European Journal of Control (2010)3:1–14

DOI:10.3166/EJC.16.1–14

European Journal of Control

Generalized Linear Dynamic Factor Models – An Approach *via* Singular Autoregressions

M. Deistler[1,*], B.D.O. Anderson[2,**], A. Filler[1,***], Ch. Zinner[3,****], W. Chen[2,*****]

Department of Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria;
[2]Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia;
[3]Group of Market Risk, BAWAG P.S.K., Vienna, Austria

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 emphas is 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. The Yule–Walker equations do not necessarily have a unique solution in the singular case, and the resulting complexities are examined with a view to find a stable and coprime system.

Keywords: High Dimensional Time Series, Generalized Dynamic Factor Models, Singular AR System, (Generalized) Yule–Walker Equations

1. Introduction

Generalized linear dynamic factor models (GDFMs) have been introduced in [9, 12], and, in a slightly different form, in [23, 24]. The idea is to generalize and combine linear dynamic factor models with strictly idiosyncratic[1] noise as analyzed in [21] and [22] and generalized linear static factor models, introduced in [5] and [6]. 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 co-movement between the single time series. GDFMs are used, both for forecasting and for analysis of high dimensional time series (see e.g. [8, 13, 25]). In forecasting, the forecasts of the latent variables are used to forecast the observed variables.

The basic idea of GDFMs 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 on the cross-sectional dimension,

*Correspondence to: M. Deistler, E-mail: manfred.Deistler@tuwien.ac.at
**E-mail: brian.anderson@anu.edu.au
***E-mail: alexander.filler@tuwien.ac.at
****E-mail: c_zinner@gmx.at
*****E-mail: weitian.chen@anu.edu.au

[1] "Idiosyncratic" means specific for a particular univariate time series, as opposed to "co-moving". We use the term "strictly idiosyncratic" noise if the noise components are mutually uncorrelated in cross-section, i.e. if the noise spectrum is diagonal.

Received 14 August 2009; Accepted 22 December 2009
Recommended by A. Karimi, L. Ljungy

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

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

$$\mathbb{E}[\hat{y}_t^N u_s^{N\prime}] = 0 \quad \forall s, t \tag{3}$$

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 dynamic factors see (8) 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.

The following assumptions constitute the class of GDFMs considered here:

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 $(y_{it} \mid i \in \mathbb{N}, t \in \mathbb{Z})$, where i is the cross-sectional index and t denotes time. Therefore we consider a *sequence* of GDFMs (1), indexed by the cross-sectional dimensions N. We assume:

Assumption 2: *The double-indexed 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 of $f_{\hat{y}}^N$ is independent of N ($N \geq$ some N_0).*

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

Next, we define weak and strong dependence as in [12]. We use e.g. $\omega_{u,r}^N$ to denote the r-th largest eigenvalue of f_u^N.

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

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

A simple example for strong dependence is a spectral density $f_{\hat{y}}^N$ where every entry is equal to one. Clearly, such a matrix has rank 1 and its largest eigenvalue is equal to N.

Contrary to the strictly idiosyncratic case 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 [12], 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 N-indexed sequence of dynamic principal component decompositions

$$\begin{aligned} f_y^N(\lambda) = {} & O_1^N(e^{-i\lambda})\Omega_1^N(\lambda)O_1^N(e^{-i\lambda})^* \\ & + O_2^N(e^{-i\lambda})\Omega_2^N(\lambda)O_2^N(e^{-i\lambda})^* \end{aligned} \tag{5}$$

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 [12], such a sequence of PCA models, for $N \to \infty$ converges to the corresponding GDFMs in the sense that e.g. the scalar components of the latent PCA variables $\hat{y}_{PCA,t}^N = O_1^N(z)O_1^N(z)^* y_t^N$ converge to the respective scalar components of $\hat{y}_t^N$. From now on, unless the contrary is stated explicitly, for the sake of simplicity of notation, we will omit the superscript N. In general terms, *this paper is concerned with identification of GDFMs, where the latent variables have a singular rational spectral density, or to be more precise, with the identification of a stable 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 and based on this structure theory an estimation algorithm is proposed. It heavily draws from previous work ([2–4]). In the structure theory considered here an idealized setting is considered, as 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 [2], our emphasis is on the zeroless case, which in our setting is generic. We extend, the results given in the previous paper. The 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), and the analysis of (generalized) Yule–Walker equations to obtain a singular autoregression. In the singular case, the Yule–Walker equations do not necessarily have a unique solution, and we must pay more attention to the nontrivial task of selecting a solution with properties such as stability and minimality.

The paper is organized as follows: In section 2, we review 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 also reviewed. Section 3 is concerned with zeroless transfer functions and their realization by (possibly singular) autoregressive systems. Section 4 is concerned with stable solutions of (generalized) Yule–Walker equations. As 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.

2. Realization of Rational Tall Transfer Functions

In this and the next section, we commence from the population spectral density $f_{\hat{y}}$ of the latent variables $(\hat{y}_t)$ rather than from data. This is justified as such an analysis gives valuable insights and as the effect of the wide sense idiosyncratic noise can be removed for letting N and T going to infinity at suitable rates. The latter has been shown in [9, 10, 23] and also follows from the discussion in Section 4.

2.1. Spectral Factorization and Wold Decomposition

We have the following result [14, 20].

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

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

where $w(z)$, $z \in \mathbb{C}$ is a $N \times q$ real rational matrix of full column rank 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

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

correspond to a causal linear finite dimensional system

$$\hat{y}_t = \sum_{j=0}^{\infty} w_j \epsilon_{t-j} \tag{8}$$

where the inputs (ϵ_t) are white noise with $\mathbb{E}[\epsilon_t \epsilon_t'] = 2\pi I_q$. This input process (ϵ_t) is a minimal dynamic factor. 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

$$w = udv \tag{9}$$

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 finite zeros of w are the finite zeros of the n_i and the poles of w are the zeros of the d_i. Note that $w(z)$ has no poles and no zeros for $|z| \leq 1$.

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

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

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 by

$$\epsilon_t = \sum_{j=0}^{\infty} w_j^- \hat{y}_{t-j} \tag{11}$$

Thus (8) corresponds to a Wold decomposition (see, e.g. [15]).

