Generalized Linear Dynamic Factor Models - An Approach via Singular Autoregressions

A. Filler, M. Deistler, B.D.O. Anderson, Ch. Zinner and W. Chen

***Abstract*— We consider generalized linear dynamic factor models. These models have been developed recently and they are used for high dimensional time series in order to overcome the "curse of dimensionality". We present a structure theory with emphasis on the zeroless case, which is generic in the setting considered. Accordingly the latent variables are modeled as a possibly singular autoregressive process and (generalized) Yule Walker equations are used for parameter estimation.**

I. Introduction

Generalized linear dynamic factor models (GDFM's) have been introduced in [1], [2], and, in a slightly different form, in [3], [4]. The idea is to generalize and combine linear dynamic factor models with strictly idiosyncratic noise as analyzed in [5] and [6] and generalized linear static factor models, introduced in [7] and [8]. Factor models in a time series setting may be used to compress information contained in the data in both the cross-sectional dimension, N say, 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 for a fixed T is linear in N. The price to be paid for overcoming this curse of dimensionality is to require a certain kind of similarity or comovement between the single time series.

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

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

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 the words weak and strong dependence will be given below.

Throughout we assume

$$\mathbb{E}\hat{y}_t^N = \mathbb{E}u_t^N = 0 \quad \forall t \tag{2}$$

$$\mathbb{E}[\hat{y}_t^N {u_s^N}'] = 0 \quad \forall s,t \tag{3}$$

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 (1), we obtain

$$f_y^N(\lambda) = f_{\hat{y}}^N(\lambda) + f_u^N(\lambda). \tag{4}$$

The latent variables are obtained from factors, see below.

Throughout, z is used for a complex variable as well as for the backward shift on $\mathbb{Z}$. This is opposite to the most common convention in control, but consistent with much econometrics literature.

In addition we assume:

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 $q < N$ 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 $(y_{it} \,|\, i \in \mathbb{N}, t \in \mathbb{Z})$, where i is the cross-sectional index and t denotes time. Therefore we consider a *sequence* of GDFM's (1). We assume:

Assumption 2: The double sequence $(y_{it} \,|\, i \in \mathbb{N}, t \in \mathbb{Z})$ 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 of $f_{\hat{y}}^N$ and its Mc Millan degree $2n$ say, are independent of N ($N \geq some\ N_0$).

Next, we define weak and strong dependence as in [2]. We use e.g. $\omega_{u,r}^N$ to denote the r-th largest eigenvalue of f_u^N.

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

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

Contrary to the strict idiosyncratic case, the case considered in "ordinary" factor models, where f_u^N is assumed to be diagonal, generalized factor models are not generically identifiable for any fixed N, no matter how large. Nevertheless, as has been shown in [2], the elements of $\hat{y}_t^N$ (and thus of u_t^N) are uniquely determined from (y_t^N)

Support by the FWF (Austrian Science Fund) under contracts P17065/N12, P17378 and P20833/N18 is gratefully acknowledged.

This work has been supported in part by ARC Discovery Project Grant DP0664427 and National ICT Australia Ltd. National ICT Australia Ltd. is funded through the Australian Governments Backing Australias Ability initiative, and in part through the Australian Research Council.

A. Filler is with the Department of Business Studies, University of Vienna, and the Department of Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria alexander.filler@tuwien.ac.at

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

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

Ch. Zinner is with the Group of Market Risk, BAWAG P.S.K. , Vienna, Austria c_zinner@gmx.at

W. Chen is with the Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia weitian.chen@anu.edu.au

ISBN 978-963-311-369-1

for $N \to \infty$. Moreover, consider the sequence of dynamic principal component decompositions

$$f_y^N(\lambda) = O_1^N(e^{-i\lambda})\Omega_1^N(\lambda)O_1^N(e^{-i\lambda})^* + O_2^N(e^{-i\lambda})\Omega_2^N(\lambda)O_2^N(e^{-i\lambda})^* \quad (5)$$

where Ω_1^N is the $q \times q$ diagonal matrix consisting of the q largest eigenvalues of f_y^N ordered as a descending sequence on its diagonal and O_1^N is the matrix whose columns are the corresponding eigenvectors; 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 [2], such a sequence of PCA models 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.

