
TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

# DISSERTATION

# Generalized Dynamic Factor Models

## Structure Theory and Estimation for Single Frequency and Mixed Frequency Data

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Em.O. Univ.Prof. Dipl.-Ing. Dr.Techn. Manfred Deistler
Inst.Nr.: E105
Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation
von

Dipl.Ing. Alexander Filler
Matr.Nr.: e0226144
Weißes-Kreuzgasse 64/12
2340 Mödling

Wien, am 3. November 2010

_____

# Deutsche Kurzfassung

Die Arbeit befasst sich mit der Theorie von Verallgemeinerten Dynamischen Faktormodellen. Diese Modelle wurden vor ungefähr 10 Jahren von zwei Gruppen gleichzeitig vorgestellt und wurden seitdem intensiv erforscht. Der Grund warum diese Modellklasse so attraktiv für mehrere Forschungsgruppen ist, ist, dass es die zurzeit allgemeinste Klasse von Faktormodellen darstellt. Der neue Aspekt dieser Modelle ist, dass das Spektrum der Fehler nicht mehr diagonal sein muss. Dies ist zulässig, da die Asymptotiken simultan, sowohl für die zeitliche als auch für die Querschnittsdimension, durchgeführt werden. Die Divergenz der Querschnittsdimension macht Sinn, da diese Modelle für hoch dimensionale Daten gedacht sind.

Ein wichtiges Kapitel in dieser Arbeit handelt über die sogenannte Strukturtheorie der latenten Variablen. Diese ist, zumindest so weit uns bewusst ist, die allgemeinste Form die latenten Variablen zu modellieren. Dies hat seinen Preis, da wir mit singulären AR Prozessen konfrontiert sind. Diese Prozesse haben eine autoregressive Darstellung, in welcher die Kovarianzmatrix der Innovation singulär ist. Theoretisch bräuchten wir lediglich singuläre AR(1) Modelle betrachten, welche die Dynamiken eines minimalen Zustandes beschreiben würden. Leider können wir nicht annehmen, dass der komplette Zustand aus den latenten Variablen rekonstruiert werden kann. Trotzdem zeigen wir, dass, solange die Dimension eines minimalen statischen Faktors größer ist als jene der Innovationen, der minimale statische Faktor generisch eine singuläre Darstellung besitzt.

Konsequenter Weise beschreiben wir die Klasse der singulären autoregressiven Prozesse ausführlich, welche in der existierenden Literatur kaum diskutiert wurde. Da die Regularität der Kovarianzmatrix der Innovationen eine zentrale Rolle in der Theorie der regulären autoregressiven Prozesse darstellt, ist die Verallgemeinerung auf den singulären Fall wesentlich schwieriger als man vielleicht vermuten möchte. Ein interessanter Fakt ist, dass nicht nur rein linear reguläre stationäre Lösungen von singulären autoregressiven System existieren, im Gegensatz zum regulären Fall.

Zusätzlich diskutieren wir die Schätzung der Koeffizientenmatrizen und der minimalen Ordnung eines singulären AR Prozesses. Wieder zerstört die Singularität der Kovarianzmatrix der Innovationen sämtliche

Resultate aus dem regulären Fall.

Zu guter letzt diskutieren wir das Problem von Daten mit unterschiedlichen Abtastzeitpunkten. Das heißt, nicht alle Variablen erscheinen mit derselben Frequenz. Möchte man zum Beispiel ein makroökonomisches Modell analysieren in welchem das BIP und Arbeitslosenzahlen vorkommen, so wird das BIP nur viermal pro Jahr berechnet wird, während die Arbeitslosenzahlen monatlich erfasst werden.

# Acknowledgment

My special thanks go to my supervisor Prof. Manfred Deistler. Without any doubt I can say, that I would not have finished this thesis if he had not been there. He was not only the first address when I needed encouragement, but he also knew how to challenge me. Furthermore, we had many exciting and humorous discussions (not only mathematical ones), and it was a pleasure for me, spending some weeks in Australia (research in Canberra) together with him.

Many thanks go to Prof. Brian Anderson, for the numerous fascinating and challenging discussions in Vienna and in Canberra. I do not have to mention that I appreciate his sense of humor, or do I? Also I want to thank my colleague Dr. Weitian Chen. I enjoyed the time spending with him in Canberra and our collaborative work.

Additionally, sincere thanks are given to Prof. Engelbert Dockner, for employing me in one of his projects such that I could continue my thesis, and for his interest in my work. Thanks also to the fin4cast group (former FSC GmbH) who guaranteed financial supply during my first months as PhD student.

Finally, I want to thank my colleagues in Vienna. Especially, Bernd Funovits for helping me to correct my thesis. And of course, my legend - wait for it - dary (office) mate Christoph Flamm, who always helped me with my nitpicking mathematical questions, and who has simply a great sense of humor.

