The Structure of Generalized Linear Dynamic Factor Models

Manfred Deistler, Wolfgang Scherrer and Brian D.O. Anderson

Abstract In this contribution we present a structure theory for generalized linear dynamic factor models. Generalized dynamic factor models have been proposed approximately a decade ago for modeling of high dimensional time series where the cross sectional dimension is of the same order of magnitude as the sample size. In these models the classical assumption for factor models, that the noise components are mutually uncorrelated, is relaxed by allowing for weak dependence. Structure theory turns out to be important for estimation and model selection. The results obtained heavily draw from linear system theory.

The contribution consists of two main parts. In the first part we deal with "denoising", i.e. with getting rid of the noise in the observations. In the second part we deal with constructing linear dynamic systems for the latent variables. Here an important result is the generic zerolessness of the transfer function relating the latent variables and the dynamic factors. This allows for modeling the latent variables by (singular) autoregressions which simplifies estimation.

1 Introduction

Analysis and forecasting of high dimensional time series is an important area in the so called "big data" revolution. High dimensional time series can be found in

Manfred Deistler
Institut für Wirtschaftsmathematik, Technische Universität Wien, e-mail: Manfred.Deistler@tuwien.ac.at

Wolfgang Scherrer
Institut für Wirtschaftsmathematik, Technische Universität Wien, e-mail: Wolfgang.Scherrer@tuwien.ac.at

Brian D.O. Anderson
Research School of Information Sciences and Engineering, Australian National University, Canberra, e-mail: brian.anderson@anu.edu.au

many fields such as econometrics, finance, genetics, environmental research and chemometrics.

The main reasons for joint modeling of high dimensional time series are:

- The analysis of dynamic relations between the time series
- Extraction of factors or features common to all time series (construction of indices)
- The improvement of forecasts, by using the past of many time series

The "traditional" approach to multivariate time series is plagued by the so called "curse of dimensionality". Let T denote the sample size and N denote the cross sectional dimension. If we perform "unstructured" AR modeling, then the dimension of the parameters space is $N^2p+N(N+1)/2$ (where p is the AR order). Thus this dimension increases with N^2, whereas the number of data points, NT, is linear in N. For this reason, for high dimensional time series, model classes with reduced complexity have been used. There are several approaches for this:

- Traditional structural macro-econometric modeling which uses over-identifying a-priori restrictions
- Hierarchical Bayesian modeling, where the prior distribution of the original parameter depends on a few hyper parameters ([15]).
- "Sparse" AR models. A particular class of such sparse models corresponds to so called graphical time series models, where e.g. the zero coefficients correspond to lack of conditional Granger causality ([20]).
- Factor models and (dynamical) principal components analysis. Such models allow for dimension reduction in cross section by using co-movements in the time series.

Here we consider a particular class of factor models, the so called generalized linear dynamic factor models (GDFMs). GDFMs generalize

- Generalized linear static factor models, as introduced in [9, 10]. These models generalize static factor models with strictly idiosyncratic noise, i.e. with uncorrelated noise components, by allowing for "weak dependence" between the noise components.
- Linear dynamic factor models with strictly idiosyncratic noise ([18, 25, 31, 32])

The main features of GDFMs are

- They allow for modeling of dynamics (here in a stationary context)
- Uncorrelatedness of the noise components is generalized to weak dependence
- "Co-movement" of the individual single time series has to be assumed.
- Whereas the observations are ordered in time the results are "permutation invariant" in cross section. Of course, here additional structure, e.g. corresponding to spatial distance, might be imposed
- Strictly speaking, for GDFMs, sequences of model classes, indexed by the cross sectional dimension N are considered.

GDFMs have been developed over the last 13 years, say, and have been successfully applied for instance in macroeconometrics since. The main early references are [21, 24, 33] and further important contributions are [7, 6, 13, 16].

The rest of the paper is organized as follows: In section 2 the model class is described. Section 3 is concerned with denoising. Section 4 deals with the realization of latent variables and static factors by state space and ARMA systems. Section 5 deals with the AR case, which is generic in the "typical" situation considered here. Section 6 treats the Yule-Walker equations for singular AR systems and section 7 outlines the relevance of our results for estimation and model selection.

2 GDFMs: The Model Class

We assume that the N-dimensional observations, y_t^N, are of the form

$$y_t^N = \hat{y}_t^N + u_t^N,\ t \in \mathbb{Z} \tag{1}$$

where $\hat{y}_t^N$ are the *latent variables* and u_t^N is the (weakly dependent) noise.

Throughout we impose the following assumptions:

(A.1) $\mathbf{E}\hat{y}_t^N = \mathbf{E}u_t^N = 0 \in \mathbb{R}^N$ for all $t \in \mathbb{Z}$

(A.2) $(\hat{y}_t^N)$ and (u_t^N) are wide sense stationary[1] with absolutely summable covariances.

(A.3) $\mathbf{E}\hat{y}_t^N (u_s^N)' = 0$ for all $t, s \in \mathbb{Z}$

Thus the spectral densities exist and, using an obvious notation, we obtain for the spectral densities

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

For GDFMs the asymptotic analysis is performed for $T \to \infty$ and $N \to \infty$; thus we consider sequences of GDFMs for $N \to \infty$. In addition we assume throughout that the entries in the vectors are nested, e.g. $\hat{y}_t^{N+1}$ is of the form $((\hat{y}_t^N)', \hat{y}_{N+1,t})'$ where $\hat{y}_{i,t}$ denotes the i-th component of $\hat{y}_t^{N+1}$.

The following assumptions constitute the core of our definition of GDFMs. Here we always assume that N is large enough:

(A.4) (Strong dependence of the latent variables): $f_{\hat{y}}^N$ is a rational spectral density matrix with constant (i.e. for all $\theta \in [-\pi, \pi]$) rank $q < N$; q does not depend on N. The first q eigenvalues of $f_{\hat{y}}^N$ diverge to infinity for all frequencies, as $N \to \infty$.

(A.5) (Weak dependence in the noise): The largest eigenvalue of $f_u^N(\theta)$ is uniformly bounded for all $\theta \in [-\pi, \pi]$ and all N.

Since we assume that the spectral density $f_{\hat{y}}^N$ of the latent variables $\hat{y}_t^N$ is rational, it can be realized by a state space or ARMA system. Here our focus is on state space systems

[1] It should be noted, however, that recently GDFMs for integrated processes have been proposed.

$$x_{t+1}^N = F^N x_t^N + G^N \varepsilon_{t+1}^N \tag{3}$$

$$\hat{y}_t^N = H^N x_t^N \tag{4}$$

where x_t^N is an, n-dimensional, say, (minimal) state and $F^N \in \mathbb{R}^{n\times n}$, $G^N \in \mathbb{R}^{n\times q}$, $H^N \in \mathbb{R}^{N\times n}$ are parameter matrices. We assume that the system is minimal, stable and miniphase and accordingly

$$\hat{y}_t^N = w^N(z)\varepsilon_t^N \tag{5}$$

where $w^N(z) = H^N(I - F^N z)^{-1}G^N$ is a rational, causal and miniphase transfer function and where z is used for the backward shift on the integers $\mathbb{Z}$ as well as for a complex variable. This will be discussed in detail in section 4. In (5), ε_t^N is a minimal *dynamic factor* of dimension q.

