Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

WeA04.5

Generalized Linear Dynamic Factor Models – A Structure Theory

B.D.O. Anderson and M. Deistler

Abstract— In this paper we present a structure theory for generalized linear dynamic factor models (GDFM's). Emphasis is laid on the so-called zeroless case. GDFM's provide a way of overcoming the "curse of dimensionality" plaguing multivariate time series modelling, provided that the single time series are similar. They are used in modelling and forecasting for financial and macroeconomic time series.

I. INTRODUCTION

Factor analysis has been developed by psychologists for measurement of intelligence in the beginning of the twentieth century. In particular Burt and Spearman [1], observing that in tests of mental ability of a person, the scores on different items tended to be correlated, developed the hypothesis of a common latent factor, called general intelligence.

The "classical" static factor model is of the form

$$y_t = \ell_0 z_t + u_t, \quad \ell_0 \in \mathbb{R}^{N \times r} \tag{1}$$

where y_t is the N-dimensional vector of observations, z_t is the $r < N$ dimensional vector of (in general unobserved) factors and u_t is the strictly idiosyncratic noise, meaning that $E u_t u_t'$ is diagonal. In the classical case the factors and the noise are independently and identically distributed in t and factors and noise are uncorrelated. The variable $\hat{y}_t = \ell_0 z_t$ is called the vector of latent variables.

The basic idea of factor models is to split the observations into two parts; in the first part the *comovement* between the components of the observations is compressed into a small number of factors and in the second part the individual movements in the observations are described by uncorrelated noise components.

In a number of cases, the assumption of uncorrelatedness of the noise components is too restrictive. *Generalized* linear static factor models, where a certain form of "weak" correlation beween the noise components is allowed, have been introduced in [2], [3]. Here, "weak" means that the noise components can be averaged out, or, in other words, since the concept was motivated by applications in finance, that noise is the component of the observations which can be completely diversified.

In linear *dynamic* factor models, the latent variables can be expressed as a linear dynamic transformation of factors.

Generalized linear dynamic factor models (GDFM's) have been introduced in [4], [5], and, in a slightly different form,

B.D.O. Anderson is with the Department of Information Engineering, Australian National University, Canberra, Australia brian.anderson@anu.edu.au

M. Deistler is with the Department of Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria deistler@tuwien.ac.at

in [6], [7]. The idea is to generalize and combine linear dynamic factor models with strictly idiosyncratic noise (in the sense that the noise components are uncorrelated at any lead and lag) as analyzed in [8] and [9] and generalized linear static factor models. Factor models in a time series setting may be used to compress information contained in the data in both the cross-sectional dimension N and in the time dimension T. In this way it is possible to overcome the "curse of dimensionality" plaguing traditional multivariate time series modeling, where e.g. in the (unrestricted) autoregressive case, the dimension of the parameter space is proportional to N^2, whereas the number of data points is linear in N. The price to be paid for overcoming this curse of dimensionality is to require a certain kind of similarity between the component processes. An important idea in this context is to collect further information not only by adding data in the time dimension, but also by adding similar time series. The situation is not symmetric in the time and the cross-sectional dimension. For the time dimension we impose stationarity, for the cross-sectional dimension similarity or comovement of the single time series, in the sense that they are driven by a small number of common factors, has to be assumed.

Dynamic factor models are used for forecasting (see, e.g. [10], [11]); their advantage is that (in a certain setting) the problem can be reduced to forecasting low dimensional factors. Of course, factor models are also used for analysis (see, e.g. [12]), one aspect being that the factors may be interpreted e.g. as certain shocks to the economy or in financial applications as "market forces", whereas noise represents individual influences.