# Abstract

The thesis deals with the theory of Generalized Dynamic Factor Models. These models have been introduced about 10 years ago by two groups simultaneously and have been discussed intensively since then. The reason why this model class attracts so many different research groups is, that it is, up to now, the most general class of factor models. The new aspect of these models is, that the spectrum of the noise does not have to be diagonal anymore. This is possible as asymptotics, in both the time and the cross-section, are regarded simultaneously. The divergence of the cross-section dimension is reasonable as these models are designed for high dimensional data sets.

An important part of this thesis is the so called structure theory for the latent variables. It is, at least to our knowledge, the most general way to model these variables. This has a price, as we are faced with singular AR processes. These processes have an autoregressive representation where the driving white noise has a singular covariance matrix. Theoretically we only need singular AR(1) models to describe the dynamics of a minimal state. Unfortunately we cannot assume that the whole state can be reconstructed by the latent variables. Nevertheless we have shown, that if the dimension of a minimal static factor is larger than the dimension of the driving white noise, the minimal static factor has generically a singular autoregressive representation.

Consequently we present a detailed description of singular autoregressive processes, which have been poorly discussed in the existing literature. As the regularity of the covariance matrix of the driving white noise plays a central role in the theory of regular autoregressive processes, the generalization to the singular case is much more complicated than one might expect. An interesting fact is, that not only purely linearly regular stationary autoregressive processes exist, contrary to the regular case.

Additionally we discuss the estimation of the coefficients and the minimal order of singular AR processes. Again the singularity of the covariance matrix of the noise destroys several results from the regular case.

Finally we discuss the problem of mixed frequency sampling data, which means that not all variables

have the same sampling frequency. Example given, if one wants to analyze a model in which the GDP and the unemployment rate are used, the GDP is calculated only four times a year, whereas the unemployment rate appears monthly.

# Contents

# Symbol Index

**vector processes:**

| | | |
|---|---|---|
| $y_t$ | ... | observations ($N$ dimensional or of infinite dimension) |
| $y_t^N$ | ... | $N$ dimensional vector of observations |
| $\chi_t$ | ... | latent variables ($N$ dimensional or of infinite dimension) |
| $\chi_t^N$ | ... | $N$ dimensional vector of latent variables |
| $u_t$ | ... | idiosyncratic noise ($N$ dimensional or of infinite dimension) |
| $u_t^N$ | ... | $N$ dimensional vector of idiosyncratic noise |
| $x_t$ | ... | $n$ dimensional (minimal) state |
| $f_t$ | ... | $r$ dimensional minimal static factor |
| $\nu_t$ | ... | $r$ dimensional minimal white noise with singular covariance matrix |
| $\varepsilon_t$ | ... | $q$ dimensional (minimal) dynamic factor/ driving white noise |
| $x_{t+1\|t}$ | ... | one-step-ahead predictor of $x_{t+1}$ given information up to time $t$ |

**second moments:**

| | | |
|---|---|---|
| $\gamma_j^x$ | $=$ | $\mathbb{E}x_t x_{t-j}'$ |
| $\Gamma_p^x$ | $=$ | $\mathbb{E}[x_{t-1}', ..., x_{t-p}']'[x_{t-1}', ..., x_{t-p}']$ |
| $f_x$ | ... | spectral density of $x_t$ |
| $\Sigma_\nu$ | ... | (singular) covariance matrix of the white noise process $\nu_t$ |
| $\Sigma$ | ... | (full rank) covariance matrix of a driving white noise process ($\varepsilon_t$ usually) |

**matrices:**

| | | |
|---|---|---|
| $I$ | ... | identity matrix of suitable dimension |
| $I_n$ | ... | identity matrix of dimension $(n \times n)$ |
| $\mathscr{O}$ | ... | Observability matrix |
| $\mathscr{R}$ | ... | Reachability matrix |
| $\mathscr{H}$ | ... | Hankel matrix |

**Hilbert space notations:**

$H_y$ ... $\overline{span}\{y_{i,t}|i \in \mathbb{N}; t \in \mathbb{Z}\}$ or $\overline{span}\{y_{i,t}|i = 1, ..., N; t \in \mathbb{Z}\}$
$H_y(t)$ ... $\overline{span}\{y_{i,t}|i \in \mathbb{N}; s \leq t\}$ or $\overline{span}\{y_{i,t}|i = 1, ..., N; s \leq t\}$

**miscellaneous:**

$z$ ... complex variable/ backward shift
$\mathscr{L}$ ... Lesbegue measure

# Chapter 1

# Introduction

*"We are facing a data tsunami."*

*Bart de Moor, Panel discussion: "Econometrics and Systems Theory - Quo Vadis?", Conference in honor of Manfred Deistler, Vienna 2009.*

In every empiric analysis collecting data is mandatory. Recently, especially high dimensional time series are becoming more and more important in many disciplines, for example in financial time series analysis, macro economic modeling, biomedicine and engineering. What the collected samples have in common is that they are huge. Huge not only in the time dimension but in the cross-sectional dimension too. That means that many sensors measure something simultaneously.