The above representation (5) shows that the latent variables are driven by the q dimensional factor process ε_t. Typically $q \ll N$ holds. This generates the co-movement of the latent variables and the observed variables. The assumption (A.5) implies that the noise components are only weakly dependent, which means that the noise can be eliminated by suitable (dynamic) cross sectional averages. This property will be used for "denoising" i.e. for getting $\hat{y}_t^N$ from y_t^N (for $N \to \infty$), as will be discussed in section 3.

In a number of applications GDFMs have been quite successfully applied which shows that the above assumptions in many cases at least provide a reasonable approximation of the true data generating mechanism. There exist (testing) procedures which try to assess whether the given data is compatible with the assumptions. In particular estimation routines (see e.g. [26]) for the number of factor q implicitly test this assumption.

In addition we assume:

(A.6) The spectral density $f_{\hat{y}}^N$ corresponds to a state space system (3),(4) with state dimension n, independent of N.

By (A.6) the McMillan degree of the spectral density $f_{\hat{y}}^N$ is smaller than or equal to $2n$, independent of N. (A.6) is an assumption of bounded complexity dynamics. It is justifiable in a number of applications, e.g. when there is a true underlying system (of finite order) and the number N of sensors is increasing (over sensoring). Recently a theory for the case where q is independent of N, but n is allowed to increase with N has been developed in [23].

As will be shown in section 4, (A.6) implies that the minimal dynamic factor ε_t, the state x_t and F, G in (3),(4) can be chosen independent of N. Furthermore this assumption implies the existence of a *static factor*, z_t say, which may be chosen independent of N and thus

$$\hat{y}_t^N = L^N z_t \tag{6}$$

holds. Note that z_t is called a static factor since the factor loading matrix $L^N \in \mathbb{R}^{N\times r}$ is a constant matrix whereas the corresponding factor loading matrix $w^N(z)$ for ε_t^N in (5) is a transfer function matrix. Let us denote the minimal dimension of such a static factor z_t by r. Then clearly

$$q \leq r$$

holds. In a certain sense the case $q < r$ is of particular interest, since it allows for further complexity reduction.

Given the assumptions, the decomposition (1) of the observed variables into latent variables and noise is unique, asymptotically with N going to infinity. However, the factor loading matrix L^N and the static factor z_t are only unique up to post- respectively pre-multiplication with nonsingular matrices. If we assume that the dynamic factors ε_t^N are the innovations of the latent variables then they are unique up to pre-multiplication by non singular matrices.

3 Denoising

In this section we consider the problem of estimating the factors and/or the latent variables $\hat{y}_{it}$, i.e. we want to eliminate the noise u_{it} from the observations y_{it}. We will concentrate on the estimation of the static factors z_t and corresponding estimates of the latent variables. The dynamic case will be shortly treated in section 3.2.

3.1 Estimation of The Static Factors z_t

Here we consider the static factor model as defined in (6). Since the spectral density f_u^N of (u_t^N) is uniformly bounded by (A.5) it follows that the covariance matrices $\gamma_u^N(0) = \mathbf{E}u_t^N(u_t^N)'$ are also bounded, i.e. there exists a constant, $\overline{\gamma} < \infty$ say, such that

$$\gamma_u^N(0) \leq \overline{\gamma} I_N \text{ for all } N \in \mathbb{N} \tag{7}$$

holds. For the latent variables $\hat{y}_t^N = L^N z_t$ we assume

(A.7) $\gamma_z(0) = \mathbf{E}z_t z_t'$ is positive definite and the minimum eigenvalue of $(L^N)'L^N$ converges to infinity for $N \to \infty$.

This assumption (together with the assumptions above) implies that

$$y_t^N = \hat{y}_t^N + u_t^N = L^N z_t + u_t^N$$

is a (static) generalized factor model as defined in [9, 10] and the denoising can be performed by a simple static PCA analysis as described below.

A sequence of row vectors $(a^N \in \mathbb{R}^{1\times N} \,|\, N \in \mathbb{N})$ with $a^N(a^N)' \to 0$ is called an *averaging sequence,* since by the property (7) it follows that $a^N u_t^N$ converges to zero in mean squares sense. This key idea has been proposed in [9], however with a different name, namely "well diversified portfolio". Therefore such sequences may be used for "denoising" purposes. If $a^N y_t^N$ has a (non zero) limit then this limit has to be an element of the space spanned by the components of the factors z_{it}, $i = 1, \ldots, r$ in the Hilbert space L_2 of the underlying probability space $(\Omega, \mathscr{A}, P)$. A

straightforward generalization is to consider sequences of matrices $(A^N \in \mathbb{R}^{r\times N} \,|\, N \in \mathbb{N})$ with $A^N(A^N)' \to 0 \in \mathbb{R}^{r\times r}$. Clearly

$$A^N y_t^N = A^N L^N z_t + A^N u_t^N \longrightarrow z_t$$

holds if and only of $(A^N L^N) \to I_r$ and thus averaging sequences with this property yield consistent estimates of the static factors z_t. There are a number of possible ways to construct such a denoising sequence A^N.

First let us assume that we know the factor loadings matrix L^N and the covariance matrices $\gamma_z(0)$ and $\gamma_u^N(0)$. The best (in the mean squares sense) linear estimate of z_t given y_t^N is

$$\begin{aligned}\hat{z}_t &= \mathbf{E}(z_t(y_t^N)')\left(\mathbf{E}(y_t^N(y_t^N)')\right)^{-1} y_t^N \\ &= \gamma_z(0)(L^N)'(L^N\gamma_z(0)(L^N)' + \gamma_u^N(0))^{-1} y_t^N \\ &= \underbrace{\left(\gamma_z(0)^{-1} + (L^N)'(\gamma_u^N(0))^{-1}L^N\right)^{-1}(L^N)'(\gamma_u^N(0))^{-1}}_{=:A^N} y_t^N \end{aligned} \tag{8}$$

This estimate is the orthogonal projection of z_t onto the space spanned by the observed variables y_t^N. If we in addition to (7) assume that the noise covariances are bounded from below by

$$\gamma_u^N(0) \geq \underline{\gamma} I_N \text{ for all } N \in \mathbb{N} \text{ with } \underline{\gamma} > 0 \tag{9}$$

then it is easy to prove that the sequence (A^N) defined above is an averaging (matrix) sequence and that $A^N L^N \to I_r$. Thus we get consistent estimates of the factor z_t. In the above formulas one may even replace $\gamma_u^N(0)$ by a rough approximation $\gamma_0 I_N$, $\gamma_0 \geq 0$ and one still gets a consistent estimate

$$\hat{z}_t = \left(\gamma_0\gamma_z^{-1} + (L^N)'L^N\right)^{-1}(L^N)'y_t^N \tag{10}$$

for the factor z_t.