In general terms, *this paper is concerned with identification of GDFM's, where the latent variables have a singular rational spectral density, or to be more precise, with the identification of the linear (state space or ARMA) system generating the latent variables, from the observations* $y_1, ..., y_T$. We neither impose additional structure on the noise, nor are we interested in estimating the noise parameters.

The emphasis of this paper is on structure theory. It heavily draws from previous work ([9],[10],[11]). In the structure theory considered here an idealized setting is considered, since we commence from the population second moments of the latent variables, rather than from the sample second moments of the observations in order to obtain the parameters of the system generating the latent variables.

As in [10] our emphasis is on the zeroless case, which in our setting is generic. We extend results given in this previous paper. New contributions in this paper include the characterization of the latent variable model as a singular autoregression (where the driving white noise has a singular variance matrix), the identification of a minimal static factor, and the use of (generalised) Yule-Walker equations to obtain a singular autoregression.

The paper is organized as follows: In section II, we consider spectral factorization and realization of tall rational transfer functions. The relation between the dimensions of (minimal) states, of minimal static factors and of minimal dynamic factors is described. Section III is concerned with zeroless transfer functions and their realization by (possibly singular) autoregressive systems. Section IV is concerned with solutions of (generalized) Yule Walker equations, their continuity and stability. Since the static factors can be obtained by a linear static transformation from the latent variables and the static factors have smaller dimension, we concentrate on realizing the static factors by autoregressive systems.

GDFM's are used, both for forecasting and for analysis of high dimensional time series (see e.g. [12],[13],[14]). In forecasting, the forecasts of the latent variables are used to forecast the observed variables.

II. Realization of Rational Tall Transfer Functions

As has been shown in [1],[15] the effect of the noise on the observations can be removed for $N \to \infty$ (e.g. by PCA), and the spectral density $f_{\hat{y}}$ of the latent variables can be estimated consistently for $N, T \to \infty$. Therefor it is reasonable to analyse an idealised setting were we commence from the population spectral density $f_{\hat{y}}$ of the latent variables $(\hat{y}_t)$ rather than from the spectral density f_y of (y_t).

A. Spectral factorization and Wold decomposition

We have the following result ([16], [17]).

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

$$f_{\hat{y}}(\lambda) = \frac{1}{2\pi} w(e^{-i\lambda}) w(e^{-i\lambda})^* \quad (6)$$

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 constant orthogonal matrices.

The spectral factors

$$w(z) = \sum_{j=0}^{\infty} w_j z^j, \quad w_j \in \mathbb{R}^{N \times q} \quad (7)$$

correspond to a causal linear finite dimensional system

$$\hat{y}_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j} \quad (8)$$

where the inputs (ε_t) are white noise with $\mathbb{E}[\varepsilon_t \varepsilon_t'] = 2\pi I_q$. We will be concerned with the case where w is tall, i.e. $N > q$ holds.

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

$$w = udv \quad (9)$$

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. 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

$$w^- = v^{-1}(d'd)^{-1}d'u^{-1} \quad (10)$$

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 (8) is uniquely determined from $\hat{y}_t, \hat{y}_{t-1}, \ldots$, independently of the particular choice of the causal inverse

$$\varepsilon_t = \sum_{j=0}^{\infty} w_j^- \hat{y}_{t-j} \quad (11)$$

Thus (8) corresponds to Wold decomposition (see, e.g. [18]).

B. ARMA representation

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:

$$a(z)\hat{y}_t = b(z)\varepsilon_t. \tag{12}$$

We assume that (a, b) are left coprime (see e.g. [18]), 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

$$\det a(z) \neq 0, \qquad |z| \leq 1 \tag{13}$$

and

$$b(z) \text{ has full rank } q, \quad |z| \leq 1. \tag{14}$$

A right MFD

$$w = dc^{-1} \tag{15}$$

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 [2].