As long as the cross-sectional dimension is not too big, classic models are still reasonable such as unrestricted AR or ARMA models. But as the parameter space of such models is proportional to $N^2$, if $N$ is the cross-section dimension, it is obvious that the estimation of such models is problematic. The problem is often called the "curse of dimensionality". To be more precise, the curse of dimensionality is not such a big problem if the time dimension of the sample, $T$ say, is far bigger than $N$, but in practice $N$ and $T$ are usually approximately of the same size.

Of course one could always estimate only univariate models or group some variables and estimate a smaller ARMA model for each group, but this means that information is lost, as not all variables would be used simultaneously. Furthermore a huge cross-sectional dimension has usually an explicitly reason namely to observe co-movement, i.e. something like a common trend or even an oscillation.

Therefore models are needed to first handle these huge data sets and second to extract the co-movement. Factor models are reasonable in this sense, as their parameter spaces grow only linearly with $N$ and extract the co-movement in (a few) factors.

This thesis is concerned with Generalized Dynamic Factor Models (GDFMs). The theory of GDFMs was developed about 10 years ago by two groups of researchers simultaneously: On the one hand Forni, Hallin, Lippi and Reichlin ([Forni et al., 2000], [Forni and Lippi, 2001], [Forni et al., 2005]) and on the other hand Stock and Watson ([Stock and Watson, 2002a], [Stock and Watson, 2002b]). Although Factor Models have been introduced already 100 years ago, they have become extremely popular in recent years for modeling macro economic and financial data, based on the results of the two groups mentioned above.

Apart from that, GDFMs have a huge potential of becoming very attractive for biomedicine and engineering as the problem called "oversensoring" is becoming more and more important these days.

A factor model is characterized by the representation of the observed process $y_t$ as a sum of two uncorrelated processes

$$y_t = \chi_t + u_t$$

where $\chi_t$ is called the process of the latent variables and $u_t$ is a so called idiosyncratic noise process. Both the latent variables and the idiosyncratic noise are not directly observable. The idea is the following: Every "co-movement" of the observed variables $y_t$ in the cross-section is collected in the latent variables $\chi_t$ and the remainder is an idiosyncratic (i.e. only relevant for each component) noise part, i.e. $\chi_t$ is a hidden unobserved process which is driven by a few factors and $u_t$ is the vector collecting the specific error in each component (which does not have to be white noise). Another explanation is that a "true" process $\chi_t$ can only be measured with noise. This is the theory of errors-in-variables models where the process $\chi_t$ has a non-trivial left kernel and the noise $u_t$ occurs as $y_t$ is a noisy measurement of $\chi_t$. In Control theory one would call $u_t$ the measurement noise rather than idiosyncratic noise, which is used by econometricians. However the observations $y_t$ have a common process $\chi_t$, which is driven by a (far) lower dimensional process $f_t$, and an error process $u_t$ which is uncorrelated with $\chi_t$ or $f_t$ respectively.

The first factor models were designed to extract an Intelligence Factor from a large number of measurements which were independently and identically distributed. These models assumed that the latent variables had one common static factor and the idiosyncratic noise had a diagonal covariance matrix.

The idea remained the same over the years, namely to compress information, in the so called cross-sectional dimension which is usually (very) large and additionally in the time dimension, into low dimensional factors. The continuative literature discussed models which allowed dynamics within the latent variables and the idiosyncratic noise but still with the restriction of a diagonal spectrum of the noise part (see for instance [Geweke, 1977], [Sargent and Sims, 1977], [Scherrer and Deistler, 1998]).

Chamberlain and Rothschild were the first who introduced models which weakened the assumption of strictly idiosyncratic noise (i.e. diagonal spectrum) [Chamberlain, 1983],
[Chamberlain and Rothschild, 1983]. They defined a static factor model with the additional assumption that an infinite number of variables is available at each point in time. Therefore they could allow the idiosyncratic noise to have (weak) correlations in the cross-section.

Generalized Dynamic Factor Models combine both innovations, namely the dynamic representation of the latent variables and the idiosyncratic noise as well as the weakening of the assumption of a diagonal spectrum of the errors, of course again under the assumption that the cross-sectional dimension tends to infinity.

The popularity of Generalized Dynamic Factor Models is caused by the use of these models by several central banks, for example given the ECB and the FED. As macro economic and financial data always have a large cross-section the asymptotic of letting N, the cross-sectional dimension, going to infinity is reasonable. Furthermore these models are potential candidates for other areas as for example given the problem of oversensoring might be handled with GDFMs.

## 1.1 General Framework

Throughout the whole thesis we will deal with (wide sense) stationary processes only. Of course, in many applications the true observations are obviously not stationary, such that de-trending and differencing procedures have to be applied first. We assume that these procedures result in a stationary process $(y_t)_{t\in\mathbb{Z}}$. For the sake of simplicity we will write only $y_t$ instead of $(y_t)_{t\in\mathbb{Z}}$.

As already mentioned before, a factor model is characterized by a decomposition of the observed process $y_t$ into two unobserved processes $\chi_t$ and $u_t$

$$y_t = \chi_t + u_t \tag{1.1}$$

The processes $y_t, \chi_t$ and $u_t$ fulfill the following assumption.