2.2. ARMA Representation

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

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

We assume that (a, b) are left coprime (see e.g. [15]); then the set of all observationally equivalent left coprime ARMA systems is obtained as (ua, ub) where u is an arbitrary unimodular matrix. Note that for such a coprime ARMA system the zeros of $w(z)$ correspond to the zeros of $b(z)$ and the poles of $w(z)$ are the zeros of

$\det a(z)$. Thus the conditions stated in Theorem 1 on the poles and zeros of the transfer function $w = a^{-1}b$ are, for left coprime a, b, equivalent to

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

and

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

A right MFD

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

where d and c are polynomial matrices of appropriate dimension, corresponds to the cascade of an (nonsingular) AR system followed by a finite impulse response system, and has been used in [12]. We do not further explore such representations in this paper.

2.3. State Space Realization

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

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

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

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 [15]; 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) below has no zeros for $|z| \leq 1$. The transfer function for (16) and (17) is given by

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

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

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

has rank less than $n + q$ at z_0. Starting with the power series expansion (7), the form (16) and (17) can be obtained by the "Akaike–Kalman procedure" [1] from the equation

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

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

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

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

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

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

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

2.4. Static Factors

A *static factor* of the latent variables $(\hat{y}_t)$ is a process (z_t) of dimension lesser than or equal to n, 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 $\mathbb{E}[\hat{y}_t\hat{y}_t'] = L\mathbb{E}[z_t z_t']L'$, 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, by considering $\mathbb{E}[(\hat{y}_t - Mz_t)(\hat{y}_t - Mz_t)']$, 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. Because x_t has a nonsingular covariance, the fact that $\text{rk}\,\mathbb{E}[\hat{y}_t\hat{y}_t'] = r$ means that $\text{rk}\,H_1 = r$ and H_1 has r columns. A particular minimal static factor is then defined by

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

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 $\text{rk}H = n$. A particular minimal static factor can be obtained from the echelon form (22) by selecting the first r linearly independent components of $\hat{y}_t$

$$z_t = S_1\hat{Y}_t. \tag{29}$$

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

Note that $n = r$ holds if and only if $\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \ldots$ do not contain further linearly independent components and x_t is a minimal static factor if and only if all Kronecker indices of $\mathcal{H}$ (see [15], chapter 2) are equal to zero or one.

Any 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,

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

$$y_t = Mk(z)\epsilon_t = w(z)\epsilon_t \tag{32}$$

for some $k(z)$ transfer function corresponding to the choice of the static factor. 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 from a minimal $\{F, G, H\}$ as $\{F, G, C\}$ and conversely, by

$$C = (M'M)^{-1}M'H \tag{33}$$

$$H = MC \tag{34}$$

Because z_t has the same dynamics as $\hat{y}_t$ and is of smaller dimension independent of N, modeling of (z_t) is more convenient. As (z_t) is an ARMA process, ARMA identification procedures, such as the autoregression-regression approach [18, 19] may be applied to this case, as has been done in [27]. Here, however, we will be able to deal with an autoregressive approach, which is simpler and can be applied in a generic situation.

3. Zeroless Transfer Functions and Autoregressive Systems

Of particular interest for us are zeroless transfer functions, because, as reviewed below, tall rational 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, as 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 also be represented by an autoregression, and again the variance matrix of the driving white noise may be singular. 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, a fact which generates additional complexity.

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 [2]:

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

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 detailed proof is given in [2].

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

Theorem 3: *Let $(\hat{y}_t)$ satisfy Assumptions 1–4 and let z_t be an associated minimal static factor, of dimension r; 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^- corresponding to (10) and thus the input ϵ_t in (31) is determined from a finite number of outputs $z_t, z_{t-1}, \ldots, z_{t-L}$, for some L*
(iii) *(z_t) is a stationary solution of a stable AR system*

$$z_t = e_1 z_{t-1} + \cdots + e_p z_{t-p} + \nu_t;\ e_i \in \mathbb{R}^{r \times r} \tag{35}$$

where

$$\det \underbrace{(I - e_1 z - \cdots - e_p z^p)}_{e(z)} \neq 0,\ |z| \leq 1$$

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

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

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

where $\tilde{e}, f$ are relatively left prime and $\tilde{e}$ is stable. As $k(z) = \tilde{e}^{-1}(z) f(z)$ is zeroless, the same holds for $f(z)$. As is well known every zeroless tall polynomial matrix f can be completed by a suitable choice of a polynomial matrix g to a unimodular matrix $u = (f, g)$ by extending the Smith–McMillan form of $f = \tilde{u}\tilde{d}\tilde{v}$ to

$$(f, g) = \tilde{u}\left(\tilde{d}, \begin{pmatrix} 0 \\ I \end{pmatrix}\right)\begin{pmatrix} \tilde{v} & 0 \\ 0 & I \end{pmatrix}$$

Then

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

and

$$u^{-1}(z) \tilde{e}(z) z_t = \begin{pmatrix} \epsilon_t \\ 0 \end{pmatrix} \tag{37}$$

gives an autoregressive representation, and premultiplying (37) by $\tilde{e}^{-1}(0) u(0)$ gives the desired form (35). The stability of $e(z)$ follows from the stability of $\tilde{e}(z)$.

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

$$\Sigma_\nu = PP', P \in \mathbb{R}^{r \times q}, \operatorname{rk}\ P = q \tag{38}$$

Then premultiplying (35) 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

$$\Gamma_m = \begin{pmatrix} \gamma_0 & \cdots & \cdots & \gamma_{m-1} \\ \vdots & \gamma_0 & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{m-1} & \cdots & \cdots & \gamma_0 \end{pmatrix}, \tag{39}$$

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) [14]. For singular AR systems, as we may have, the situation is more subtle.