C. State Space realization

We can also consider state space realizations of w of the form

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{16}$$

$$\hat{y}_t = Hx_t \tag{17}$$

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 (16)–(17) is different from the form considered in [18]; we have chosen this form because of its convenience for our purposes. We assume that the system is minimal, stable, i.e.

$$|\lambda_{max}(F)| < 1 \tag{18}$$

(where $\lambda_{max}(F)$ denotes an eigenvalue of maximum modulus) and mini-phase, i.e. the right side of (20) has no zeros for $|z| \leq 1$. The transfer function for (16)–(17) is given by

$$w(z) = HG + \sum_{j=1}^{\infty} HF^jGz^j. \tag{19}$$

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

$$M(z) = \begin{pmatrix} I - Fz & -G \\ H & 0 \end{pmatrix} \tag{20}$$

has rank less than $n+q$ at z_0. Starting with the power series expansion (7), the form (16)–(17) is obtained by the "Akaike-Kalman procedure" [19] from the equation

$$\underbrace{\begin{pmatrix} \hat{y}_t \\ \hat{y}_{t+1|t} \\ \hat{y}_{t+2|t} \\ \vdots \end{pmatrix}}_{\hat{Y}_t} = \underbrace{\begin{pmatrix} w_0 & w_1 & \cdots \\ w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{\mathscr{H}} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \end{pmatrix} \tag{21}$$

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 (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 basis in terms of the components of $\hat{Y}_t$ from $\hat{Y}_t$. Then the equations

$$x_t = S\hat{Y}_t \tag{22}$$

$$S\begin{pmatrix} w_1 & w_2 & \cdots \\ w_2 & w_3 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = FS\mathscr{H} \tag{23}$$

$$G = S\left(w_0', w_1', \ldots\right)' \tag{24}$$

$$(w_0, w_1, \ldots) = HS\mathscr{H} \tag{25}$$

(compare [20]) define a (minimal) state space system (16)–(17) 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.

D. Static Factors

A *static factor* of the latent variables $(\hat{y}_t)$ is a process (z_t) of lesser vector dimension, with the property that for some constant matrix L, there holds $\hat{y}_t = Lz_t$ for all t. A *minimal static factor* is one for which z_t has least dimension. It is obvious that x_t itself is a static factor, but our interest is in studying minimal static factors.

We note the standard result:

Lemma 1: Let $\hat{y}_t$ be a stationary vector process. Then the dimension of a minimal static factor is the rank, call it r, of the zero-lag variance matrix $\mathbb{E}[\hat{y}_t\hat{y}_t']$.

Proof: Suppose z_t is a static factor, with $\hat{y}_t = Lz_t$. Then $\mathbb{E}[\hat{y}_t\hat{y}_t'] = L\mathbb{E}[z_tz_t']L'$, and it follows that there can be no static factor of dimension less than the rank of $\mathbb{E}[\hat{y}_t\hat{y}_t']$. To show that there is indeed a static factor with this dimension, let M be any matrix with least number of columns such that $\mathbb{E}[\hat{y}_t\hat{y}_t'] = MM'$. Notice that M is unique up to right multiplication by an orthogonal matrix. Make the definition

$$z_t = (M'M)^{-1}M'\hat{y}_t \tag{26}$$

which means that z_t has dimension equal to the rank of $\mathbb{E}[\hat{y}_t\hat{y}_t']$ and has variance I. It is trivial to verify also that

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

Thus z_t is indeed a minimal static factor. ■

Observe that a minimal static factor is not required to have a unit variance matrix. For any nonsingular R, Rz_t with z_t as just defined is a minimal static factor (and indeed all minimal static factors are obtained this way). Also, there is an infinite family of minimal static factors of unit variance, obtained by multiplying z_t by an arbitrary orthogonal matrix.