The latent variables are given by $\hat{y}_t^N = L^N z_t$. Therefore an obvious estimate of $\hat{y}_t^N$, for known L^N, is

$$\hat{\hat{y}}_t^N = L^N \hat{z}_t \tag{11}$$

Clearly this estimate is consistent if $\hat{z}_t$ is a consistent estimate of the true factor z_t. To be more precise if $\hat{z}_t \to z_t$ in mean squares sense then $\hat{\hat{y}}_{it} \to \hat{y}_{it}$ in mean squares sense where $\hat{y}_{it}$ and $\hat{\hat{y}}_{it}$ denote the i-th component of $\hat{y}_t^N$ and $\hat{\hat{y}}_t^N$ respectively. If we use the estimate $\hat{z}_t$ defined in (8) then $\hat{\hat{y}}_t^N = L^N\hat{z}_t$ equals the projection of $\hat{y}_t^N$ onto the space spanned by the observed variables y_t^N, i.e. $\hat{\hat{y}}_t^N = L^N\hat{z}_t$ is the best (in a mean squares sense) estimate of $\hat{y}_t^N$.

Of course in practice the above estimates are not operational because the parameters, in particular the loading matrix L^N, are not known. For many (operational) estimates the principal component analysis (PCA) is a starting point. The PCA is a decomposition of the covariance matrix $\gamma_y^N(0) = \mathbf{E}y_t^N(y_t^N)'$ of the form

$$\gamma_y^N(0) = U_1 \Lambda_1 U_1' + U_2 \Lambda_2 U_2'$$

where $U = (U_1, U_2) \in \mathbb{R}^{N \times N}$ is an orthogonal matrix whose columns are the eigenvectors of $\gamma_y^N(0)$ and $\Lambda_1 \in \mathbb{R}^{r \times r}$, $\Lambda_2 \in \mathbb{R}^{(N-r)\times(N-r)}$ are diagonal matrices with diagonal elements equal to the eigenvalues of $\gamma_y^N(0)$. The eigenvalues (and thus the diagonal entries of Λ_1 and Λ_2) are arranged in decreasing order which in particular implies that the minimal diagonal element of Λ_1 is larger than or equal to the maximum diagonal element of Λ_2. Note that U_i and Λ_i depend on the cross sectional dimension N. However, for simplicity we do not use an explicit notation for this dependence. Our assumptions together with basic properties of eigenvalues of symmetric matrices imply

$$\lambda_r(\Lambda_1) = \lambda_r(\gamma_y^N(0)) \geq \lambda_r(L^N \gamma_z(0)(L^N)') \to \infty$$

$$\lambda_1(\Lambda_2) = \lambda_{r+1}(\gamma_y^N(0)) \leq \lambda_1(\gamma_u^N(0)) \leq \overline{\gamma}$$

Here $\lambda_k(M)$ denotes the k-th eigenvalue of a symmetric matrix $M = M'$ where the eigenvalues are ordered as $\lambda_1(M) \geq \lambda_2(M) \geq \cdots$.

An estimate of z_t now is defined as

$$\hat{z}_t = \underbrace{\Lambda_1^{-1/2} U_1'}_{=:A^N} y_t^N \tag{12}$$

where $\Lambda_1^{-1/2}$ is the diagonal matrix defined by $(\Lambda_1^{-1/2})(\Lambda_1^{-1/2}) = \Lambda_1^{-1}$. This estimate, in general, is not consistent for z_t, but gives a consistent estimate for the space spanned by the components of z_t in the following sense. Let $T^N = A^N L^N$ then

$$((T^N)^{-1}\hat{z}_t - z_t) \to 0 \text{ for } N \to \infty$$

and $T^N(T^N)'$ is bounded from below and from above from a certain N_0 onwards, i.e. there exists constants $0 < \underline{c} \leq \overline{c} < \infty$ such that

$$\underline{c} I_r \leq T^N(T^N)' \leq \overline{c} I_r \text{ for all } N \geq N_0$$

First note that $A^N(A^N)' = \Lambda_1^{-1} \to 0$, i.e. (A^N) is an averaging sequence which implies that $A^N u_t^N \to 0$ and thus

$$(\hat{z}_t - T^N z_t) = A^N y_t^N - A^N L^N z_t = A^N u_t^N \to 0$$

Furthermore this implies

$$\mathbf{E} T^N z_t z_t' (T^N)' - \mathbf{E}\hat{z}_t \hat{z}_t' = T^N \gamma_z(0)(T^N)' - I_r \to 0$$

Together these two statements prove the above claim.

The latent variables then are estimated as follows. Note that $\hat{y}_{it}$ is the projection of y_{it} onto the space spanned by the components of the factor z_t since u_{it} is orthogonal to z_t. Correspondingly one may estimate the latent variables by the projection of the

observed variables y_{it} onto the estimated factor $\hat{z}_t$. For the PCA estimate $\hat{z}_t$ defined in (12) we get

$$\hat{\hat{y}}_t^N = \mathbf{E} y_t^N \hat{z}_t' \left(\mathbf{E}\hat{z}_t\hat{z}_t'\right)^{-1} \hat{z}_t = \gamma_y^N(0)(A^N)' \left(A^N \gamma_y^N(0)(A^N)'\right)^{-1} A^N y_t^N = U_1 U_1' y_t^N$$

Since the PCA based estimate $\hat{z}_t$ gives a consistent estimate of the space spanned by z_t one can easily show that the above estimate of the latent variables is consistent too, i.e. $\hat{\hat{y}}_{it} \longrightarrow \hat{y}_{it}$ for $N \to \infty$.

Up to now we have assumed that we have given the covariance matrix $\gamma_y^N(0)$. However, given suitable regularity assumptions which guarantee consistency of the sample covariances one can show that PCA gives consistent estimates of the factors and of the latent variables if one replaces in the above formulas the population moments with sample moments. See e.g. [6, 7, 33].

A slightly different route for the estimation of the factors and the latent variables is taken in [22]. Suppose for the moment that we have given the covariance matrices $\gamma_{\hat{y}}^N(0)$ and $\gamma_u^N(0)$ of the latent variables and the noise respectively. A linear combination $a^N y_t^N = a^N \hat{y}_t^N + a^N u_t^N$ is close to the factor space if the variance of $a^N \hat{y}_t^N$ is large compared to the variance of $a^N u_t^N$. Therefore it makes sense to determine the weights a^N as the solution of the optimization problem

$$\max_{a \in \mathbb{R}^N} a\gamma_{\hat{y}}^N(0)a' \text{ s.t. } a\gamma_u^N(0)a' = 1$$

Iterating this argument one determines r such weighting vectors a_j, $j = 1, \ldots, r$ recursively by

$$a_j = \arg\max_{a \in \mathbb{R}^N} a\gamma_{\hat{y}}^N(0)a' \text{ s.t. } a\gamma_u^N(0)a' = 1 \text{ and } a\gamma_u^N(0)a_i' = 0 \text{ for } 1 \le i < j$$