Consider the following Yule–Walker equations (see [14] pages 326–327):

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

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

Formula (40) 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 predictor $z_{t|t-1}$ 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 [15]). 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*

$$e(z)z_t = f\epsilon_t, f \in \mathbb{R}^{r\times q} \tag{42}$$

where (ϵ_t) *is white noise with* $\mathbb{E}[\epsilon_t\epsilon_t'] = I_q$, $e(0) = I$ *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* $\delta(\bar{e}(z)) \leq p$, $e(0) = I$ *and* $\delta(\bar{f}) = 0$, *is given by*

$$(\bar{e}(z), \bar{f}) = u(z)(e(z), f) \tag{43}$$

where the polynomial matrix $u(z)$ *satisfies*

$$\det u(z) \neq 0, |z| \leq 1 \tag{44}$$

$$u(0) = I \tag{45}$$

$$\delta(u(z)e(z)) \leq p \tag{46}$$

$$\delta(u(z)f) = 0 \tag{47}$$

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* $e(z)$ *with* $e(0) = I$ *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. ■

4. The Yule–Walker Equations

As the set of the solutions of (40) is a nontrivial affine space in the case of Γ_p being singular, we want to study the properties of certain solutions. In particular, we want to ensure that we can pick a solution which will provide a stable system. Should such a solution not be coprime, then we need to be able to determine from it a coprime stable solution. We distinguish between population results and sample results.

Population Results

In this section, we examine the extent to which the solution of the (generalized) Yule–Walker equation (40) automatically gives a stable matrix polynomial.

For future reference, let us define the block companion matrix of the polynomial $e(z) = I - e_1 z - \ldots - e_p z^p$ of (35)

$$E = \begin{pmatrix} e_1 & e_2 & \cdots & e_{p-1} & e_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & I & 0 \end{pmatrix} \tag{48}$$

and

$$Q = \operatorname{diag}(\Sigma_\nu, 0, \ldots, 0) \tag{49}$$

Note that the zeros of $e(z)$ (i.e. the roots of $det(e(z))$) correspond to the inverted eigenvalues of E so that by the zero condition of (35), E has all eigenvalues inside the unit circle. Note further that (as can be checked by straightforward calculations using (39) through (41)) the following discrete-time Lyapunov equation holds

$$\Gamma_p - E\Gamma_p E' = Q \tag{49}$$

Observe that for any solution of the (generalized) Yule–Walker equation (40) $(\bar{e}_1\bar{e}_1 \ldots \bar{e}_p)$, if we define $\bar{E}$ as E with e_i replaced by $\bar{e}_i$, then (49) holds with $\bar{E}$ replacing E.

Before stating the main result, we prove two lemmas.

Lemma 2: *Let* A *be a square real matrix. Let* $\lambda(A)$ *be an arbitrary eigenvalue of* A. *Then there holds*

$$|\lambda(A)| \leq \max_{||x||=1} |x^* A x|$$

Proof: Let y be a unit length eigenvector corresponding to $\lambda(A)$. There holds $Ay = \lambda(A)y$ and therefore

$$|\lambda(A)| = |y^*\lambda(A)y| = |y^*Ay| \leq \max_{||x||=1} |x^*Ax|$$

Lemma 3: *Let A be a matrix defined by*

$$A = (0_{a\times b}\ I_a)T\begin{pmatrix} 0 & 0 \\ I_c & 0 \end{pmatrix}T'\begin{pmatrix} 0_{b\times a} \\ I_a \end{pmatrix}$$

with T an orthogonal matrix. Then all eigenvalues of A have magnitude less than 1.

Proof: Observe first that if x is a unit norm vector of dimension a, then (with obvious definition of U)

$$y = T'\begin{pmatrix} 0_{b\times a} \\ I_a \end{pmatrix}x = Ux$$

also has unit norm, as easily can be seen. Evidently,

$$\max_{||x||=1} |x^*Ax| = \max_{||y||=1, y=Ux} |y^*\begin{pmatrix} 0 & 0 \\ I_c & 0 \end{pmatrix}y|$$
$$\leq \max_{||y||=1} |y^*\begin{pmatrix} 0 & 0 \\ I_c & 0 \end{pmatrix}y|$$

Now observe that for any unit norm y,

$$|y^*\begin{pmatrix} 0 & 0 \\ I_c & 0 \end{pmatrix}y| = |(y_1^*\ y_2^* \dots y_{a+b}^*)\begin{pmatrix} 0 \\ \vdots \\ 0 \\ y_1 \\ y_2 \\ \vdots \\ y_c \end{pmatrix}|$$

By the Cauchy-Schwarz inequality, we see that the right side of this equation is bounded by the product of $||y||$, which is 1, and $||(y_1'\ y_2' \dots y_c'\ 0 \dots 0)||$; the second of these two norms is less than 1 unless $y_{c+1} = y_{c+2} = \dots y_{a+b} = 0$. But if it happens that $y_{c+1} = y_{c+2} = \dots y_{a+b} = 0$ then the Cauchy-Schwarz inequality can only hold with equality in case all y_i are zero, which is a contradiction. Hence we know that the Cauchy-Schwarz inequality is strict.

Consequently, we have

$$\max_{||x||=1} |x^*Ax| \leq \max_{||y||=1} |y^*\begin{pmatrix} 0 & 0 \\ I_c & 0 \end{pmatrix}y| < 1$$

and the result follows by Lemma 2. ■

Theorem 5: *Let $(\hat{y}_t)$ satisfy Assumptions 1–4, let (z_t) be a minimal static factor satisfying condition (iii) of Theorem 3. Let Γ_m be as in* (39) *and let $\bar{e}_i, i = 1, \dots, p$ denote the solution of the Yule-Walker equation (40) defined by*

$$(\bar{e}_1, \bar{e}_2, \dots, \bar{e}_p) = (\gamma_1, \gamma_2, \dots, \gamma_p)\Gamma_p^{\#} \tag{50}$$

where the superscript # denotes the matrix pseudo inverse. Then the system defined by the $\bar{e}_i$ is stable, and there are s eigenvalues of $\bar{E}$, the block companion matrix associated with the $\bar{e}_i$, which are identical with eigenvalues of E defined in (48).