The basic idea of GDFM's is that the N-dimensional observation at time t, y_t^N say, can be represented as

$$y_t^N = \hat{y}_t^N + u_t^N \tag{2}$$

where $(\hat{y}_t^N)$ is the process of latent variables, which are strongly dependent in the cross-sectional dimension, and where (u_t^N) is the wide sense idiosyncratic noise, i.e. (u_t^N) is weakly dependent in the cross-sectional dimension. The precise meaning of these terms is given in Assumptions 4 and 5 below. For reasons which will be made clear below, we have added the superscript N. Throughout we assume

$$E \hat{y}_t^N = E u_t^N = 0 \quad \forall t \tag{3}$$

$$E \hat{y}_t^N u_s^{N'} = 0 \quad \forall s, t \tag{4}$$

and that $(\hat{y}_t^N)$ and (u_t^N) are wide sense stationary with absolutely summable covariances. Thus, using an obvious notation for the spectral densities corresponding to (2), we

obtain

$$f_y^N(\lambda) = f_{\hat{y}}^N(\lambda) + f_u^N(\lambda). \tag{5}$$

For dynamic factor models, the latent variables are obtained from

$$\hat{y}_t^N = \ell^N(z)f_t \tag{6}$$

where f_t are q-dimensional factors (with, normally, N much larger than q) and

$$\ell^N(z) = \sum_{j=-\infty}^{\infty} \ell_j^N z^j; \quad \ell_j^N \in \mathbb{R}^{N \times q} \tag{7}$$

are the factor loadings. Throughout, z is used for a complex variable as well as for the backward shift on $\mathbb{Z}$.

Parametric models have a number of advantages. We assume:

Assumption 1: $f_{\hat{y}}^N$ is a rational spectral density with constant rank $q < N$ on $[-\pi, \pi]$.

Since we are considering high dimensional time series, for asymptotic analysis, not only sample size T, but also the cross-sectional dimension N is tending to infinity; thus we consider a doubly indexed stochastic process $(y_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$, where i is the cross-sectional index and t denotes time. Therefore we consider a sequence of GDFM's (2). We assume:

Assumption 2: The double sequence $(y_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$ corresponds to a nested sequence of models, in the sense that $\hat{y}_{it}$ and u_{it} do not depend on N for $i \leq N$.

Assumption 3: The rank q and the McMillan degree, $2n$ say, of $f_{\hat{y}}^N$ are independent of N (from a certain N onwards).

The first part of assumption 3 is achieved by assuming that the factors in (6) do not depend on N.

Next, we define weak and strong dependence as in [5]. We use e.g. $\omega_{u,r}^N$ for the r-th largest eigenvalue of f_u^N.

Assumption 4 (Weak dependence): $\omega_{u,1}^N(\lambda)$ is uniformly bounded in λ and N.

Assumption 5 (Strong dependence): The first q eigenvalues of $f_{\hat{y}}^N$ diverge to infinity for all frequencies, as $N \to \infty$.

Contrary to the strict idiosyncratic case, generalized factor models are not generically identifiable for any fixed N, no matter how large. However, as has been shown in [5], the elements of $\hat{y}_t^N$ (and thus of u_t^N) are uniquely determined from (y_t^N) for $N \to \infty$. Moreover, consider the sequence of dynamic principal component decompositions (PCAs)

$$f_y^N(\lambda) = O_1^N(e^{-i\lambda})\Omega_1^N(\lambda)O_1^N(e^{-i\lambda})^* \tag{8}$$
$$+ O_2^N(e^{-i\lambda})\Omega_2^N(\lambda)O_2^N(e^{-i\lambda})^*$$

where Ω_1^N is the $q \times q$ diagonal matrix consisting of the q largest eigenvalues of f_y^N ordered according to size on its diagonal and O_1^N is the matrix whose columns are the corresponding eigenvectors and where the second part on the right hand side of the above equation is defined analogously for the smallest eigenvalues. Here, e.g. $O_1^N(z)^*$ denotes $O_1^N(z^{-1})'$. As has been shown in [5], such a sequence of PCAs for $N \to \infty$ converges to the corresponding GDFM's in the sense that e.g. the scalar components of the latent PCA variables $\hat{y}_{PCA,t}^N = O_1(z)O_1^*(z)y_t$ converge to the respective

scalar components of $\hat{y}_t^N$. From now on, for the sake of simplicity of notation, we will omit the superscript N, unless the contrary is stated explicitly.

