of Minneapolis, 1977.

[23] W. Scherrer and M. Deistler. A structure theory for linear dynamic errors-in-variables models. *SIAM, J. on Control Optim.*, 36(6):2148–2175, 1998.

[24] J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the Americal Statistical Association*, 97:1167–79, 2002.

[25] J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20:147–62, 2002.

[26] L.L. Thurstone. Multiple factor analysis. *Psychological Review*, 38:406–427, 1931.

[27] C. Zinner. *Modeling of High-Dimensional Time Series by Generalized Dynamic Factor Models*. PhD thesis, Vienna University of Technology, 2008.