Proof: Let O_p denote an orthogonal matrix such that

$$O_p\Gamma_pO_p' = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$$

with Λ_1 diagonal and nonsingular. For the purposes of the proof, most of our calculations will be carried out in a changed coordinate basis defined by O_p. Accordingly, define

$$(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p) = (\bar{e}_1, \bar{e}_2, \dots, \bar{e}_p)O_p'$$
$$(f_1, f_2, \dots, f_p) = (e_1, e_2, \dots, e_p)O_p'$$

Since

$$\begin{aligned}(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p)O_p\Gamma_pO_p' &= (\bar{e}_1, \bar{e}_2, \dots, \bar{e}_p)\Gamma_pO_p' \\ &= (\gamma_1, \gamma_2, \dots, \gamma_p)O_p' \\ &= (e_1, e_2, \dots, e_p)\Gamma_pO_p' \\ &= (f_1, f_2, \dots, f_p)O_p\Gamma_pO_p'\end{aligned}$$

or equivalently

$$(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p)\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} = (f_1, f_2, \dots, f_p)\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$$

we see that the first s columns of the block row matrix $(f_1, f_2, \dots, f_p)$ are identical with those of $(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p)$. Also, we can argue that the last $rp - s$ columns are 0:

$$(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p) = (\gamma_1, \gamma_2, \dots, \gamma_p)\Gamma_p^{\#}O_p'$$
$$= (\gamma_1, \gamma_2, \dots, \gamma_p)O_p'\begin{pmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Now consider the block companion matrices $\bar{E}, E$ defined by the $(\bar{e}_i), (e_i)$, together with their transforms

$$\bar{F} = O_p\bar{E}O_p'$$
$$F = O_pEO_p'$$

Partition these two matrices in the same manner as the right side of $O_p\Gamma_pO_p'$, so that $\bar{F}_{11}, F_{11}$ are $s \times s$:

$$\bar{F} = \begin{pmatrix} \bar{F}_{11} & \bar{F}_{12} \\ \bar{F}_{21} & \bar{F}_{22} \end{pmatrix}$$

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}$$

Now the matrix F satisfies the following transformed version of (49):

$$\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} - F\begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}F' = O_p Q O_p'$$

and similarly for $\bar{F}$. The 22 block term on the left side is $-F_{21}\Lambda_1 F_{21}'$ while the 22 block term on the right is nonnegative definite. It follows that F_{21} is zero. Likewise, $\bar{F}_{21}$ is zero, and so $\bar{F}, F$ are both upper triangular. Now consider

$$F - \bar{F} = O_p(E - \bar{E})O_p'$$

$$= O_p \begin{pmatrix} e_1 - \bar{e}_1 & e_2 - \bar{e}_2 & \dots & e_{p-1} - \bar{e}_{p-1} & e_p - \bar{e}_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} O_p'$$

$$= O_p \begin{pmatrix} f_1 - \bar{f}_1 & f_2 - \bar{f}_2 & \dots & f_{p-1} - \bar{f}_{p-1} & f_p - \bar{f}_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

We have shown above that the first s columns of $(f_1, f_2, \dots f_p)$ are identical with those of $(\bar{f}_1, \bar{f}_2, \dots \bar{f}_p)$, and so the first s columns of the matrix on the right of the above equation are zero. This means that the first s columns of $\bar{F}$ and F are the same, i.e. $\bar{F}_{11} = F_{11}$. Since E and therefore F has all eigenvalues in $|\lambda| < 1$, the same is true of $F_{11} = \bar{F}_{11}$. We now have to examine the last $rp - s$ columns of $\bar{F}$, and in particular the last $rp - s$ rows of these columns; this is because $\bar{F}$ is block triangular, and the lower triangular block has yet to be proved to have eigenvalues in $|\lambda| < 1$. Observe that

$$\bar{F} = O_p \bar{E} O_p' = O_p \begin{pmatrix} \bar{e}_1 & \bar{e}_2 & \dots & \bar{e}_{p-1} & \bar{e}_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix} O_p'$$

$$= O_p \begin{pmatrix} \bar{f}_1 & \bar{f}_2 & \dots & \bar{f}_{p-1} & \bar{f}_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} + O_p \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix} O_p'$$

Now our interest is in the last $rp - s$ columns, and the entries of the last $rp - s$ columns of $(\bar{f}_1\, \bar{f}_2 \dots \bar{f}_{p-1}\, \bar{f}_p)$ have all been shown to be zero. Since what is in the first s columns is immaterial, we can say that the last $rp - s$ columns of $\bar{F}$ are actually identical with the last $rp - s$ columns of the matrix

$$O_p \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix} O_p'$$

Thus

$$\bar{F}_{22} = (0_{(rp-s)\times s}\; I_{rp-s}) O_p \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}$$

$$O_p' \begin{pmatrix} 0_{s\times(rp-s)} \\ I_{rp-s} \end{pmatrix}$$

By the previous lemma, we know that $|\lambda_i(\bar{F}_{22})| < 1$ for every eigenvalue of $\bar{F}_{22}$. In summary,

$$\bar{F} = \begin{bmatrix} F_{11} & \bar{F}_{12} \\ 0 & \bar{F}_{22} \end{bmatrix} \qquad ()$$

where $\bar{F}_{12}$ is irrelevant, F_{11} has eigenvalues of magnitude less than 1 by hypothesis and the triangularity of F, and $\bar{F}_{22}$ has just been proved to have the same property. ■

4.1. Sample Results

In this section, we briefly discuss *estimation* in the framework of GDFMs. We do not intend to give a detailed account here, this will be done in a subsequent paper. The purpose of this subsection is twofold: First to shortly describe an estimation procedure, in order to demonstrate the usefulness of the "structural" results, described above. Second to analyze the (generalized) Yule–Walker equations for the case where the population second moments have been replaced by sample counterparts.

As has been shown in [11] and [23] a consistent estimator $\hat{z}_t$ of z_t can be obtained by a static PCA of y_t as follows: Consider the eigenvalue decomposition

$$T^{-1}\sum_{t=1}^{T} y_t^N y_t^{N'} = O_1^N \Lambda_1^N O_1^{N'} + O_2^N \Lambda_2^N O_2^{N'} \qquad (51)$$

where Λ_1^N is the diagonal $r \times r$ matrix, whose diagonal elements are the largest r eigenvalues of the matrix on the l.h.s. and O_1^N is the $N \times r$ matrix of the corresponding eigenvectors. Then define

$$\hat{z}_t^N := N^{-1/2} O_1^{N'} y_t^N \qquad (52)$$

Now the following can be shown: Under the additional assumptions that the r largest eigenvalues of $\mathbb{E}[y_t^N y_t^{N'}]$