In general terms, in this paper we are concerned with the structure of GDFM's where the latent variables have a singular rational spectral density, using and further developing results from system theory. This structure is useful for gaining insight and, even more, for subsequent estimation. Our ultimate goal is to obtain a model for the latent variables $\hat{y}_t$ from the observations y_t, without estimating the noise parameters.

In a first step we consider an idealized problem; we commence from the rational and singular population spectral density of the latent variables, rather than from observed data or the corresponding sample second moments. A state space realization for a spectral factor of such a spectral density is given and the notions of minimal static factors and minimal dynamic factors and their relation to the state are clarified.

Our emphasis is on the zeroless case, which in our setting is generic. In this case the latent variables as well as the minimal static factors are autoregressive processes. However, if the dimension of the minimal static factors exceeds the dimension of the minimal dynamic factors, the autoregressive processes are non standard, since their driving white noise has a singular variance matrix. Since the relation between latent variables and static factors is static and straightforward to determine, we may restrict ourselves to modeling of the static factors. An algorithm based on Yule-Walker equations is proposed. In our setting the solution of the Yule-Walker equations is not necessarily unique. In any case, only a *finite number* of population covariances of the latent variables are needed for realizing the system.

In section IV we discuss the relevance for actual estimation, of the results obtained for the idealized setting. By letting the cross sectional dimension N and the sample size T diverge to infinity, the effects of the weakly idiosyncratic noise on the latent variables can be filtered out and the sample second moments of the observations converge to the population second moments of the latent variables. One way to show this is to use PCA for estimating the minimal static factors and their second moments. To the best of our knowledge this approach is novel.

II. SPECTRAL FACTORIZATION OF SINGULAR RATIONAL SPECTRAL DENSITIES AND REALIZATION OF TALL TRANSFER FUNCTIONS

As has been stated above, here, dealing with an idealized setting, we commence from the rational, singular population spectral density of the latent variables, rather than directly from the data or their sample second moments, in order to find the underlying system generating the latent variables. The special feature here is that the spectral density is singular, thus we can find a tall spectral factor. The stable and minimum phase factor considered is the transfer function corresponding to Wold decomposition. In this Wold decomposition the white noise inputs are at the same time dynamic factors.

After having obtained such a transfer function, we discuss its realization as an ARMA or state space system. The particular form of the state space representation used in this paper provides valuable insights. We show how static factors and the state can be obtained using the so-called Hankel matrix of the transfer function.

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

$$f_{\hat{y}}(\lambda) = \frac{1}{2\pi} w(e^{-i\lambda}) w(e^{-i\lambda})^* \qquad (9)$$

where $w(z)$ is a $N \times q$ real rational matrix which has no poles and no zeros for $|z| \leq 1$.

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

The spectral factors

$$w(z) = \sum_{j=0}^{\infty} w_j z^j, \quad w_j \in \mathbb{R}^{N \times q} \qquad (10)$$

correspond to a causal linear finite dimensional system

$$\hat{y}_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j} \qquad (11)$$

where the inputs (ε_t) are white noise with $E\varepsilon_t\varepsilon_t' = 2\pi I_q$. Let us repeat, our particular focus will be on the case where w is tall, i.e. $N > q$ holds.

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

$$w = udv \qquad (12)$$

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 zeros of w are the zeros of the n_i and the poles of w are the zeros of the d_i.

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

$$w^- = v^{-1}(d'd)^{-1}d'u^{-1} \qquad (13)$$

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 (11) is uniquely determined, independently of the particular choice of the causal inverse from $\hat{y}_t, \hat{y}_{t-1}, \ldots$; thus (11) corresponds to Wold decomposition (see, e.g. [13]).

Every rational causal transfer function can be realized by an ARMA system, by a right MFD or by a state space system. Let us start with ARMA systems:

$$a(z)\hat{y}_t = b(z)\varepsilon_t. \qquad (14)$$

We assume that (a, b) are left coprime (see e.g. [13]), then the set of all observationally equivalent left coprime ARMA systems is obtained as (ua, ub) where u is an arbitrary unimodular matrix.