The solutions a_j are generalized eigenvectors of the pair $(\gamma_{\hat{y}}^N(0), \gamma_u^N(0))$, i.e. they satisfy

$$a_j \gamma_{\hat{y}}^N(0) = \lambda_j a_j \gamma_u^N(0), \;\; j = 1, \ldots, N$$

with the normalization constraints $a_j \gamma_u^N(0) a_j' = 1$ and $a_j \gamma_u^N(0) a_i' = 0$ for $i \neq j$. The λ_j's are the associated generalized eigenvalues. Now let $A^N = ((1+\lambda_1)^{-1/2} a_1', (1+\lambda_2)^{-1/2} a_2', \ldots, (1+\lambda_1)^{-1/2} a_r')'$ and define

$$\hat{z}_t = A^N y_t^N \tag{13}$$

as an estimate for the factors z_t. It is immediate to see that $\mathbf{E}\hat{z}_t\hat{z}_t' = I_r$ and that A^N is an averaging sequence if the noise variances are bounded from below as in (9). One can also show that $\hat{z}_t$ is a consistent estimate for the factor space. The latent variables then are estimated by the projection of the latent variables onto the space spanned by the estimated factors, i.e.

$$\hat{\hat{y}}_t^N = \mathbf{E}\hat{y}_t^N \hat{z}_t' \left(\mathbf{E}\hat{z}_t\hat{z}_t'\right)^{-1} \hat{z}_t = \gamma_{\hat{y}}^N(0)(A^N)' A^N y_t^N \tag{14}$$

This estimation scheme gives the same factor space as the estimate (8) and thus the corresponding estimates for the latent variables coincide, provided that $\gamma_{\hat{y}}^N(0) = L^N\gamma_z(0)(L^N)'$ holds. In order to get a feasible estimate one has first to estimate the covariance matrices $\gamma_{\hat{y}}^N(0)$ and $\gamma_u^N(0)$. The authors [22] obtain such estimates via the dynamic principal component analysis as will be outlined at the end of the next sub section 3.2. Since this procedure incorporates information about the underlying (dynamic) factor model one may hope for an improvement as compared to the (static) PCA scheme.

The above estimates for the factors and the latent variables ignore possible serial correlations which might help to improve the estimates. A possible strategy for doing so was introduced by [16]. Suppose that the factor process (z_t) is an AR(p) process[2] of the form $z_t = a_1 z_{t-1} + \cdots + a_p z_{t-p} + \nu_t$ where (ν_t) is a white noise process and that the noise u_t^N is a (spherical) white noise with $\gamma_u^N(0) = \mathbf{E}u_t^N(u_t^N)' = \gamma_0 I_N$. A state space model for the observed variables y_t^N is as follows

$$\underbrace{\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix}}_{x_t} = \begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ I & & & 0 \\ & \ddots & & \\ & & I & 0 \end{pmatrix} \underbrace{\begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{pmatrix}}_{x_{t-1}} + \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix} \nu_t$$

$$y_t^N = (L^N, 0, \ldots, 0)x_t + u_t^N$$

With a Kalman smoother one then may obtain the best linear estimate of z_t (rsp. the state x_t) given a sample $y_1^N, \ldots, y_T^N$, i.e. the linear projection of z_t onto the space spanned by the components of $y_1^N, \ldots, y_T^N$. If the above model is correct then this dynamic estimate of course improves the "static" estimate (8) since more information is used to construct the estimates for z_t. Furthermore one can show that this estimate is consistent even in the case that the above state space model is (slightly) miss specified, in particular even if the noise u_t^N does not satisfy the very restrictive assumptions above. The drawback of this approach is that the estimates for z_t depend both on future and past values of the observed variables y_t^N and thus this estimate is not suited for prediction purposes. Secondly the quality of the estimate depends on time t, i.e. it makes a difference whether the factors at the start ($t \approx 1$), at the end ($t \approx T$) or in the middle of the sample ($t \approx T/2$) are considered.

In order to obtain a feasible estimation algorithm one first has to estimate the parameters of the above state space model. In [16] the following procedure is proposed to this end. The PCA procedure gives initial estimates $\hat{z}_t$ for the factors z_t, an estimate $\hat{L}^N = U_1 \Lambda^{1/2}$ for the factor loading matrix and $\hat{\gamma}_0 = \frac{1}{N}\mathrm{tr}(U_2 \Lambda_2 U_2')$ is an estimate for the variance of the errors u_{it}. (Note that here U_i, Λ_i are computed from the sample covariance matrix $\hat{\gamma}_y^N(0)$.) Then an AR model is fitted to the estimated factors $\hat{z}_1, \ldots, \hat{z}_T$ yielding estimates for the AR parameters $a_1, \ldots, a_p$ and the covari-

[2] A motivation for the choice of an AR model for the static factors z_t is given in section 5.

ance matrix of the noise v_t. It is shown in [16] that (given suitable assumptions) this procedure gives consistent estimates for the factors (resp. the factor space).

3.2 Estimation of the Dynamic Factors

In order to estimate the dynamic factors ε_t, see equation (5), the concept of averaging sequences is generalized to so called *dynamic averaging sequence*s as follows, see [24]. Let $a^N(z)$ denote a sequence of $1 \times N$ dimensional filters for which

$$\int_{-\pi}^{\pi} a^N(e^{-i\theta})(a^N(e^{-i\theta}))^* d\theta \longrightarrow 0 \text{ for } N \longrightarrow \infty$$

holds. Then by assumption (A.5) the filtered noise $a^N(z)u_t^N$ converges in mean squares to zero and if $a^N(z)y_t^N$ has a limit then this limit is an element of the space spanned by the factor process (ε_t).

The starting point for the estimation of ε_t and of the latent variables $\hat{y}_{it}$ is the dynamic principal component analysis as described in [8, ch. 9]. Let $\lambda_j(\theta)$ and $u_j(\theta)$ denote the j-largest eigenvalue of $f_y^N(\theta)$ and $u_j(\theta)$ be the corresponding (left) eigenvector. This means we have $f_y^N(\theta) = \sum_{j=1}^N \lambda_j(\theta)u_j^*(\theta)u_j(\theta)$. (Again for simplicity we do not explicitly notate the dependence of the eigenvectors and eigenvalues on N.) Note that by assumption (A.4) $\lambda_j(\theta)$ converges to infinity for $1 \leq j \leq q$ and $N \to \infty$ and that $\lambda_j(\theta)$ is bounded for $j > q$. Analogously to the static PCA then estimates of the factor ε_t are defined as

$$\hat{\varepsilon}_t = A^N(z)y_t^N$$

where the (dynamic averaging) filter $A^N(z)$ are computed by

$$A^N = \sum_{k=-\infty}^{\infty} a_k^N z^k \,,\, (a_k^N)^* = \left[(a_{1k}^N)^*, \ldots, (a_{rk}^N)^*\right] \text{ and } a_{sk}^N = \int_{-\pi}^{\pi} \lambda_s^{-1/2}(\theta)u_s(\theta)e^{ik\theta}d\theta \tag{15}$$