diverge to infinity, for $N \to \infty$, and the other eigenvalues are bounded, the estimators $\hat{z}_t^N$ converge for $N, T \to \infty$ to a suitably normalized version of z_t. Let $\hat{\gamma}_j = T^{-1} \sum_{t=1}^{T-j} \hat{z}_{t+j}\hat{z}_t'$, $j \geq 0$ and let $\hat{\Gamma}_p$ be defined analogously. Note that the $\hat{\gamma}_j$ are not the usual sample covariances corresponding to γ_j, because the z_t are not directly observed. Note also, that, if the rank of Γ_p is equal to $s < rp$ holds, then "typically" $\hat{\Gamma}_p$ will be of rank pr. Nevertheless, in such a case, for the purpose of "regularization" a truncation setting the "small" eigenvalues of $\hat{\Gamma}_p$ equal to zero will give a numerically stable procedure. Define $\hat{\Gamma}_p^s = O_p^s \Lambda_p^s O_p^{s'}$ where $\Lambda_p^s \in \mathbb{R}^{s \times s}$ is a diagonal matrix consisting of the s largest eigenvalues of $\hat{\Gamma}_p$, and $O_p^s \in \mathbb{R}^{pr \times s}$ is the matrix of the corresponding eigenvectors. The next theorems show that the minimal norm solution of eventually truncated sample Yule–Walker equations yields consistent estimators corresponding to stable autoregressions.

Theorem 6: *If* $\mathrm{rk}\,\Gamma_p = s < pr$ *and if all nonzero eigenvalues of* Γ_p *are distinct, then the minimal norm solution of the Yule–Walker equations corresponding to* (40) *defined by*

$$(\hat{e}_1, ...\hat{e}_p) = (\hat{\gamma}_1, ..., \hat{\gamma}_p) O_p^s (\Lambda_p^s)^{-1} O_p^{s'} \tag{53}$$

defines a function of $\hat{\gamma}_0, ..., \hat{\gamma}_p$ *which is continuous at* $\gamma_0, ..., \gamma_p$ *(here the* $\hat{\gamma}_j$ *as well as the* $\hat{e}_j$ *are considered to be matrices with real entries).* ■

Proof: As the eigenvalues and the corresponding suitable normalized eigenvectors of a symmetric matrix are locally continuous functions of the entries of the matrix, the right side of (53) is obviously a continuous function at $\gamma_0, ..., \gamma_p$. ■

Remark 1: *Under suitable assumptions the sample covariances* $\hat{\gamma}_j$ *are consistent estimators of the population counterparts* γ_j*. Therefore the continuity result of Theorem 6 implies that the* $(\hat{e}_1, ...\hat{e}_p)$ *are consistent estimators of the minimal norm solution* (50) *of the Yule–Walker equations* (40).

Remark 2: *The condition that all nonzero eigenvalues of* $\hat{\Gamma}_p$ *are distinct is imposed for convenience and can be relaxed.*

Theorem 7:

(i) *If* $rk\,\Gamma_p = pr$ *holds, then the Yule–Walker estimator corresponding to* (40) *(i.e. when the* γ_j *in* (40) *are replaced by* $\hat{\gamma}_j$*) yields a stable autoregression*

(ii) *For* $rk\,\Gamma_p = s < pr$*, the solution* (53) *corresponds to a stable autoregression*

Proof: (i) As we need to show $\det \hat{e}(z) \neq 0$, $|z| \leq 1$ we proceed as follows: Let

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

then $\det \hat{e}(z) \neq 0$, $|z| \leq 1$ is equivalent to postulating that the roots of

$$\det(\hat{E} - zI_{rp}) \tag{54}$$

are within the unit circle. Define the $(T+p) \times rp$ matrices

$$Z = \begin{pmatrix} 0 & \cdots & & 0 \\ \hat{z}_1' & 0 & \cdots & 0 \\ \hat{z}_2' & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \hat{z}_1' \\ \vdots & \vdots & \vdots & \vdots \\ \hat{z}_T' & \hat{z}_{T-1}' & \cdots & \hat{z}_{T-p+1}' \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \hat{z}_T' & \hat{z}_{T-1}' \\ 0 & \cdots & 0 & \hat{z}_T' \end{pmatrix} \tag{55}$$

and

$$Y = \begin{pmatrix} \hat{z}_1' & 0 & \cdots & 0 \\ \hat{z}_2' & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \hat{z}_1' \\ \vdots & \vdots & \vdots & \vdots \\ \hat{z}_T' & \hat{z}_{T-1}' & \cdots & \hat{z}_{T-p+1}' \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \hat{z}_T' & \hat{z}_{T-1}' \\ \vdots & \cdots & 0 & \hat{z}_T' \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

Consider the "thin" singular value decompositions[2] $Z = U_1 \Sigma_1 V_1$ and $Y = U_2 \Sigma_2 V_2'$. Note that $\Sigma_1 = \Sigma_2 =: \Sigma$ and $V_1 = V_2 =: V$ can be chosen because $Z'Z = Y'Y$ holds. Furthermore $U_1 = (0, U')'$ and $U_2 = (U', 0)'$ can be chosen because of the form of Z and Y. It is straightforward to show that

$$\hat{E} = Y'Z(Z'Z)^{-1}$$

[2] The singular value decomposition of $Y \in \mathbb{R}^{m \times h}, m \geq h$ is defined as $Y = U\Sigma V'$ with $U \in \mathbb{R}^{m \times m}, \Sigma \in \mathbb{R}^{m \times h}, V \in \mathbb{R}^{h \times h}$, whereas the "thin" singular value decomposition of $Y \in \mathbb{R}^{m \times h}, m \geq h$ is defined as $Y = U\Sigma V'$ with $U \in \mathbb{R}^{m \times h}, \Sigma \in \mathbb{R}^{h \times h}, V \in \mathbb{R}^{h \times h}$.

and therefore

$$\hat{E} = V\Sigma U_2' U_1 \Sigma^{-1} V'$$

holds. The roots of (54) are the same as the roots of

$$\det(\Sigma^{-1} V'(\hat{E} - zI_{rp})V\Sigma) = \det(U_2' U_1 - zI_{rp})$$

Now we have

$$|\lambda_{max}(U_2' U_1)| \leq max_{\|x\|=1}|x^* U_2' U_1 x| < max_{\|x\|=1}|x^* U_1' U_1 x| = 1$$

where the first inequality holds due to Lemma 2 and the strict inequality is valid because of the form of $U_1 = (0, U')'$ and $U_2 = (U', 0)'$ (compare with the proof of Lemma 3, strict Cauchy–Schwarz inequality).