Knowing H and x_t, we can also construct a minimal static factor. Let T be any nonsingular matrix such that

$HT = [H_1\ 0]$ where H_1 has full column rank. Since x_t has a nonsingular covariance, the fact that $\mathrm{rk}\,\mathbb{E}[\hat{y}_t\hat{y}_t'] = r$ means that $\mathrm{rk}\,H_1 = r$. A particular minimal static factor is then defined by

$$\bar{z}_t = [I\ \ 0]T^{-1}x_t \tag{28}$$

for it is trivial to verify that $\hat{y}_t = H_1\bar{z}_t$, and $\bar{z}_t$ has dimension r. Evidently, $n \geq r \geq q$ and x_t is a minimal static factor if and only if $\mathrm{rk}\,H = n$. Another way to obtain a minimal static factor is to select the first linearly independent components of $\hat{y}_t$, and $n = r$ holds if and only if $\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \ldots$ do not contain further linearly independent components. Thus in general we may write

$$S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}, S_1 \in \mathbb{R}^{r\times\infty}, S_2 \in \mathbb{R}^{(n-r)\times\infty} \tag{29}$$

where

$$z_t = S_1\hat{Y}_t. \tag{30}$$

and x_t is a minimal static factor if and only if all Kronecker indices of $\mathscr{H}$ (see [18], chapter 2) are equal to zero or one. The Kronecker indices equal to one define a static factor.
A minimal static factor is obtainable by a simple static linear transformation of the latent variables $\hat{y}_t$ and vice versa. This means that, using the minimal static factor z_t above as an example,

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

for some $k(z)$ transfer function corresponding to the static factors. As with $w(z)$, $k(z)$ has no poles and zeros in $|z| \leq 1$. It is also easy to show that a minimal state space realization for z_t is obtainable as $\{F, G, C\}$ where

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

Since z_t has the same dynamics as $\hat{y}_t$ and is of smaller dimension, modeling of (z_t) is more convenient. As (z_t) is an ARMA process, ARMA identification procedures, such as the autoregression-regression approach ([21], [22]) may be applied to this case, as has been done in [23]. Here, however, we deal with an autoregressive approach, which is simpler and can be applied in a generic situation.

III. Zeroless Transfer Functions and Autoregressive Systems

Of particular interest for us are zeroless transfer functions, because, as argued below, tall transfer functions are generically zeroless. As we will show, 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, also the static factors may be represented by an autoregression, and again the variance matrix of the driving white noise may be singular. We propose to use an AR model for the static factors in order to avoid "redundant" dimensions. Such an AR model will be obtained by solving the Yule-Walker equations. These equations commence from a finite number of second moments of the static factors (and thus of the latent variables), they are linear in the unknown parameters and they give the correct spectral factors. However, 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 the numerator polynomials of the diagonal matrix in its Smith-McMillan form (9) are all equal to one.
For $N = q$, the zeroless case is nongeneric; in the tall case however, the zeroless case is generic. We have ([10]):

Theorem 2: Consider an $N\times q$ 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 [10].
As is easily seen from (31), $k(z)$ is zeroless if and only if $w(z)$ is zeroless. In the zeroless case, the numerator polynomials of the diagonal matrix in the Smith-McMillan form (9) are all equal to one and thus k^- coresponding to (10) is *polynomial*. Then *the input ε_t is determined from a finite number of outputs $z_t, z_{t-1}, \ldots, z_{t-L}$, for some L.*

Note that $\mathrm{rk}\,H = n$ implies that w and thus k are zeroless. This is easily seen from (20) since always $\mathrm{rk}\,G = q$ holds. However, for zeroless transfer functions w, $\mathrm{rk}\,H < n$ may hold; in other words, assuming that $w(z)$ is zeroless is more general than assuming $\mathrm{rk}\,H = n$.

Theorem 3: Let $(\hat{y}_t)$ satisfy Assumptions 1-3; then the following statements for (z_t) 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^- for k and thus the input ε_t in (31) is determined from a finite number of output values
(iii) (z_t) is a stable AR process, i.e.

$$z_t = e_1z_{t-1} + \cdots + e_pz_{t-p} + \nu_t \tag{33}$$

where

$$\det \underbrace{(I - e_1z - \cdots - e_pz^p)}_{e(z)} \neq 0,\ |z| \leq 1$$

ν_t is a zero mean white noise process with $\mathrm{rk}\,\Sigma_\nu = q$, $\Sigma_\nu = \mathbb{E}[\nu_t\nu_t']$.