The conditions on the poles and zeros of the transfer function $w = a^{-1}b$ are, for left coprime a, b, equivalent to

$$\det a(z) \neq 0, \qquad |z| \leq 1 \qquad (15)$$

and

$$b(z) \text{ has full rank } q, \quad |z| \leq 1. \qquad (16)$$

A right matrix fraction description (MFD)

$$w = dc^{-1} \qquad (17)$$

where d and c are polynomial matrices of appropriate dimension, corresponds to an AR process applied to a finite impulse response and has been used in [5].

We consider state space realizations of w of the form

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \qquad (18)$$
$$\hat{y}_t = Hx_t \qquad (19)$$

where x_t is the n-dimensional state and $F \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{n \times q}$, $H \in \mathbb{R}^{N \times n}$. Note that the state space form (18)–(19) is different from the form considered in [13]; we have chosen this form because of its convenience for our purposes. The basic ideas and concepts are completely analogous to the usual form. We assume that the system is minimal and stable, the latter meaning that

$$|\lambda_{max}(F)| < 1 \qquad (20)$$

(where $\lambda_{max}(F)$ denotes an eigenvalue of maximum modulus). The transfer function for (18)–(19) is given by

$$w(z) = HG + \sum_{j=1}^{\infty} HF^j Gz^j. \qquad (21)$$

Note that $\mathrm{rk}(HG) = q$ implies $\mathrm{rk}\, G = q$. If (F, G, H) is minimal, then the transfer function w has a zero for some z_0 if and only if the matrix

$$M(z) = \begin{pmatrix} I - Fz & -G \\ H & 0 \end{pmatrix} \qquad (22)$$

has rank less than $n + q$ at z_0 [14]. The form (18)–(19) is obtained by the "Akaike-Kalman procedure" [15] from the equation

$$\underbrace{\begin{pmatrix} \hat{y}_t \\ \hat{y}_{t+1|t} \\ \hat{y}_{t+2|t} \\ \vdots \end{pmatrix}}_{\hat{Y}_t} = \underbrace{\begin{pmatrix} w_0 & w_1 & \cdots \\ w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{\mathscr{H}} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \end{pmatrix} \qquad (23)$$

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 $\mathscr{H}$ is called the 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 basis in terms of the components of $\hat{Y}_t$ from $\hat{Y}_t$. Then the equations

$$x_t = S\hat{Y}_t \tag{24}$$

$$S\begin{pmatrix} w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = FS\mathscr{H} \tag{25}$$

$$G = S\left(w_0', w_1', \dots\right)' \tag{26}$$

$$(w_0, w_1, \dots) = HS\mathscr{H} \tag{27}$$

(compare [16]) define a (minimal) state space system (18)–(19) in echelon form. From now on, we mainly consider echelon forms. Every other minimal state is obtained by premultiplying the echelon state by a constant nonsingular matrix.

From (19) we see that x_t is a static factor. Note that the variance matrix of a minimal state is always nonsingular. The state x_t is not necessarily a minimal static factor, i.e. a static factor of minimal dimension. The state x_t is a minimal static factor if and only if $\operatorname{rk} H = n$ holds, otherwise (in case that x_t is not an echelon state, after premultiplying x_t with a suitable matrix) we obtain

$$\hat{y}_t = (H_1, 0)x_t = H_1 z_t \tag{28}$$

where z_t is a, r-dimensional say, minimal static factor and $\operatorname{rk} H_1 = r$ holds.

The system

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{29}$$

is an example of a minimal system where $q = 1, n = 3$ and the dimension of a minimal static factor is $r = 2 < 3$.

In general we always have $n \geq r \geq q$. From what was said above, a minimal static factor can be obtained by selecting the first linear independent components from $\hat{y}_t$ and $n = r$ holds if and only if $\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \dots$ do not contain further linearly independent components. Thus, in general, we may write

$$S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}, S_1 \in \mathbb{R}^{r \times \infty}, S_2 \in \mathbb{R}^{(n-r) \times \infty} \tag{30}$$

where

$$z_t = S_1 \hat{Y}_t. \tag{31}$$

(in other words x_t is a minimal static factor if and only if all Kronecker indices of $\mathscr{H}$ (see [13], chapter 2) are equal to zero or one and the Kronecker indices equal to one define the static factor.)