(ii) First we want to repeat that, as is well known, the sample covariance matrix $\hat{\Gamma}_p$ is "typically" non-singular. We assume that we know the rank s of the true matrix Γ_p and set the smallest singular values of Y and Z which are "typically" not zero to zero: Using an evident notation we define

$$\tilde{Z} = (U_{1s}, U_x)\begin{bmatrix} \Sigma_{1s} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_{1s}' \\ V_x' \end{bmatrix} = U_{1s}\Sigma_{1s}V_{1s}' \tag{56}$$

where

$$Z = U_1 \Sigma V' = (\underbrace{U_{1s}}_{(T+p)\times s}, \underbrace{U_x}_{(T+p)\times(rp-s)})\begin{bmatrix} \Sigma_{1s} & 0 \\ 0 & \Sigma_x \end{bmatrix}\begin{bmatrix} V_{1s}' \\ V_x' \end{bmatrix} \tag{57}$$

We call $\tilde{Z} = U_{1s}\Sigma_{1s}V_{1s}'$ the "very thin" singular value decomposition of $\tilde{Z}$ (note $s < rp$) and define $\tilde{Y}$ analogously, i.e. $\tilde{Y} = U_{2s}\Sigma_{2s}V_{2s}'$. Then, as $Y'Y = Z'Z$ holds, we have $\tilde{Y}'\tilde{Y} = \tilde{Z}'\tilde{Z}$ and thus we can choose $\Sigma_{1s} = \Sigma_{2s} =: \Sigma_s$, $V_{1s} = V_{2s} =: V_s$, $U_{1s} = (0, U_s')'$ and $U_{2s} = (U_s', 0)'$. From now on, for the simplicity of notation, we us Z and Y for $\tilde{Z}$ and $\tilde{Y}$, respectively. As can be easily seen $V_{[,-r]}V' = (I_{r(p-1)}, 0)$, where $V_{[,-r]}$ is V is defined by omitting the last r rows in V from (56). Now observe that

$$(Y')_{[,-r(p-1)]}Z = (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s V_s' \tag{58}$$

holds, which is an estimate of $T(\gamma_1, \ldots, \gamma_p)'$ (note that if $\Sigma_x \neq 0$ holds, then (58) does not correspond to the sample covariances $T(\hat{\gamma}_1, \ldots, \hat{\gamma}_p)'$). It follows that

$$\begin{aligned}
\hat{e} &= (\hat{e}_1, ..., \hat{e}_p) \\
&= (Y)_{[,-r(p-1)]}' Z(Z'Z)^{\#} \\
&= (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s V_s' V_s \Sigma_s^{-2} V_s' \\
&= (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s \Sigma_s^{-2} V_s' \\
&= (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s^{-1} V_s'
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{E} &= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s^{-1} V_s' \\ V_{[,-r]}V' \end{pmatrix} \\
&= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s^{-1} V_s' \\ ((V_s)_{[,-r]}, (V_x)_{[,-r]})\begin{bmatrix} V_s' \\ V_x' \end{bmatrix} \end{pmatrix} \\
&= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s^{-1} V_s' \\ (V_s)_{[,-r]}V_s' + (V_x)_{[,-r]})V_x' \end{pmatrix} \\
&= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U_{2s}' U_{1s}\Sigma_s^{-1} V_s' \\ (V_s)_{[,-r]}V_s' \end{pmatrix} \\
&\quad + \begin{pmatrix} 0 \\ (V_x)_{[,-r]}V_x' \end{pmatrix}
\end{aligned}$$

holds. Call the first summand on the r.h.s. in the last equality above B and the second summand $\tilde{V}$. Recall that we want to show that the roots of $det(\hat{E} - zI_{rp})$ are within the unit circle. Now we have (with the penultimate equality following using (59) and (60) below)

$$\begin{aligned}
det(\hat{E} - zI_{rp}) &= det(B + \tilde{V} - zI_{rp}) = det(V'(B + \tilde{V} - zI_{rp})V) \\
&= det\left(\overbrace{\begin{pmatrix} \Sigma_s^{-1} & 0 \\ 0 & I_{rp-s} \end{pmatrix}}^{\tilde{\Sigma}^{-1}} V'(B + \tilde{V} - zI_{rp})V\begin{pmatrix} \Sigma_s & 0 \\ 0 & I_{rp-s} \end{pmatrix}\right) = det(\tilde{\Sigma}^{-1}V'(B + \tilde{V})V\tilde{\Sigma} - zI_{rp}) \\
&= det\left(\begin{pmatrix} U_{2s}' U_{1s} & \Sigma_s^{-1} V_s' \bar{V} \\ 0 & V_x' \bar{V} \end{pmatrix} - zI_{rp}\right) = det(U_{2s}' U_{1s} - zI_s)det(V_x' \bar{V} - zI_{rp-s})
\end{aligned}$$

with $\bar{V} := \begin{pmatrix} 0 \\ (V_x)_{[,-r]} \end{pmatrix}$. As the matrices $V'_x\bar{V} = V'_x\begin{pmatrix} 0 \\ (V_x)_{[,-r]} \end{pmatrix} = (V'_x, 0)\begin{pmatrix} 0 \\ V_x \end{pmatrix}$ and $U'_{2s}U_{1s}$ (compare with the proof of Lemma 3, strict Cauchy–Schwarz inequality) have all eigenvalues of magnitude less than 1 the result follows.