Proof: (i) $\Rightarrow$ (ii) follows from Definition1 above. In order to show (i) $\Rightarrow$ (iii), we commence from an ARMA representation for z_t

$$\tilde{e}(z)z_t = f(z)\varepsilon_t \tag{34}$$

where $\tilde{e}, f$ are relatively left prime. Since $k(z) = \tilde{e}^{-1}(z)f(z)$ is zeroless, the same holds for $f(z)$. Now, as easily can be seen, every zeroless tall polynomial matrix can be completed by a suitable choice of a polynomial matrix g to a unimodular matrix $u = (f, g)$. Then

$$\tilde{e}(z)z_t = u(z)\begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

and

$$u^{-1}(z)\tilde{e}(z)z_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix} \tag{35}$$

gives an autoregressive representation, and premultiplying (35) by $\tilde{e}^{-1}(0)u(0)$ gives the desired form (33).

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

$$\Sigma_\nu = PP', P \in \mathbb{R}^{r\times q}, \operatorname{rk} P = q \tag{36}$$

Then premultiplying (33) by $(P'P)^{-1}P'$ yields a k^- of the desired form. ∎

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

$$\Gamma_m = \begin{pmatrix} \gamma_0 & \cdots & \cdots & \gamma_{m-1} \\ \vdots & \gamma_0 & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma'_{m-1} & \cdots & \cdots & \gamma_0 \end{pmatrix}, \tag{37}$$

(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 (population) second moments of (z_t) [18]. For singular AR systems, things are more subtle.

Consider the following Yule-Walker equations (see [17] pages 326-327):

$$(e_1, \ldots, e_p)\Gamma_p = (\gamma_1, \ldots, \gamma_p) \tag{38}$$

$$\Sigma_\nu = \gamma_0 - (e_1, \ldots, e_p)(\gamma_1, \ldots, \gamma_p)' \tag{39}$$

Formula (38) may be used to determine $(e_1, \ldots, e_p)$. Note that in the case $q < r$, as opposed to the regular case $r = q$, the matrix Γ_{p+1} will be singular and the matrix Γ_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 for $(e_1, \ldots, e_p)$ may not be unique. However, by the projection theorem, every solution determines the same $z_{t|t-1}$ and ν_t.

The possible nonuniqueness of the solutions of the (generalized) 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 [18]). We obtain the following result (in which we denote by $\delta(e(z))$ the degree of the polynomial matrix $e(z)$):

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

$$e(z)z_t = f\varepsilon_t, f \in \mathbb{R}^{r\times q} \tag{40}$$

where (ε_t) is white noise with $\mathbb{E}[\varepsilon_t\varepsilon_t'] = I_q$ and where $e(z)$ and f are relatively left prime.

(ii) Let $(e(z), f)$ be relatively left prime; then the class of all observationally equivalent $(\bar{e}(z), \bar{f})$ satisfying the degree restrictions $\delta(\bar{e}(z)) \le p$ and $\delta(\bar{f}) = 0$, is given by

$$\left(\bar{e}(z), \bar{f}\right) = u(z)(e(z), f) \tag{41}$$

where the polynomial matrix $u(z)$ satisfies

$$\det u(z) \neq 0, |z| \le 1 \tag{42}$$

$$u(0) = I \tag{43}$$

$$\delta(u(z)e(z)) \le p \tag{44}$$

$$\delta(u(z)f) = 0 \tag{45}$$

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

(iii) Let $(e(z), f)$ be relatively left prime; then $e(z)$ is unique if and only if $\operatorname{rk}(e_p, f) = r$ holds.