**Assumption 1.**
*$x_t$ is a (wide sense) stationary zero mean linearly regular process with absolutely summable covariances, such that the spectral density exists and it is the Fourier transformation of its covariance function*

$$f_x(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{is\lambda} \gamma^x(s) \tag{1.2}$$

*where $f_x(\lambda)$ denotes the spectral density and $\gamma^x(s)$ the covariance function of $x_t$.*

Assumption 1 implies for example that $\chi_t$ has a Wold representation

$$\chi_t = w(z)\varepsilon_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j}, \sum_{j=0}^{\infty} \|w_j\| < \infty$$

where $z$ denotes, throughout the whole thesis, a complex variable as well as the backward shift. Moreover the processes $y_t$, $\chi_t$ and $u_t$ fulfill

**Assumption 2.**

1. $y_{i,t} = \chi_{i,t} + u_{i,t}, i = 1, ..., N$, *where* $N$ *is the dimension of* $y_t$.
2. $\mathbb{E}\chi_t u_s' = 0, \forall s, t \in \mathbb{Z}$, *where* $(\cdot)'$ *denotes the transpose of* $(\cdot)$.
3. *The spectral density* $f_\chi(\lambda)$ *is rational and has rank* $q$ $\forall \lambda \in [-\pi, \pi]$.

Obviously

$$\gamma_s^y = \mathbb{E}y_t y_{t-s}' = \mathbb{E}\chi_t \chi_{t-s}' + \mathbb{E}u_t u_{t-s}' = \gamma_s^\chi + \gamma_s^u \tag{1.3}$$

holds and therefore the analogous statement holds for the spectral densities

$$f_y(\lambda) = f_\chi(\lambda) + f_u(\lambda) \tag{1.4}$$

In the following we give a very short overview about other models related to Generalized Dynamic Factor Models.

## 1.2 Principal Component Models

A very popular class of models for multivariate time series are Principal Component Models (PC models). The idea is to compress as much information of the $N$ dimensional process $y_t$ as possible into a $q$ dimensional process, $x_t$ say, which is a linear transformation of $y_t$, i.e. $x_t = B(z)y_t$, where $B(z)$ is a linear filter. With compressing as much information as possible it is meant, that the covariance matrix of the remainder $u_t = y_t - C(z)B(z)y_t = y_t - C(z)x_t = y_t - \chi_t$ is minimized, where $C(z)$ is also a linear filter.

### Static Principal Component Models

A static PCA model is characterized by $B(z) = B...constant$, $C(z) = C...constant$. That means, that the process $x_t$ is a static transformation of the process $y_t$. Thus, for given $q$, the goal is to minimize

$$\mathbb{E}[y_t - CBy_t]'[y_t - CBy_t] \tag{1.5}$$

where $C, B' \in \mathbb{R}^{N\times q}$. It is a well known result that the minimum is achieved for $C = B' = O_1$, where $\mathbb{E}y_t y_t' = O\Lambda O' = [O_1, O_2]\begin{bmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{bmatrix}\begin{bmatrix}O_1'\\ O_2'\end{bmatrix}$ is the eigenvalue decomposition of $\mathbb{E}y_t y_t'$ and $O_1$ is the matrix consisting of the normalized eigenvectors corresponding to the $q$ largest eigenvalues of $\mathbb{E}y_t y_t'$ (see e.g. [Darroch, 1965]). The components $O[,i]'y_t, i = 1, ..., N$ are called the static Principal Components of $y_t$ and $O[,i]$ denotes the $i$-th column of $O$.

### Dynamic Principal Component Models

Dynamic PC models are a generalization of static PC models, in the sense that now $B(z) = \sum_{j=-\infty}^{\infty} B_j z^j$ and $C(z) = \sum_{j=-\infty}^{\infty} C_j z^j$ are linear filters, and (1.5) has to be minimized. Thus, for given $q$, the task is to minimize

$$\int_{-\pi}^{\pi} [I - C(e^{-i\lambda})B(e^{-i\lambda})]^* f_y [I - C(e^{-i\lambda})B(e^{-i\lambda})] d\lambda$$

where $^*$ denotes conjugate and transpose. The solution is similar to the static case, namely $C(e^{-i\lambda}) = B(e^{-i\lambda})^* = O_1(e^{-i\lambda})$ where $O_1(e^{-i\lambda})$ denotes the matrix consisting of the normalized eigenvectors corresponding to the $q$ largest eigenvalues of the spectral density matrix $f_y$. The components $O[,i](e^{-i\lambda})^* y_t, i = 1, ..., N$ are called the dynamic Principal Components of $y_t$ and $O[,i](e^{-i\lambda})$ denotes the $i$-th column of $O(e^{-i\lambda})$.