For given $(\hat{y}_t)$, the minimal static factors are unique up to premultiplication by a constant nonsingular matrix and a static factor may be obtained from

$$E\hat{y}_t\hat{y}_t' = MM', M \in \mathbb{R}^{N \times r}, \operatorname{rk} M = r \tag{32}$$

by

$$z_t = (M'M)^{-1}M'\hat{y}_t. \tag{33}$$

Thus the static factors z_t are obtained by a simple static linear transformation of the latent variables $\hat{y}_t$ (and vice versa). We have

$$z_t = (M'M)^{-1}M'w(z)\varepsilon_t = k(z)\varepsilon_t \tag{34}$$

where $k(z)$ is the transfer function corresponding to the static factors.

III. ZEROLESS TRANSFER FUNCTIONS AND AUTOREGRESSIVE SYSTEMS

Of particular interest for us are zeroless transfer functions. In this case the latent variables may be represented by an AR system. However, these AR systems differ from the usual ones, since they may be singular in the sense that their driving white noise may have a singular variance matrix.

In this case, the static factors may be represented by an autoregression, again the variance matix of the driving white noise may be singular. Since the latent variables can be obtained from the static factors in a straightforward way by a linear static transformation, we propose to use an AR model for the static factors in order to avoid "redundant" dimensions. Such an AR model is obtained by solving the Yule-Walker equations. These equations commence from a finite number of second moments of the static factors (and of the latent variables) and they give the correct spectral factors. As opposed to the usual case, for singular AR systems, the solutions of the Yule-Walker equations may not be unique.

Definition 1: An $N \times q$ transfer function $w(z)$ is called zeroless if $w(z)$ has full column rank q for all $z \in \mathbb{C}$.

For $N = q$, the zeroless case is nongeneric; in the tall case however, the zeroless case is generic. We have ([17]):

Theorem 2: Consider a rational transfer function w with a minimal state space realization (F, G, H) with state dimension n. If $N > q$ holds, then for generic values of (F, G, H), the transfer function w is zeroless.

This can be seen from the fact that the zeros of w are the intersection of the sets of zeros of the determinants of all $q \times q$ submatrices of w. A more precise proof is given in [17]. In the zeroless case, the numerator polynomials n_i of d in (12) are all equal to one and thus w^- given by (13) is polynomial. Then the input ε_t is determined from a finite number of outputs $\hat{y}_t, \hat{y}_{t-1}, \dots, \hat{y}_{t-L}$, for some L.

Note that $\operatorname{rk} H = n$ implies that w is zeroless. This is easily seen from (22) since always $\operatorname{rk} G = q$ holds. However, as the example in the previous section shows, for zeroless transfer functions $\operatorname{rk} H < n$ may hold; in other words, assuming that $w(z)$ is zeroless is more general than assuming $\operatorname{rk} H = n$. The case where w is not necessarily zeroless has been treated in [18], [19], [20].

Theorem 3: Let $(\hat{y}_t)$ satisfy Assumptions 1 through 5, then the following statements are equivalent:

(i) The spectral factors w satisfying the properties listed in Theorem 1 are zeroless

(ii) There exists a polynomial left inverse w^- for w

(iii) $(\hat{y}_t)$ is a stable AR process, i.e.

$$\hat{y}_t = a_1\hat{y}_{t-1} + \cdots + a_p\hat{y}_{t-p} + \nu_t, \qquad (35)$$
$$a_j \in \mathbb{R}^{N \times N}, E\nu_t = 0, E\nu_s\nu_t = \delta_{st}\Sigma_\nu$$

where

$$\det(I - a_1 z - \cdots - a_p z^p) \neq 0, |z| \leq 1$$

and $\operatorname{rk}\Sigma_\nu = q$.

Proof: (i) $\Rightarrow$ (ii) has been shown above. For showing (i) $\Rightarrow$ (iii), we commence from (14); since w is zeroless, the same holds for b. Now, as easily seen, every tall zeroless matrix can be completed by a suitable choice of a polynomial matrix c to a unimodular matrix $u = (b, c)$. Then

$$a(z)\hat{y}_t = u(z)\begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

and

$$u^{-1}(z)a(z)\hat{y}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} = \nu_t \qquad (36)$$

gives an autoregressive representation.