Furthermore let

$$\hat{y}_t^N = (B^N(z))^* B^N(z)y_t^N \tag{16}$$

where

$$B^N(z) = \sum_{k=-\infty}^{\infty} b_k^N z^k \,,\, (b_k^N)^* = \left[(b_{1k}^N)^*, \ldots, (b_{rk}^N)^*\right] \text{ and } b_{sk}^N = \int_{-\pi}^{\pi} u_s(\theta)e^{ik\theta}d\theta \tag{17}$$

It is proven in [24] that these estimates are consistent for the factors and for the latent variables. In the paper [21] a feasible estimation scheme is constructed based on the above ideas. First the population spectral density $f_y^N(\theta)$ is replaced by a consistent estimate, $\hat{f}_y^N(\theta)$ say. From the eigenvalue decomposition of this estimated spectral density then estimates for the filter A^N and B^N are computed as in (15) and (17). In

order to get operational estimates the infinite filters furthermore are approximated by finite order filters of the form

$$\hat{A}^N = \sum_{k=-M}^{M} \hat{a}_k^N z^k \text{ and } \hat{B}^N = \sum_{k=-M}^{M} \hat{b}_k^N z^k \tag{18}$$

where the order M converges to infinity with increasing sample size. Note that the above filters (and the estimated filters) are in general two sided. This holds in particular for the filters related to the latent variables. Therefore these estimates for the latent variables are not suited for prediction purposes.

At the end of this sub section we shortly explain the estimation of the covariance matrix of the latent variables and of the noise which is used in one of the denoising schemes explained in section 3.1, see equations (13) and (14). The spectral density of the estimated latent variables, see (16), is equal to $\sum_{j=1}^{r} \lambda_j(\theta) u_j^*(\theta) u_j(\theta)$ and thus the covariance matrices are estimated through

$$\hat{\gamma}_{\hat{y}}^N(0) = \int_{-\pi}^{\pi} \left[\sum_{j=1}^{r} \lambda_j(\theta) u_j^*(\theta) u_j(\theta) \right] d\theta \text{ and } \hat{\gamma}_u^N(0) = \int_{-\pi}^{\pi} \left[\sum_{j=r+1}^{N} \lambda_j(\theta) u_j^*(\theta) u_j(\theta) \right] d\theta$$

Of course starting from a sample the eigenvalues and eigenvalues are computed from an eigenvalue decomposition of an (consistent) estimate of the spectral density f_y^N.

4 Structure Theory for the Latent Process

In this and the next section we deal with structure theory for the latent process. In this idealized setting we assume that the observations have been completely denoised and we commence from the population spectral density $f_{\hat{y}}^N$ of the latent variables $\hat{y}_t^N$, We proceed in three steps

- Spectral factorization
- Construction of a minimal static factor z_t
- Construction of a model for the dynamics of (z_t) with the dynamic factors ε_t as innovations

4.1 The Spectral Factorization and the Wold Representation of the Latent Process

The following result is well known ([27, 30]. Here we omit the superscript N if no confusion can arise.

Theorem 1. *Every ($N \times N$-dimensional) rational spectral density $f_{\hat{y}}$ of constant rank q can be factorized as*

$$f_{\hat{y}}(\lambda) = w(e^{-i\lambda})w(e^{-i\lambda})^* \tag{19}$$

where

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

is rational, analytic in $|z| \leq 1$ *and has rank q for all* $|z| \leq 1$. *Here* * *denotes the conjugate transpose. In addition such a w is unique up to post multiplication by constant orthogonal matrices.*

A transfer function matrix w with the above properties is called a stable, miniphase factor of $f_{\hat{y}}$. There exist q dimensional white noise (ε_t) with $\mathbf{E}\varepsilon_t\varepsilon_t' = 2\pi I_q$ such that

$$\hat{y}_t = w(z)\varepsilon_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j} \tag{20}$$

Now let

$$w = ulv \tag{21}$$

where u and v are unimodular polynomial matrices and where l is an $N \times q$-dimensional, quasi diagonal rational matrix whose (i,i)-th element is of the form $n_i(z)/d_i(z)$ where n_i, d_i are coprime and monic polynomials and n_i divides n_{i+1} and d_{i+1} divides d_i. Then (21) is called the Smith-McMillan form of $w(z)$, [28, ch. 2]. As easily seen a particular left inverse of w is

$$w^- = v^{-1}(l'l)^{-1}l'u^{-1} \tag{22}$$

where w^- is rational and has no poles and zeros for $|z| \leq 1$. This implies that (20) is already a Wold representation and the innovations ε_t in (20) are minimal dynamic factors.

A transfer function matrix w is called *zeroless* if all numerator polynomials of the diagonal elements of l are equal to one. In this case w^- is a polynomial matrix.

4.2 Minimal Static Factor

From (4) it is clear that x_t^N is a static factor, which is not necessarily minimal, as discussed below. Then

$$\mathrm{rk}\underbrace{\mathbf{E}\hat{y}_t^N(\hat{y}_t^N)'}_{=\gamma_{\hat{y}}^N(0)} = \mathrm{rk}H^N\mathbf{E}x_t^N(x_t^N)'(H^N) \leq n$$

and therefore, by (A.6), $\mathrm{rk}\gamma_{\hat{y}}^N(0)$ is bounded by n, independent of N. This implies that the rank of $\gamma_{\hat{y}}^N(0)$ is constant from a certain N_0 onwards. Let r denote this rank, i.e. let $\mathrm{rk}(\gamma_{\hat{y}}^N(0)) = r$ for all $N \geq N_0$. Furthermore, we see that also the minimal

static factor z_t can be chosen independent of N. Take for instance a z_t consisting of the first basis elements in $\hat{y}_t^N$ spanning the space generated by the components of $\hat{y}_t^N$ in the Hilbert space L_2. Minimal static factors are unique up to premultiplication by constant nonsingular matrices. They may be obtained via a factorization

$$\gamma_{\hat{y}}^N(0) = L^N (L^N)', \; L^N \in \mathbb{R}^{N\times r}, \; \mathrm{rk}(L^N) = r$$

of the covariance matrix $\gamma_{\hat{y}}^N(0)$ as

$$z_t = \underbrace{\left((L^N)' L^N\right)^{-1} (L^N)'}_{L^{N-}} \hat{y}_t^N \tag{23}$$

Clearly z_t has a rational spectral density of the form

$$f_z(\theta) = L^{N-} f_{\hat{y}}^N (L^{N-})'$$

and, for $q < r$, f_z is singular. Note that

$$z_t = \underbrace{L^{N-} w^N(z)}_{=k^N(z)} \varepsilon_t^N = k^N(z)\varepsilon_t^N \tag{24}$$

is a Wold representation of (z_t), because $w^{N-} L^N z_t = \varepsilon_t^N$ and $w^{N-} L^N$ is a causal transfer function. Thus $k^N(z)$ is a causal miniphase spectral factor of f_z. Since such a spectral factor is unique up to post multiplication by non singular matrices it follows that we may chose ε_t and k independent of N.