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

IV. The Yule Walker Equations

In this section we analyse the (generalized) Yule Walker equations from the point of view of estimation of AR systems. We propose in Theorem 5 a solution procedure for the case where the population second moments (37) are singular and show in Theorem 6 stability of the estimated AR systems. Let $\hat{z}_t$ be a consistent estimator of z_t (for $N, T \to \infty$). Let $\hat{\gamma}_j = \sum_{t=1}^{T} \hat{z}_{t+j}\hat{z}_t'$ and let $\hat{\Gamma}_p$ be defined analogously. Finally define $\hat{\Gamma}_p^s = O_s\Lambda_s O_s'$ where Λ_s consists of the s largest eigenvalues, $O_s \in \mathbb{R}^{pr\times s}$, $\Lambda_s \in \mathbb{R}^{s\times s}$.

Theorem 5: If $\operatorname{rk}\Gamma_p = s < pr$ holds, then the Yule Walker estimator for the solution of (38) defined by

$$(\hat{e}_1, \ldots \hat{e}_p) = (\hat{\gamma}_1, \ldots, \hat{\gamma}_p)O_s\Lambda_s^{-1}O_s' \tag{46}$$

is a continuous function of $\hat{\gamma}_0, \ldots, \hat{\gamma}_p$ if all eigenvalues of $\hat{\Gamma}_p$ are distinct.

Proof: As the eigenvalues and the corresponding suitable normalized eigenvectors of a symmetric matrix are continuous functions of the entries of the matrix, the right side of (46) is obviously a continuous function of the entries of $\hat{\Gamma}_p$. The condition that all eigenvalues of $\hat{\Gamma}_p$ are distinct is imposed for convenience and can be generalized. ∎

Remark 1: Even if Γ_p is singular, "typically" the corresponding sample second moment, $\hat{\Gamma}_p$ say, will not be singular. Nevertheless a truncation, setting the "small" eigenvalues of $\hat{\Gamma}_p$ equal to zero will give a numerically more reliable procedure.

Theorem 6: (i) If $\operatorname{rk}\Gamma_p = pr$ holds, then the Yule Walker estimator corresponding to (38) (i.e. when the γ_j in (38) are replaced by $\hat{\gamma}_j$) yields a stable autoregression

(ii) For $\operatorname{rk}\Gamma_p = s < pr$, the solution (46) corresponds to a stable autoregression

Proof: (i) As we need to show $\det\hat{e}(z) \neq 0$, $|z| \le 1$ we proceed as follows: Through row and column interchanges of

Γ_p we rewrite (38) as $\tilde{\Gamma}_p(e_p,\ldots,e_1)' = (\gamma_p,\ldots,\gamma_1)'$ with

$$\tilde{\Gamma}_p = \begin{pmatrix} \gamma_0 & \cdots & \cdots & \gamma'_{p-1} \\ \vdots & & & \vdots \\ \vdots & \gamma_0 & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \cdots & \gamma_0 \end{pmatrix}.$$

As can be easily seen, $\det \hat{e}(z) \neq 0$, $|z| \leq 1$ is equivalent to $\lambda_{max}(\hat{E}) < 1$ with

$$\hat{E} = \begin{pmatrix} 0 & \cdots & \cdots & 0 & \hat{e}'_p \\ I_r & 0 & \cdots & 0 & \hat{e}'_{p-1} \\ & \ddots & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & & \\ \vdots & \cdots & I_r & 0 & \hat{e}'_2 \\ 0 & \cdots & 0 & I_r & \hat{e}'_1 \end{pmatrix}$$