From a geometric point of view we can say, that in (static/ dynamic) PC models the observed process $y_t$ is projected on the space spanned by its (static/ dynamic) principal components corresponding to the largest eigenvalues of the covariance/ spectral density matrix of $y_t$, for more details the reader is referred to for instance [Jolliffe, 2002]. As this projection is an orthogonal one, we result in a pair $\chi_t = O_1(e^{-i\lambda})O_1(e^{-i\lambda})^* y_t, u_t = [I - O_1(e^{-i\lambda})O_1(e^{-i\lambda})^*]y_t = O_2(e^{-i\lambda})O_2(e^{-i\lambda})^* y_t$ which is orthogonal. Furthermore, as the orthogonal projection is unique, the process $\chi_t$ and thus $u_t$ are uniquely determined.

## 1.3 Factor Models with strictly idiosyncratic noise

Factor Models with strictly idiosyncratic noise commence from a diagonal spectrum of the noise. This implies that the process $\chi_t$ which contains most of the information (variance) of the observed process $y_t$ is **not** a linear transformation of $y_t$ contrary to PC models. Thus, things are much more complicate, e.g. to identify the processes $\chi_t$ and $u_t$.

The observed process $y_t$ satiesfies Assumption 1 and has a representation

$$y_t = \chi_t + u_t$$

such that $\chi_t$ and $u_t$ fulfill Assumptions 1 and 2. Moreover the spectral density of $u_t$ is diagonal, i.e. the noise $u_t$ is strictly idiosyncratic.

### Static Factor Models with strictly idiosyncratic noise

This model class is the simplest and best known class of factor models. One assumes additionally that $y_t$ is iid, such that the covariances are the spectral densities times a factor $2\pi$, which implies that $\gamma^u = \mathbb{E}u_t u_t'$ is diagonal. As the spectral density and thus the covariance of $\chi_t$ has rank $q$ we have

$$\gamma^y = \gamma^\chi + \gamma^u = W\gamma^\varepsilon W' + \gamma^u \tag{1.6}$$

where $W \in \mathbb{R}^{N\times q}$. Thus for identifying the model we have to identify $\gamma^\chi$ and $\gamma^u$ first and have to select $W$ and $\gamma^\varepsilon$ afterwards.

It has been shown in [Scherrer and Deistler, 1998] that $\gamma^\chi$ and $\gamma^u$ can be identified generically if $q \leq q_{max}$ holds, where $q_{max}$ is the so called Lederman bound. The second selection can be easily done by setting $\gamma^\varepsilon = I_q$ such that $W$ is identified up to post-multiplication by orthogonal matrices.

### Dynamic Factor Models with strictly idiosyncratic noise

This model class is the extension to the dynamic case of the models described above. The task is to identify the spectral densities $f_\chi$ and $f_u$ where $f_u$ is diagonal and $f_\chi$ has rank $q$. Similar to the static case a bound for $q$ exists such that $f_\chi$ and $f_u$ are generically identifiable (see [Scherrer and Deistler, 1998]).

# Chapter 2

# Generalized Dynamic Factor Models

In this section we want to introduce and discuss the so called Generalized Dynamic Factor Models (GDFMs). In the previous section we have already mentioned static factor models and dynamic factor models with strictly idiosyncratic noise. The essential feature of factor models with strictly idiosyncratic noise is a diagonal spectrum of the noise. Unfortunately this feature is unrealistic in many applications [Scherrer and Deistler, 1998]. The first models which weaken this assumption have been introduced by Chamberlain and Rothschild [Chamberlain, 1983], [Chamberlain and Rothschild, 1983]. These models are in a static framework and use the additional asymptotic of $N$ going to infinity, to avoid the need of a diagonal error covariance. GDFMs extend these models to the dynamic case and have been introduced by two groups of researchers namely by Forni, Hallin, Lippi and Reichlin (see [Forni et al., 2000], [Forni and Lippi, 2001], [Forni et al., 2005]) and by Stock and Watson (see [Stock and Watson, 2002a], [Stock and Watson, 2002b]).

As already mentioned, the weakening of the assumption on the spectrum of the noise (diagonal spectrum) is of big theoretical value as an uncorrelated noise process is highly unrealistic in almost all applications. This together with the fact that a big cross-sectional dimension is not a drawback anymore, actually a big cross-sectional dimension is theoretically even needed, makes these models so popular.

In the following we will first give a definition of a Generalized Dynamic Factor Model and second present a characterization of GDFMs by using the eigenvalues of the spectral density of the observations.

Before we can define a GDFM we need some basics and preliminary definitions as the asymptotics of GDFMs are with respect to the time dimension *and* the cross-sectional dimension.

The following sections we will repeat the major lemmas and theorems of the paper [Forni and Lippi, 2001] which yields the basement of the theory of GDFMs.

## 2.1 Basics

Let us recall that the linear space $L_2(\mathscr{P})$ of all zero mean complex valued square integrable random variables, defined on a set $\Omega$ and $\mathscr{P} = (\Omega, F, P)$ is a probability space, is a Hilbert space with the inner product $\langle x, y\rangle = \mathbb{E}x\bar{y}$, where $^-$means conjugation. Strictly speaking we have to factorize the space by the equivalence class of functions which are 0 almost everywhere first.