Remark 1. As shown above (A.6) implies that the rank of $\gamma_{\hat{y}}^N(0)$ is bounded. Vice versa it is easy to see that a bound on the rank of $\gamma_{\hat{y}}^N(0)$ implies (A.6) under our assumptions (A.1)-(A.5).

4.3 State Space Realizations for the Latent Process and the Minimal Static Factors

The problem of realization is to find a system for a given transfer function. Here our focus is on state space systems, for instance for a minimal static factor $z_t = k(z)\varepsilon_t$ we have

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{25}$$

$$z_t = Cx_t \tag{26}$$

where x_t is an, n-dimensional, say, (minimal) state and $F \in \mathbb{R}^{n\times n}$, $G \in \mathbb{R}^{n\times q}$, $C \in \mathbb{R}^{r\times n}$ are parameter matrices. We assume minimality (i.e. the system (25), (26) is controllable and observable), stability, i.e.

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

where $\lambda_{\max}(F)$ is an eigenvalue of F of maximum modulus and the miniphase condition i.e. that

$$M(z) = \begin{pmatrix} I - Fz & -G \\ C & 0 \end{pmatrix} \tag{27}$$

has rank $n+q$ for $|z| \leq 1$. Note that the zeros of $M(z)$ are the zeros of $k(z) = C(I - Fz)^{-1}G$ as defined via its Smith McMillan form (see [29]) . For given k a unique state space realization may be obtained e.g. by echelon forms, see [12].

From this state space realization we immediately get a state space realization for the latent variables $\hat{y}_t^N = L^N z_t$

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{28}$$

$$\hat{y}_t^N = \underbrace{L^N C}_{H^N} x_t = H^N x_t \tag{29}$$

This state space system is minimal, stable and miniphase. We also see that due to our assumptions the state x_t, the innovations ε_t and the matrices F and G may be chosen independent of N, compare (3) and (4). Only the matrix H^N depends on N, however note that the H^N's are nested i.e. the first N rows of H^{N+1} coincide with H^N.

The rational transfer function $k(z)$ in (24) may be written as a left matrix fraction $k(z) = a^{-1}(z)b(z)$ ([28, ch. 2]) giving rise to an ARMA realization

$$a(z)z_t = b(z)\varepsilon_t$$

where we w.r.o.g. assume that $(a(z), b(z))$ are left coprime, stable and miniphase.

Alternatively, we may write a right matrix fraction

$$k(z) = d(z)c^{-1}(z)$$

see [22]. This gives rise to a factor representation of $\hat{y}_t^N$ of the form

$$\hat{y}_t^N = D^N(z)\mu_t, \; D^N(z) = L^N d(z) \text{ and } c(z)\mu_t = \varepsilon_t$$

i.e. with a minimal dynamic factor μ_t which is an AR process and a factor loading matrix $D^N(z) = L^N d(z)$ which is a finite impulse response filter.

5 Zeroless Transfer Functions and Singular AR Systems

In this section we will show that for the case $r > q$, generically, the static factor can be modeled by an AR system. This is important because estimation of AR systems is much easier compared to the ARMA case.

Let us repeat, that a transfer function is called zeroless if all numerator polynomials in l in its Smith McMillan form are equal to one.

Lemma 1. *The transfer function $w(z)$ is zeroless if and only if $k(z)$ is zeroless.*

Proof. As is shown in [29], $w(z)$ is zeroless if and only if the associated $M(z)$ in (27) has rank $n+q$ for all $z \in \mathbb{C}$. An analogous statement holds for $k(z)$ and thus

$$\begin{pmatrix} I-Fz & -G \\ H^N & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & L^N \end{pmatrix} \begin{pmatrix} I-Fz & -G \\ C & 0 \end{pmatrix} \tag{30}$$

together with $\mathrm{rk}(L^N) = r$ yields the desired result. □

The proof of the Theorem below is given in [2] and [5]. Note that a property is said to hold generically on a given set, if it holds on an open and dense subset.

Theorem 2. *Consider the set of all minimal state space systems (F,G,C) for given state dimension n, output dimension r and input dimension q, where $r > q$ holds. Then the corresponding transfer functions are zeroless for generic values of $(F,G,C) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times q} \times \mathbb{R}^{r\times n}$.*

The theorem states, that, in a certain setting, tall transfer functions are generically zeroless. For simple special cases this is immediate. Consider e.g. the MA(1) case for $r = 2, q = 1$:

$$\begin{aligned} z_{1t} &= b_{11}\varepsilon_t + b_{12}\varepsilon_{t-1} \\ z_{2t} &= b_{21}\varepsilon_t + b_{22}\varepsilon_{t-1} \end{aligned}$$

then the system is zeroless whenever

$$b_{11}b_{22} - b_{12}b_{21} \neq 0$$

holds.

We have (see [1])

Theorem 3. *Let k denote a stable miniphase factor of the spectral density f_z of (z_t). (Remember that such a factor is unique up to post multiplication with non singular matrices.) The following statements are equivalent*

1. *k is zeroless.*
2. *There exist a polynomial left inverse k^- for k.*
3. *(z_t) is an AR process, i.e.*

$$z_t = a_1 z_{t-1} + \cdots + a_p z_{t-p} + \nu_t \tag{31}$$

where

$$\det(I - a_1 z - \cdots - a_p z^p) \neq 0 \text{ for } |z| \leq 1 \tag{32}$$

and (ν_t) is white noise with $\gamma_\nu(0) = \mathbf{E}\nu_t\nu_t'$ and $\mathrm{rk}(\gamma_\nu(0)) = q$.

Remark 2. We call an AR system (31) *regular* if $\mathrm{rk}(\gamma_v(0)) = r$ and *singular* if $\mathrm{rk}(\gamma_v(0)) < r$ holds. This means that in the case $q < r$ we have to deal with singular AR systems. Clearly we may write (31) as

$$a(z)z_t = b\varepsilon_t \tag{33}$$

where $\gamma_\varepsilon(0) = \mathbf{E}\varepsilon_t\varepsilon_t' = I_q$ and $\mathrm{rk}(b) = q$, and where $\gamma_v(0) = bb'$ holds.

Remark 3. Two AR systems $(a(z), b)$ and $(\bar{a}(z), \bar{b})$ are called *observationally equivalent* if their transfer functions $a^{-1}(z)b$ and $\bar{a}^{-1}(z)\bar{b}$ are the same up to post multiplication by an orthogonal matrix. Let $\delta(a(z))$ denote the degree of $a(z)$. We assume throughout that the specified degree of $a(z)$ is given by p, i.e. $\delta(a(z)) \leq p$, the stability condition (32), and $a(0) = I$. Note that our notion of observational equivalence is based on the stationary solution

$$z_t = a^{-1}(z)b\varepsilon_t$$

and does not take into account other solutions, compare [14].