or equivalent to postulate that the roots of

$$\det(\hat{E} - zI_{rp}) \tag{47}$$

are within the unit circle. Define Z = $\begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \hat{z}'_1 \\ \vdots & & & \ddots \\ \vdots & & & \hat{z}'_2 \\ & & \ddots & \vdots \\ \hat{z}'_1 & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{z}'_{T-p+1} & \cdots & \hat{z}'_{T-1} & \hat{z}'_T \\ \vdots & & \ddots & \\ \hat{z}'_T & 0 & \cdots & 0 \end{pmatrix}$ and Y = $\begin{pmatrix} 0 & \cdots & 0 & \hat{z}'_1 \\ \vdots & & & \ddots \\ \vdots & & & \hat{z}'_2 \\ & & \ddots & \vdots \\ \hat{z}'_1 & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \hat{z}'_{T-p+1} & \cdots & \hat{z}'_{T-1} & \hat{z}'_T \\ \vdots & & \ddots & \\ \hat{z}'_T & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}$ Considering the "thin" singular value decompositions of $Z = U_1\Sigma_1V_1'$ and $Y = U_2\Sigma_2V_2'$ note that $\Sigma_1 = \Sigma_2 =: \Sigma$ and $V_1 = V_2 =: V$ can be choosen because of $Z'Z = Y'Y$. Furthermore $U_1 = (0, U')'$ and $U_2 = (U', 0)'$ can be choosen because of the form of Z and Y. It is straightforward to show that $\hat{E} = (Z'Z)^{-1}Z'Y = V\Sigma^{-1}U_1'U_2\Sigma V'$ holds. The roots of (47) are the same as the roots of $\det[\Sigma V'(\hat{E} - zI_{rp})V\Sigma^{-1}] = \det(U_1'U_2 - zI_{rp})$. Since one can show that a q with $\|q\| = 1$ exists which fulfills $|\lambda_{max}(U_1'U_2)| = |q^*U_1'U_2q|$, finally the proof is completed by using the following straightforward inequality $|\lambda_{max}(U_1'U_2)| \leq max_{\|x\|=1}|x^*U_1'U_2x| < max_{\|x\|=1}|x^*U_1'U_1x| = 1$.

(ii) Defining Y and Z the same way as in (i) we set the smallest singular values of Y and Z, which are "typically" not zero (although $\operatorname{rk}\Gamma_p = s < rp$) to zero. So we get the "very thin" singular value decompositions of $\tilde{Z} = U_{1s}\Sigma_{1s}V_{1s}'$ and $\tilde{Y} = U_{2s}\Sigma_{2s}V_{2s}'$. As $Y'Y = Z'Z$ still holds we once again can choose $\Sigma_{1s} = \Sigma_{2s} =: \Sigma_s$, $V_{1s} = V_{2s} =: V_s$, $U_{1s} = (0, U_s')'$ and $U_{2s} = (U_s', 0)'$. As can be easily seen $VV_{-r}' = (0, I_{rp})$, where V_{-r}' is V' without its first r rows. Furthermore $\hat{e}' = (\hat{e}'_p, \ldots, \hat{e}'_1)' = V_s\Sigma_s^{-1}U_{1s}'U_{2s}\Sigma_sV_{s,-r(p-1)}'$ with $V_{s,-r(p-1)}'$ defined the same way as V_{-r}'. Finally, from the eigenvalues of $\tilde{\Sigma}V'\hat{E}V\tilde{\Sigma}^{-1}$, where $\tilde{\Sigma} = \begin{pmatrix} \Sigma_s & 0 \\ 0 & I_{rp-s} \end{pmatrix}$ and $\hat{E} = (VV_{-r}', \hat{e})$, it can be seen that $\lambda_{max}(\hat{E}) < 1$ holds. ■

A related theorem dealing with population as opposed to sample statistics will be included in an extended version of this paper.

REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

[6] 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.

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

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

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

[10] ——, "Generalized linear dynamic factor models a structure theory," *CDC*, 2008.

[11] ——, "Properties of zero-free spectral matrices," *forthcoming in IEEE Tr.AC*, 2009.

[12] D. Giannone, L. Reichlin, and L. Sala, "Monetary policy in real time," *NBER Macroeconomic Annual 2004*, 2004.

[13] J. H. Stock and M. W. Watson, "An empirical comparison of methods for forecasting using many predictors," *Mimeo, Princeton University*, 2004.

[14] M. Forni, D. Giannone, M. Lippi, and L. Reichlin, "Opening the black box: structural factor models versus structural vars," *Manuscript, Universite Libre de Bruxelles*, 2005.

[15] M. Forni, M. Hallin, M. Lippi, and L. Reichlin, "The generalized dynamic factor model: consistency and rates," *Journal of Econometrics*, vol. 119, pp. 231–255, 2004.

[16] Y. A. Rozanov, *Stationary Random Processes*. San Francisco: Holden-Day, 1967.

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

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

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

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

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

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

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