That (ii) implies (i) is straightforward and that (iii) implies (ii) can be seen as follows: Write

$$\Sigma_\nu = NN', N \in \mathbb{R}^{N \times q}, \operatorname{rk} N = q \qquad (37)$$

then premultiplying (35) by $(N'N)^{-1}N'$ yields a w^- of the desired form. ∎

Let $z_t = k(z)\varepsilon_t$ denote a minimal static factor. Then as a direct consequence of (33), $k(z)$ is zeroless if and only if $w(z)$ is zeroless. Thus, in the zeroless case, (z_t) has an autoregressive representation as well:

$$b(z)z_t = \mu_t \qquad (38)$$
$$b(z) = I - b_1 z - \cdots - b_p z^p, \quad b_j \in \mathbb{R}^{r \times r}$$

satisfying

$$b(z) \neq 0, \quad |z| \leq 1 \qquad (39)$$

and where (μ_t) is white noise satisfying

$$\operatorname{rk}\Sigma_\mu = \operatorname{rk} E z_t z_t' = q. \qquad (40)$$

To repeat, since the latent variables are obtained by a simple static linear transformation from the static factors and since the dimension r of the minimal static factors is in general much smaller than N, estimation will be performed for (38) rather than for (35).

As is well known, in the regular case, i.e. when Σ_μ is nonsingular, the matrices

$$\underline{\Gamma}_p = \begin{pmatrix} \gamma_0 & \cdots & \cdots & \gamma_{p-1}' \\ \vdots & \gamma_0 & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \cdots & \gamma_0 \end{pmatrix}, \qquad (41)$$

where $\gamma_j = E z_{t+j}z_t'$, are nonsingular for all $p \in \mathbb{N}$ and $b(z)$ is uniquely defined from the (population) second moments of (z_t). For singular AR systems, things are more subtle.

In the zeroless case, Yule-Walker equations can be used to obtain an autoregression for (z_t) and thus a spectral factor w from a finite number of covariances of $(\hat{y}_t)$:

Equation (38) gives the following Yule-Walker equations:

$$(b_1, \ldots, b_p)\underline{\Gamma}_p' = (0, \ldots, 0) \qquad (42)$$
$$\Sigma_\mu = \gamma_0 - (b_1, \ldots, b_p)(\gamma_1', \ldots, \gamma_p') \qquad (43)$$

Formula (42) may be used to determine $(b_1, \ldots, b_p)$. Note that in the case $q < r$, as opposed to the regular case $r = q$, the matrix $\underline{\Gamma}_{p+1}'$ will be singular and the matrix $\underline{\Gamma}_p$ may be singular, i.e. the components of the vectors $(z_{t-1}', \ldots, z_{t-p-1}')'$ and $(z_{t-1}', \ldots, z_{t-p}')'$ will, or may be, respectively, linearly dependent and thus the solution may not be unique. However, by the projection theorem, every solution determines the same $z_{t|t-1}$ and μ_t.

The possible nonuniqueness of the Yule-Walker equations can be seen from a description of the class of observationally equivalent systems. The idea is to relate the singular AR case to the ARMA case (see [13]). We obtain the following result (by $\delta(b(z))$ we denote the degree of the polynomial matrix $b(z)$):

Theorem 4: (i) Every singular AR system with $\operatorname{rk}\Sigma_\mu = q$ can be written as

$$b(z)z_t = c\varepsilon_t, c \in \mathbb{R}^{r \times q} \qquad (44)$$

where (ε_t) is white noise with $E\varepsilon_t\varepsilon_t' = I_q$ and where $b(z)$ and c are relatively left prime.