Proof. Let

$$k = u \begin{bmatrix} l \\ 0_{r-q\times q} \end{bmatrix} v$$

denote the Smith McMillan form of k, where u, v are two unimodular matrices, l is a $q \times q$ diagonal matrix and $0_{r-q\times q}$ denotes a zero matrix of suitable dimension. The i-th diagonal entry of l is $n_i(z)/d_i(z)$ where n_i, d_i are coprime and monic polynomials. The spectral factor k is zeroless if and only if $n_i = 1$ holds for $i = 1, \ldots, q$.

Clearly

$$k^- = v^{-1} \begin{bmatrix} l^{-1} & 0 \end{bmatrix} u^{-1}$$

is a left inverse of k and k^- is polynomial if and only if k is zeroless. Note that this left inverse corresponds to (22). This proves 1) $\Rightarrow$ 2). Conversely, if there exist a polynomial left inverse, k^- say, then $k^-k = I_q$ implies $k^-u_1l = v^{-1}$ where $u = [u_1, u_2]$ has been partitioned conformingly. This implies $n_i(z) \neq 0$ for all $z \in \mathbb{C}$ and thus k must be zeroless.

Next we define

$$\bar{k} = u \begin{bmatrix} l & 0_{q\times r-q} \\ 0_{r-q\times q} & I_q \end{bmatrix} \begin{bmatrix} v & 0_{q\times r-q} \\ 0_{r-q\times q} & I_{r-q} \end{bmatrix}$$

which gives

$$z_t = k(z)\varepsilon_t = \bar{k}(z) \begin{bmatrix} \varepsilon_t \\ 0_{r-q} \end{bmatrix}$$

and thus

$$\begin{bmatrix} \varepsilon_t \\ 0_{r-q} \end{bmatrix} = \bar{k}^{-1}(z)z_t = \begin{bmatrix} v^{-1}(z) & 0_{q\times r-q} \\ 0_{r-q\times q} & I_{r-q} \end{bmatrix} \begin{bmatrix} l^{-1}(z) & 0_{q\times r-q} \\ 0_{r-q\times q} & I_q \end{bmatrix} u^{-1}(z)z_t$$

If k is zeroless then $\bar{k}^{-1}$ is polynomial and premultiplying the above equation with $\bar{k}(0)$ yields a *stable* AR system

$$\underbrace{\bar{k}(0)\bar{k}(z)^{-1}}_{a(z)} z_t = a(z)z_t = \underbrace{\bar{k}(0)\begin{bmatrix}\varepsilon_t\\ 0_{r-q}\end{bmatrix}}_{v_t} = v_t$$

with $a(0) = I_r$. Thus 1) $\Rightarrow$ 3).

To prove the converse first note that (see e.g. [2]) k is zeroless if and only if for every left coprime MFD $\bar{a}^{-1}\bar{b} = k$, the polynomial matrix $\bar{b}$ is zeroless. If we start with an AR system (33) then for $q < r$, the pair $(a(z), b)$ is not necessarily left coprime. By $a(0) = I$, the greatest common divisor $r(z)$ of $(a(z), b)$ may be chosen with $r(0) = I$ and thus extracting such a common divisor (see [28, 29]) we obtain a left coprime system $(\tilde{a}(z), b)$ (where b remains the same) and thus, as $\mathrm{rk}(b) = q$, $k = a(z)^{-1}b = \tilde{a}^{-1}(z)b$ is zeroless. □

The next Theorem (see [3]) states that for $(\bar{a}(z), b)$ non necessarily left coprime, there is an observationally equivalent left coprime pair $(a(z), b)$ satisfying the same degree restriction p.

Theorem 4. *Every AR system $(\bar{a}(z), b)$ with $\delta(\bar{a}(z)) \leq p$ can be transformed to an observationally equivalent AR system $(a(z), b)$ such that $\delta(a(z)) \leq p$ and $(a(z), b)$ are left coprime.*

6 The Yule Walker Equations

As is well known and has been stated before, in the usual (regular) case, estimation of AR systems is much easier than estimation of ARMA systems or state space systems, because AR systems can be estimated e.g. by the Yule-Walker equations, which are linear in the AR parameters, whereas in the ARMA (or state space) case usually numerical optimization procedures are applied. This also holds for the singular case and shows the importance of Theorem 2.

The Yule Walker equations are of the form

$$(a_1, \ldots, a_p)\Gamma_p = (\gamma_z(1), \ldots, \gamma_z(p)) \tag{34}$$

$$\gamma_v(0) = \gamma_z(0) - (a_1, \ldots, a_p)(\gamma_z(1), \ldots, \gamma_z(p))' \tag{35}$$

where $\gamma_z(j) = \mathbf{E} z_{t+j} z_t'$ and

$$\Gamma_m = (\gamma_z(j-i))_{i,j=1,\ldots,m}$$

are the population moments of (z_t). From an estimator $\hat{z}_t$ of z_t, these second moments can be estimated and yield a Yule Walker estimator of $(a_1, \ldots, a_p)$ and $\gamma_v(0)$ via (34) and (35).

As is well known and easy to see, Γ_m for a regular AR process is non singular for all m. On the other hand for a singular AR process, premultiplying (31) by a vector $a \neq 0$ such that $a\gamma_v(0) = 0$ yields a dependence relation between the components in $(z_t', z_{t-1}', \ldots, z_{t-p}')$ and thus Γ_{p+1} is singular. However, the matrix Γ_p may be singular or non singular. Now (33) may be written in companion form as

$$\underbrace{\begin{pmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t-p+1} \end{pmatrix}}_{x_t} = \underbrace{\begin{pmatrix} a_1 & \cdots & a_{p-1} & a_p \\ I & & & 0 \\ & \ddots & & \\ & & I & 0 \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} z_{t-1} \\ z_{t-2} \\ \vdots \\ z_{t-p} \end{pmatrix}}_{x_{t-1}} + \underbrace{\begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{G} \varepsilon_t$$

$$z_t = (I, 0, \ldots, 0) x_t$$

As is well known and easy to see,

$$\Gamma_p = \mathbf{E} x_t x_t'$$

is non singular if and only if (F, G) is controllable. For this case, the Yule Walker equations (34) and (35) have a unique solution. As shown in [4] (F, G) is generically controllable in the parameter space. However, in this context, the notion of genericity has to be used with care, as it depends on the choice of p. If Γ_p is singular, of course the solution of (34) are not unique, where $\gamma_v(0)$ remains unique. Because of the linearity the solution set of (34) has an affine structure in the sense that every row of $(a_1, \ldots, a_p)$ is of the form one particular solution plus the left kernel of Γ_p. Since $\gamma_z(j)$, $j > p$ are uniquely determined by $\gamma_z(k)$, $k = 0, \ldots, p$ the solution set of (34) is the set of all observationally equivalent AR systems (without imposing the stability condition (32)). The structure of the solution set has been discussed in [11]. In case of singular Γ_p, uniqueness in (34) may be achieved by taking the minimum norm solution (see [13, 19]) or by describing column degrees in a(z) (see [14]).