We will deal with doubly indexed sequences

$$(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}} \tag{2.1}$$

For the sake of simplicity we will denote with $y_t$ the doubly indexed sequence $(y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$. Let $y_t^N$ be an $N$ dimensional vector process $y_t^N = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix}$, with components $y_{i,t} \in L_2(\mathscr{P})$, then we can write (2.1) also as

$$((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}} \tag{2.2}$$

Furthermore we denote by $H_{y^N}$ the Hilbert space spanned by all components of $y_t^N$, i.e. $H_{y^N} = \overline{span}\{y_{i,t}|1 \le i \le N, t \in \mathbb{Z}\}$, where $\overline{span}\{\cdot\}$ denotes the closure of the linear hull of $\{\cdot\}$ and by $H_y$ the Hilbert space $\overline{span}\{y_{i,t}|i \in \mathbb{N}, t \in \mathbb{Z}\}$. Of course $H_{y^N} \subseteq H_y$ holds.

Let us denote by $f_y$ the infinite dimensional matrix whose upper left $N \times N$ submatrix is the spectral density $f_y^N$ of $y_t^N$. We will denote by $\lambda_i^{y^N}(\theta)$ the $i$-th largest (dynamic) eigenvalue of the ($N$ dimensional) spectral density matrix $f_{y^N}(\theta)$ and $o_i^N(\theta)$ will be the corresponding eigenvector, i.e.

$$f_{y^N}(\theta) = [o_1^N(\theta), ..., o_N^N(\theta)]diag[\lambda_1^{y^N}(\theta), ..., \lambda_N^{y^N}(\theta)][o_1^N(\theta), ..., o_N^N(\theta)]^*$$

**Lemma 2.1.1.** *For $i$ fixed $\lambda_i^{y^N}(\theta)$ is a non-decreasing sequence for growing $N$ for all $\theta \in [-\pi, \pi]$, i.e. $\lambda_i^{y^N}(\theta) \le \lambda_i^{y^{N+1}}(\theta)$*

The Lemma can be proofed by Corollary 1 on page 293 in [Lancaster and Tismenetsky, 1985]. As a consequence of this Lemma we can define $\lambda_i^y(\theta) = \lim_{N\to\infty} \lambda_i^{y^N}(\theta) = \sup_N \lambda_i^{y^N}(\theta)$. Note that $\lambda_i^y(\theta) = \infty$ is not excluded.

Let $\mathbf{a}_N = [a_1, ..., a_N]$ be an $N$ dimensional row vector, then the complex linear space $L_2^N(\Theta, f_{y^N})$ of all $N$ dimensional vectors $\mathbf{a}_N$, where the components $a_i(\theta)$ are measurable complex functions defined on $\Theta = [-\pi, \pi]$ and $\int_{-\pi}^{\pi} \mathbf{a}_N(\theta) f_{y^N}(\theta) \mathbf{a}_N^*(\theta) d\theta < \infty$ is a Hilbert space with the inner product $\langle \mathbf{a}_N, \mathbf{b}_N\rangle_{f_{y^N}} = \int_{-\pi}^{\pi} \mathbf{a}_N(\theta) f_{y^N}(\theta) \mathbf{b}_N^*(\theta) d\theta$.

$L_2^N(\Theta, f_{y^N})$ is also called the frequency domain of $y_t^N$ and $H_{y^N}$ the time domain. For every $\mathbf{a}_N \in L_2^N(\Theta, f_{y^N})$ $a_N(z)y_t^N$ is well defined and has spectral density $\mathbf{a}_N f_{y^N} \mathbf{a}_N^*$.

Let $\mathscr{L}$ be the Lebesgue measure on $\mathbb{R}$. Let us recall that a real function $f : \Theta \to \mathbb{R}$ is essentially bounded if there exists a real $M$ and a subset $D$ of $\Theta$ such that $\mathscr{L}(D) = 0$ and $|f(\theta)| < M, \forall \theta \in \Theta \backslash D$. Moreover $f$ is essentially bounded if and only if ess sup $(f) = \inf\{M | \mathscr{L}(\{y | f(y) > M\}) = 0\} < \infty$. Now we are ready to define a GDFM.

## 2.2 The Definition of a GDFM

We start with a sequence of vectors of observations

$$((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}} \tag{2.3}$$

which is nested, i.e. the first $N-1$ components of $y_t^N$ form $y_t^{N-1}$. For all $N \in \mathbb{N}$, $y_t^N$ has a representation as

$$y_t^N = \chi_t^N + u_t^N = w^N(z)\varepsilon_t + u_t^N \tag{2.4}$$

and $y_t^N, \chi_t^N, u_t^N$ fulfill Assumption 1 and Assumption 2 respectively. The process $\varepsilon_t$ is called a dynamic factor process. Note that $\varepsilon_t$ is not indexed by $N$ which means that the finite, $q$ say, dimensional process $\varepsilon_t$ drives the underlying latent process $\chi_t^N$. As the sequence of $\chi_t^N$ is nested, the transfer functions $w^N(z)$ and the sequence of the corresponding spectral densities are nested too. Another way of thinking about the nestedness and non-dependence of $\varepsilon_t$ on $N$ is that an additional component of $y_t$ gives more information on the process $\varepsilon_t$. Later we will see that $\varepsilon_t$ is actually a limit of linear combinations of $y$s.