To obtain the background equalities, observe that (using a comparison of Z and Y to justify the third equality)

$$\begin{aligned} B &= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_s \\ (V_s)_{[,-r]}V'_s \end{pmatrix} \\ &= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_s \\ (V_s)_{[,-r]}\Sigma_s U'_{1s}U_{1s}\Sigma_s^{-1}V'_s \end{pmatrix} \\ &= \begin{pmatrix} (V_s)_{[,-r(p-1)]}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_s \\ (V_s)_{r(p-1)}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_s \end{pmatrix} \\ &= V_s\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_s \end{aligned}$$

where $(V_s)_{r(p-1)}$ is the matrix consisting of the last $r(p-1)$ rows of V_s. Therefore

$$\begin{aligned} \tilde{\Sigma}^{-1}V'BV\tilde{\Sigma} &= \tilde{\Sigma}^{-1}V'V_s\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}V'_sV\tilde{\Sigma} \\ &= \tilde{\Sigma}^{-1}\begin{pmatrix} I_s \\ 0 \end{pmatrix}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}(I_s, 0)\tilde{\Sigma} \\ &= \begin{pmatrix} \Sigma_s^{-1}\Sigma_s U'_{2s}U_{1s}\Sigma_s^{-1}\Sigma_s & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} U'_{2s}U_{1s} & 0 \\ 0 & 0 \end{pmatrix} \end{aligned} \tag{59}$$

and

$$\begin{aligned} &\tilde{\Sigma}^{-1}V'\tilde{V}V\tilde{\Sigma} \\ &= \begin{pmatrix} \Sigma_s^{-1} & 0 \\ 0 & I_{rp-s} \end{pmatrix} V'\tilde{V}V \begin{pmatrix} \Sigma_s & 0 \\ 0 & I_{rp-s} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_s^{-1} & 0 \\ 0 & I_{rp-s} \end{pmatrix} V' \begin{pmatrix} 0 \\ (V_x)_{[,-r]}V'_x \end{pmatrix} V \begin{pmatrix} \Sigma_s & 0 \\ 0 & I_{rp-s} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_s^{-1} & 0 \\ 0 & I_{rp-s} \end{pmatrix} V' \underbrace{\begin{pmatrix} 0 \\ (V_x)_{[,-r]} \end{pmatrix}}_{\bar{V}} V'_xV \begin{pmatrix} \Sigma_s & 0 \\ 0 & I_{rp-s} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_s^{-1}V'_s\bar{V} \\ V'_x\bar{V} \end{pmatrix}(0, I_{rp-s}) \\ &= \begin{pmatrix} 0 & \Sigma_s^{-1}V'_s\bar{V} \\ 0 & V'_x\bar{V} \end{pmatrix} \end{aligned} \tag{60}$$

5. Changing from Non-coprime Fractional Descriptions to Coprime Fractional Descriptions

In this section, we postulate that we have solved the Yule–Walker equations to determine a polynomial $e(z) = I - e_1z - e_2z^2 - \cdots - e_pz^p$, $e_p \neq 0$ and we have a matrix f for which $ff' = \Sigma_\nu$. However, the pair $(e(z), f)$ may *not* be left coprime; we shall show how one can obtain a left coprime pair $(\tilde{e}(z), \tilde{f})$ generating almost the same spectrum; moreover, and this is the key point, in this if it occurs, coprime pair, $\tilde{e}(0) = I$, $\tilde{e}(z)$ has *maximum* degree p, and $\tilde{f}$ is constant. The determinantal zeros of $\tilde{e}(z)$ are always a subset of those of $e(z)$; so if the noncoprime pair defines a stable AR system, the coprime pair will also define a stable system.

5.1. A Preliminary Transformation

Let T be an orthogonal matrix such that

$$Tf = \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix} \tag{61}$$

where $\bar{f}$ is square and nonsingular. Define $\tilde{f} = Tf$. Define also

$$Te(z)T' = \bar{e}(z) = \begin{pmatrix} \bar{e}_1(z) \\ \bar{e}_2(z) \end{pmatrix} \tag{62}$$

Observe that $\bar{e}(0) = I$ and $\bar{e}(z)$ has degree p. Of course, $e(z)$ and $\bar{e}(z)$ have the same determinant. Observe further that

$$e^{-1}(z)f = T'\bar{e}^{-1}(z)\tilde{f} \tag{63}$$

We shall now work with $\bar{e}^{-1}(z)\tilde{f}$ without any significant loss of generality.

5.2. Identifying the Consequences of Lack of Coprimeness

It is trivial that the pair $(e(z), f)$ is not coprime if and only if the same is true of $(\bar{e}(z), \tilde{f})$. Because

$$(\bar{e}(z)\ \tilde{f}) = \begin{pmatrix} \bar{e}_1(z) & \bar{f} \\ \bar{e}_2(z) & 0 \end{pmatrix} \tag{64}$$

and because $\bar{f}$ is square and nonsingular, it is evident that $\bar{e}(z), \tilde{f}$ is not coprime if and only if for some $z_0 \neq 0$, the matrix $\bar{e}_2(z_0)$ has less than full (row) rank. (That $z_0 \neq 0$ follows from the fact that $\bar{e}_2(0)$ includes the identity as a submatrix). Equivalently, there exists a square non-unimodular $D(z)$ with $D(0)$ nonsingular and a matrix $\tilde{e}_2(z)$ with full rank for all z such that

$$\bar{e}_2(z) = D(z)\tilde{e}_2(z) \tag{65}$$

Without loss of generality, we can suppose that $\tilde{e}_2(z)$ is row proper. For if not, it can be multiplied on the left by a unimodular matrix to produce this property, and then the inverse of the unimodular matrix can be incorporated into $D(z)$. We now aim to look at the degree of $\tilde{e}_2(z)$ and at $\tilde{e}_2(0)$.

Lemma 4: *The above construction results in $\tilde{e}_2(z)$ having degree at most p*

Proof: Denote the i-th row of $\bar{e}_2(z)$ by $e_{2i}(z)$, and the ij element of $D(z)$ by $d_{ij}(z)$. Let the i-th row degree of $\tilde{e}_2(z)$ be k_i. Then by the predictable degree property ([17], p. 387, attributed to [16] and [26]), there holds

$$\deg\, e_{2i}(z) = \max_j(\deg\, d_{ij}(z) + k_j) \tag{66}$$

In this equation as is usual the degree of a zero element is set equal to $-\infty$. Let k_{max} be the largest of the k_j. Consider the corresponding column of $D(z)$. Since this column cannot contain all zero elements, it follows that for some i, we have

$$\deg\, e_{2i}(z) \geq k_{max} = \deg\, \tilde{e}_2(z) \tag{67}$$

from which the claim is immediate by maximizing over i.