7 Estimation and Model Selection

Structure theory as described in the previous sections is important for understanding data driven modeling for GDFMs. In general terms, here, data driven modeling consists of two main parts, parameter estimation and model selection, where the latter is performed by estimation of integers such as r or q. We do not treat estimation (and in particular properties of estimators) in detail here. Let us only mention that the estimation procedure we have in mind (and we had also in mind in [13]) is of the following form:

1. Estimation of the (minimal) dimension r of the static factors z_t as well as the static factor itself using a PCA on $\gamma_y^N(0)$, as described in section 3.1.

2. Estimate the maximum lag p in the AR system (33) from the given $\hat{z}_t$ using an information criterion and estimate $a_1, \ldots, a_p$ from the Yule Walker equations as described in section 6.
   If the estimate of Γ_p is "close to being singular", a truncation procedure as described in [13, 19] or a specification of columns degrees in $a(z)$ as in [14] may be used.
3. Using the Yule-Walker estimate for the covariance of ν_t in (31) a PCA is performed to estimate the dimension of the minimal dynamic factors, as well as ε_t itself and thus of b in (33).

Remark 4. Under a number of additional assumptions, for instance assumptions guaranteeing that the sample second moments converge to their population counter parts, the procedure sketched above can be shown to be consistent. There is a number of alternative procedures available (see e.g. [17, 33]) and in addition our procedure may be improved, for instance by iterations.

Remark 5. Our focus is on data driven modeling of the latent variables. However, sometimes also noise models, e.g. univariate AR systems for the components of u_t, are used.

8 Summary

Forecasting and analysis of high dimensional time series is part of the "big data revolution" and it is important in many areas of application. Generalized linear dynamic factor models which have been proposed slightly more than a decade ago [21, 24, 33] are an important tool for this task. In our contribution we present a structure theory for this model class. In particular we consider denoising and realization, the latter in the sense of finding a system representation for the spectral density of the latent variables. The importance of AR modeling is emphasized. Yule Walker equations and their solutions are discussed. Finally an estimation procedure making use of the structure theory is described.

References

1. B. D. O. Anderson and M. Deistler. Generalized linear dynamic factor models - a structure theory. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 1980–1985, 2008.
2. B. D. O. Anderson and M. Deistler. Properties of Zero-free Transfer Function Matrices. *SICE Journal of Control, Measurement, and System Integration*, 1(4):284–292, July 2008.
3. B. D. O. Anderson, M. Deistler, W. Chen, and A. Filler. Autoregressive models of singular spectral matrices. *Automatica*, 48(11):2843 – 2849, 2012.
4. B. D. O. Anderson, M. Deistler, E. Felsenstein, B. Funovits, P. Zadrozny, M. Eichler, W. Chen, and M. Zamani. Identifiability of regular and singular multivariate autoregressive models from

mixed frequency data. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 184–189, 2012.
5. B. D. O. Anderson, M. Deistler, and A. Filler. Generic properties of system zeros. Manuscript, Research Group Econometrics and System Theory, TU Vienna, 2013.
6. J. Bai. Inferential Theory for Factor Models of Large Dimension. *Econometrica*, 71(1):135–171, Jan. 2003.
7. J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.
8. D. R. Brillinger. *Time Series Data Analysis and Theory*. Holden-Day, San-Francisco, London, 1981.
9. G. Chamberlain. Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51(5):1305–1323, Sept. 1983.
10. G. Chamberlain and M. Rothschild. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, Sept. 1983.
11. W. Chen, B. D. Anderson, M. Deistler, and A. Filler. Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis*, 32(5):531–538, 2011.
12. M. Deistler, B. D. O. Anderson, A. Filler, and W. Chen. Modelling high dimensional time series by generalized factor models. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2010 - July, 2010 - Budapest, Hungary)*, pages 323–329, Piscataway USA, 2010. Institute of Electrical and Electronics Engineers (IEEE Inc).
13. M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen. Generalized Linear Dynamic Factor Models - An Approach via Singular Autoregression. *European Journal of Control*, 16(3):211–224, 2010.
14. M. Deistler, A. Filler, and M. Funovits. AR Systems and AR Processes: the Singular Case. *Communications in Informatics and Systems*, 11(3):225–236, 2011.
15. T. Doan, R. Litterman, and C. Sims. Forecasting and conditional projection using realistic prior distribution. *Econometric Reviews*, 3:1–100, 1984.
16. C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on Kalman filtering . *Journal of Econometrics*, 164(1):188 – 205, 2011. Annals Issue on Forecasting.
17. C. Doz, D. Giannone, and L. Reichlin. A quasi - maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):pp. 1014–1024, 2012.
18. R. Engle and M. Watson. A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates. *Journal of the American Statistical Association*, 76(376):774–781, Dec. 1981.
19. A. Filler. *Generalized Dynamic Factor Models; Structure Theory and Estimation for Single Frequency and Mixed Frequency Data*. PhD thesis, Technische Universität Wien, 210.
20. C. Flamm, U. Kalliauer, M. Deistler, M. Waser, and A. Graef. Graphs for dependence and causality in multivariate time series. In L. Wang and G. Hugues, editors, *System Identification, Environmental Modelling, and Control System Design*, chapter 7, pages 133–151. Springer Verlag, London, 2012.
21. M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, November 2000.
22. M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic factor model. *Journal of the American Statistical Association*, 100(471):830–840, 2005.
23. M. Forni, M. Hallin, M. Lippi, and P. Zaffaroni. One-Sided Representations of Generalized Dynamic Factor Models. Working Papers ECARES ECARES 2011-019, ULB – Universite Libre de Bruxelles, Aug. 2011.
24. M. Forni and M. Lippi. The generalized dynamic factor model: representation theory. *Econometric Theory*, 17:1113–1141, February 2001. JEL Classification: C13, C33, C43.
25. J. F. Geweke. The dynamic factor analysis of economic time series. In D. Aigner and A. Goldberger, editors, *Latent variables in socioeconomic models*. North Holland, Amsterdam, 1977.

26. M. Hallin and R. Liška. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617, 2007.
27. E. J. Hannan. *Multiple Time Series*. John Wiley & Sons Inc., New York, 1970.
28. E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Classics in Applied Mathematics. SIAM, Philadelphia, 2012. Originally published: John Wiley & Sons, New York, 1988.
29. T. Kailath. *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1980.
30. Y. A. Rozanov. *Stationary Random Processes*. Holden-Day, San Francisco, 1967.
31. T. J. Sargent and C. A. Sims. Business cycle modeling without pretending to have too much a priori economic theory. In C. A. Sims, editor, *New Methods in Business Cycle Research: Proceedings from a Conference*, pages 45–109. Federal Reserve Bank of Minneapolis, Minneapolis, Jan. 1977.
32. W. Scherrer and M. Deistler. A Structure Theory for Linear Dynamic Errors-in-Variables Models. *SIAM Journal on Control and Optimization*, 36(6):2148–2175, Nov. 1998.
33. J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.*, 97(460):1167–1179, 2002.