**Definition 2.2.1. (dynamic averaging sequence).** *Let $\mathbf{a}_N \in L_2^N(\Theta, I_N) \cap L_2^N(\Theta, f_{y^N})$ for $N \in \mathbb{N}$, where $\Theta = [-\pi, \pi]$ and $I_N$ is an $N$ dimensional Identity matrix. Then $(\mathbf{a}_N)_{N\in\mathbb{N}}$ is a **dynamic averaging sequence** (DAS), if $\lim_N \|\mathbf{a}_N\|_{I_N} = 0$, where $\|\cdot\|_{I_N}$ denotes the norm induced by the inner product on $L_2^N(\Theta, I_N)$.*

A very simple example for a DAS is the following.

**Example 2.2.2.** *Let $\mathbf{a}_N = \frac{1}{N}[\underbrace{1, 1, ..., 1}_{N}]$ then $\lim_{N\to\infty} \|\mathbf{a}_N\|_{I_N} = 0$ obviously holds. This DAS produces simply the arithmetic averages of the first $N$ components for $N \to \infty$.*

**Definition 2.2.3. (weakly dependent).** *We say that $u_t = ((u_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$, where $u_t^N$ fulfills Assumption 1 for all $N \in \mathbb{N}$, is **weakly dependent** if $\lim_N a_N(z)u_t^N = 0$ in mean square sense for any DAS $(\mathbf{a}_N)_{N\in\mathbb{N}}$, i.e. $\lim_N Var(a_N(z)u_t^N) = \lim_N \|a_N(z)u_t^N\|_{I_N} = \lim_N \int_{-\pi}^{\pi} \mathbf{a}_N f_{u^N} \mathbf{a}_N^* = \lim_N \|\mathbf{a}_N\|_{f_{u^N}} = 0$.*

Note that strictly speaking we should use the wording weakly correlated as this definition only requires properties of the spectral density of $u_t$ and therefore only the second moments are used. Another wording for weakly dependent would be idiosyncratic as Forni and Lippi [Forni and Lippi, 2001] do and this would also perfectly fit as a definition of an idiosyncratic noise, but we want to use the word idiosyncratic independently from letting the cross-section go to infinity.

**Theorem 2.2.4.** *$u_t$ is weakly dependent if and only if $\lambda_1^u$ is essentially bounded.*

*Proof. (Theorem 2.2.4.)* See also [Forni and Lippi, 2001].
"if": For any DAS $(\mathbf{a}_N)_{N\in\mathbb{N}}$ we have:

$$\lim_N Var(a_N(z)u_t^N) = \lim_N \|\mathbf{a}_N(z)u_t^N\|_{I_N} = \lim_N \int_{-\pi}^{\pi} \mathbf{a}_N f_{u^N} \mathbf{a}_N^*$$
$$\leq \lim_N \int_{-\pi}^{\pi} \lambda_1^{u^N} \mathbf{a}_N \mathbf{a}_N^* \leq \lim_N \int_{-\pi}^{\pi} \lambda_1^u \mathbf{a}_N \mathbf{a}_N^* = \lim_N \text{ess sup}(\lambda_1^u) \int_{-\pi}^{\pi} \mathbf{a}_N \mathbf{a}_N^* = 0 \quad (2.5)$$

"only if": If $\lambda_1^u$ is not essentially bounded, then it is not hard to see that there exists a sequence $(N_s)_{s\in\mathbb{N}}$ and a corresponding sequence $(\mathbf{a}_{N_s})_{s\in\mathbb{N}}$, with $\mathbf{a}_{N_s} \in L_2^{N_s}(\Theta, I_{N_s}) \cap L_2^{N_s}(\Theta, f_{u^{N_s}})$ with $\|\mathbf{a}_{N_s}\|_{I_{N_s}} = 1$ and $\|a_{N_s}(z)u_t^N\|_{I_{N_s}}^2 \geq M_s$ with $M_s \to \infty$. Thus $\mathbf{b}_{N_s} = \mathbf{a}_{N_s}/\|a_{N_s}(z)u_t^N\|_{I_{N_s}}$ is a DAS, but $\|b_{N_s}(z)u_t^N\|_{I_{N_s}} = 1$ and therefore $u_t$ is not weakly dependent. □

Now we can define a Generalized Dynamic Factor Model.