(ii) Let (44) be relatively left prime, then the class of all observationally equivalent $(\bar{b}(z), \bar{c})$ satisfying the degree restriction $\delta(\bar{b}(z)) \leq p$ and that $\bar{c}$ is constant, is given by

$$(\bar{b}(z), \bar{c}) = u(z)(b(z), c) \qquad (45)$$

where the polynomial matrix $u(z)$ satisfies

$$\det u(z) \neq 0, |z| \leq 1 \qquad (46)$$
$$u(0) = I \qquad (47)$$
$$\delta(u(z)b(z)) \leq p \qquad (48)$$
$$\delta(u(z)c) = 0 \qquad (49)$$

In addition, $(\bar{b}(z), \bar{c})$ is relatively left prime if and only if $u(z)$ is unimodular.

(iii) $b(z)$ is unique if and only if $\operatorname{rk}(b_p, c) = r$ holds.

Proof: For (i) it only remains to show that $(b(z), c)$ can be chosen as relatively left prime. Assume that $(b(z), c)$ are not relatively left prime, then we can always find a relatively left prime observationally equivalent system $(\bar{b}(z), \bar{c}(z))$, where the degree of $\bar{c}(z)$ is not necessarily zero. By Theorem 3, $\bar{c}(z)$ must be zeroless and thus can be extended to a unimodular matrix. Premultiplying $(\bar{b}(z), \bar{c}(z))$ by the inverse of this unimodular matrix yields the desired result. (ii) and (iii) are straightforward. ∎

IV. ESTIMATION

Here we briefly describe how the results from structure theory can be used for estimation, i.e. how to proceed if we

commence from the observations of (y_t), rather than from the population moments of $(\hat{y}_t)$.

A main issue is to remove the effect of the noise u_t^N by letting $N \to \infty$. There are several ways to do this ([6], [11]). The simplest way is to perform a principal component analysis on

$$\bar{\Sigma}_y^N = T^{-1} \sum_{t=1}^{T} y_t^N y_t^N{}' \qquad (50)$$

in order to obtain an estimate of z_t, which will be consistent for $N \to \infty$, $T \to \infty$.

V. SUMMARY AND OUTLOOK

This contribution is concerned with generalized linear dynamic factor models. We present a structure theory. Based on the observation that tall transfer functions are generically zeroless, we propose Yule-Walker equations for estimating an autoregressive system for the minimal static factor and for the latent variables. This approach is more general than the approach used up to now, where the autoregressive system has been assumed to be of order one. Important issues not treated in this paper are estimation of integers, namely estimation of the dimension q of the dynamic factors, the dimension r of the static factors and the state dimension n.

REFERENCES

[1] C. Burt, "Experimental tests of general intelligence," *British Journal of Psychology*, vol. 3, pp. 94–177, 1909.

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

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

[4] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, "The generalized dynamic factor model: identification and estimation," *The Review of Economic Studies*, vol. 65, pp. 453–473, 2000.

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

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

[7] ——, "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, vol. 20, pp. 147–62, 2002.

[8] T. J. Sargent and C. A. Sims, *Business cycle modelling without pretending to have too much a priori economic theory*. Minneapolis: Federal Reserve Bank of Minneapolis, 1977.

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

[10] M. Deistler and E. Hamann, "Identification of factor models for forecasting returns," *J. of Financial Econometrics*, vol. 3, pp. 256–281, 2005.

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

[12] M. Forni, D. Giannone, M. Lippi, and L. Reichlin, "Opening the black box: structural factor models with large cross-sections," *Manuscript, Universite Libre de Bruxelles*, 2008.

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

[14] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

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

[16] M. Deistler, "Linear dynamic multiinput/multioutput systems," in *A Course in Time Series Analysis*, D. Pena, G. Tiao, and R. Tsay, Eds. Wiley, 2001.

[17] B. D. O. Anderson and M. Deistler, "Properties of zero-free transfer function matrices," 2008, mimeo.

[18] C. Zinner, "Modeling of high-dimensional time series by generalized dynamic factor models," Ph.D. dissertation, Vienna University of Technology, 2008.

[19] C. Heaton, "Factor analysis of high dimensional time series," Ph.D. dissertation, University of New South Wales, 2008.

[20] C. Heaton and V. Solo, "Identification of causal models for stationary time series," *The Econometrics Journal*, vol. 7, pp. 618–627, 2004.