Next, if $D(0)$ is not the identity matrix, replace $D(z)$ by $D(z)D^{-1}(0)$ and replace $\bar{e}_2(z)$ by $D(0)\tilde{e}_2(z)$. This will result in $\bar{e}_2(0) = \tilde{e}_2(0)$.

5.3. The Coprime Fractional Description

Let us observe now that

$$\begin{aligned} \bar{e}^{-1}(z)\bar{f} &= \begin{pmatrix} \bar{e}_1(z) \\ \bar{e}_2(z) \end{pmatrix}^{-1} \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \bar{e}_1(z) \\ D(z)\tilde{e}_2(z) \end{pmatrix}^{-1} \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \bar{e}_1(z) \\ \tilde{e}_2(z) \end{pmatrix}^{-1} \begin{pmatrix} \bar{f} \\ 0 \end{pmatrix} \end{aligned} \tag{68}$$

By setting

$$\tilde{e}(z) = \begin{pmatrix} \bar{e}_1(z) \\ \tilde{e}_2(z) \end{pmatrix} \tag{69}$$

we have a left coprime fraction $\tilde{e}(z), \tilde{f}$ in which $\tilde{e}(z)$ has degree bounded by p and $\tilde{e}(0) = I$, as required.

6. Conclusions

In this paper, we have focused on Generalized Dynamic Factor Models and in particular with finding models for the latent variables. These latent variables can be generated by dynamic or static factors. The (minimal) static factors exhibit all the dynamics of the latent variables. Because static factors have smaller dimension than the latent variables and in addition this dimension does not depend on the cross-sectional dimension N it is more convenient to model static factors than latent variables. The emphasis in our paper is on AR models for the static factors. This restriction is justified as the autoregressive case in our setting is generic. A key advantage of parameter estimation in AR models compared to ARMA or state space models is that the estimates can be obtained by solving linear equations, viz. the Yule–Walker equations. For the problem of interest, the AR model may be singular and the Yule–Walker equations may have no unique solution. However, even in the nonunique case the minimal norm solution of the Yule–Walker equations, both for population and sample covariances, is shown to be stable. In the paper we proposed an estimation algorithm. However we did not discuss the estimation of the integer parameters q, r and s.

Acknowledgments

This work was supported by the FWF (Austrian Science Fund under contracts P17065/N12, P17378 and P20833/N18), the ARC (Australian Research Council under Discovery Project Grant DP0664427) and National ICT Australia Ltd. is gratefully acknowledged. Ltd. is funded through the Australian Government's Backing Australia's Ability initiative, and in part through the Australian Research Council.

References

1. Akaike H. Stochastic theory of minimal realization. *IEEE Trans Autom Control* 1974; 19(6): 667–674
2. Anderson BDO, Deistler M. Generalized linear dynamic factor models a structure theory. *Decision Control* 2008
3. Anderson BDO, Deistler M. Properties of zero-free transfer function matrices. *SICE J Control* 2008
4. Anderson BDO, Deistler M. Properties of zero-free spectral matrices. Forthcoming IEEE Tr.AC, 2009

5. Chamberlain G. Funds, factors and diversification in arbitrage pricing models. *Econometrica* 1983; 51(5): 1305–1324
6. Chamberlain G, Rothschild M. Arbitrage, factor structure and meanvariance analysis on large asset markets. *Econometrica* 1983; 51(5): 1281–1304
7. Deistler M. Linear dynamic multiinput/multioutput systems. In: Pena D, Tiao GC, Tsay RS (eds), A Course in Time Series Analysis, Wiley, 2001
8. Forni M, Giannone D, Lippi M, Reichlin L. Opening the black box: structural factor models versus structural VARs. Forthcoming in *Econ Theory*
9. Forni M, Hallin M, Lippi M, Reichlin L. The generalized dynamic factor model: identification and estimation. *Rev Econ Stat* 2000; 65: 453–473
10. Forni M, Hallin M, Lippi M, Reichlin L. The generalized dynamic factor model: consistency and rates. *J Econ* 2004; 119: 231–255
11. Forni M, Hallin M, Lippi M, Reichlin L. The generalized dynamic factor model: one-sided estimation and forecasting. *J Am Stat Assoc* 2005; 100: 830–840
12. Forni M, Lippi M. The generalized dynamic factor model: representation theory. *Econ Theory* 2001; 17: 1113–1141
13. Giannone D, Reichlin L, Sala L. Monetary policy in real time. NBER Macroeconomic Annual 2004, 2004
14. Hannan EJ. *Multiple Time Series*. Wiley, 1970
15. Hannan EJ, Deistler M. *The Statistical Theory of Linear Systems*. Wiley, New York, 1988
16. Forney Jr GD. Minimal bases of rational vector spaces, with applications to multivariable linear systems. *SIAM J Control* 1975; 13: 493–520
17. Kailath T. *Linear Systems*. Prentice Hall, Englewood Cliffs, NJ, 1980
18. Mayne DQ, Astrom KJ, Clark JMC. A new algorithm for recursive estimation of parameters in controlled ARMA processes. *Automatica* 1984; 20(6): 751–760
19. Poskitt D. Identification of echelon canonical forms for vector linear processes using least squares. *Annals Stat* 1992; 20: 195–215
20. Rozanov YA. Stationary Random Processes. Holden-Day, San Francisco, 1967
21. Sargent TJ, Sims CA. Business cycle modelling without pretending to have too much a priori economic theory. Federal Reserve Bank of Minneapolis, Minneapolis, 1977
22. Scherrer W, Deistler M. A structure theory for linear dynamic errors-in-variables models. *SIAM J Control Optim* 1998; 36(6): 2148–2175
23. Stock JH, Watson MW. Forecasting using principal components from a large number of predictors. *J Am Stat Assoc* 2002; 97: 1167–1179
24. Stock JH, Watson MW. Macroeconomic forecasting using diffusion indexes. *J Bus Econ Stat* 2002; 20: 147–162
25. Stock JH, Watson MW. An empirical comparison of methods for forecasting using many predictors. Mimeo, Princeton University, 2004
26. Vekua NP. Systems of Singular Integral Equations. Noordhoff, The Netherlands, 1967
27. Zinner C. Modeling of High-Dimensional Time Series by Generalized Dynamic Factor Models. PhD thesis, Vienna University of Technology, 2008