**Definition 2.2.5. (Generalized Dynamic Factor Model).** *We say that $y_t = ((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ has a representation as a **Generalized Dynamic Factor Model**, if there exists a $q$ dimensional white noise process $\varepsilon_t = [\varepsilon_{1,t}, ..., \varepsilon_{q,t}]'$ and a doubly indexed sequence $((u_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$, such that*

1. $$y_{i,t} = \chi_{i,t} + u_{i,t} = w_{i,1}\varepsilon_{1,t} + .. + w_{i,q}\varepsilon_{q,t} + u_{i,t} \quad (2.6)$$

   *holds $\forall i \in \mathbb{N}$, and*

   $$y_t^N = \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix} = \chi_t^N + u_t^N = \begin{bmatrix} \chi_{1,t} \\ \vdots \\ \chi_{N,t} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix} \quad (2.7)$$

   *fulfills Assumptions 1 and 2 for all $N \in \mathbb{N}$*

2. *$u_t$ is weakly dependent*

3. *$\lambda_q^\chi(\theta) = \infty$, a.e. in $[-\pi, \pi]$*

Let us give a very simple example of a Generalized Dynamic Factor Model.

**Example 2.2.6.** *Let*

$$y_t^N = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \varepsilon_t + \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix}, \forall N \in \mathbb{N}$$

*with* $\mathbb{E}\varepsilon_t = \mathbb{E}u_{i,t} = 0, \forall i \in \mathbb{N}$, $Var(\varepsilon_t) = Var(u_{i,t}) = 1 \forall i \in \mathbb{N}$, *then*

$$f_{y^N} = \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{bmatrix}}_{f_{\chi^N}} + \underbrace{diag[1, ..., 1]}_{f_{u^N}}$$

*where* $\lambda_1^{\chi^N} = N$, $\lambda_1^{u^N} = 1, \forall N \in \mathbb{N}$.

## 2.3 Characterization of GDFMs

In the previous section we figured out which properties a doubly indexed sequence $y_t = ((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ has to satisfy to have a Generalized Dynamic Factor Model representation, or sloppy speaking to be a GDFM. Now we want to give necessary and sufficient conditions, on some eigenvalues of the spectral density $f_y$, such that $y_t$ is a Generalized Dynamic Factor Model. These conditions have been established by Forni and Lippi in [Forni and Lippi, 2001], and will repeat the major theorems and proofs of this paper here.

**Theorem 2.3.1.** *The doubly indexed sequence* $y_t = ((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ *has a representation as a Generalized Dynamic Factor Model if and only if:*

*(I)* $\lambda_q^y(\theta) = \infty$ *a.e. in* $[-\pi, \pi]$

*(II)* $\lambda_{q+1}^y(\theta)$ *is essentially bounded*

The proof of the Theorem is not really easy and rather technical. The following definition is of central importance and explains a lot.

**Definition 2.3.2. (aggregate).** *We call* $v_t \in H_y$ *an* ***aggregate*** *if there exists a DAS* $(\mathbf{a}_N)_{N\in\mathbb{N}}$ *such that* $\lim_N a_N(z)y_t^N = v_t$ *holds. The space spanned by all aggregates will be denoted by* $\mathscr{A}(y)$ *and we will call it the* ***aggregation space*** *of* $H_y$.

Note that $\mathscr{A}(y)$ is a closed subspace of $H_y$ (for a proof see Lemma 6 in [Forni and Lippi, 2001]). The aggregation space is the set of random variables we are really interested in as it collects the whole information of the latent variables. As we will see below the observed process can be split in the projection on the aggregation space and a remainder which is weakly dependent. The outline of the proof

of Theorem 2.3.1 is the following. First it is shown that if $y_t$ has a Generalized Dynamic Factor Model representation then (I) and (II) hold, which is easy. Then the other direction is shown within some steps. The first will be to show that $\mathscr{A}(y)$ contains a $q$ dimensional white noise process, $(z_t)_{t\in\mathbb{Z}}$ say. The second that $\mathscr{A}(y) = H_z$, and the third one will complete the proof by showing that the remainder $\tilde{u}_t$ of the element wise projection

$$y_{i,t} = y_{i,t}|_{H_z} + \tilde{u}_{i,t}, i \in \mathbb{N} \tag{2.8}$$

is weakly dependent.

**Assumption 3.** $\lambda_i^{y^N} \geq 1$ *holds* $\forall N, i \leq N$

Note that Assumption 3 is no restriction at all as it is seen in the following. Let $\eta = (\eta_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ with $Var(\eta_{i,t}) = 1, \forall i \in \mathbb{N}, t \in \mathbb{Z}$, $\mathbb{E}\eta_{i,t}y_{j,s} = 0, \forall i, j \in \mathbb{N}, s, t \in \mathbb{Z}$ and $\mathbb{E}\eta_{i,t}\eta_{j,s} = 0, \forall i, j \in \mathbb{N}, s, t \in \mathbb{Z}$ and define $\tilde{y}_{i,t} = y_{i,t} + \eta_{i,t}$. Let us assume that Theorem 2.3.1 holds for $\tilde{y}_t$. It is obvious that (I) and (II) hold for $\tilde{y}_t$ if and only if they hold for $y_t$ as $\lambda_i^{\tilde{y}^N} = \lambda_i^{y^N} + 1$. Furthermore it is easy to see that if $y_t$ has a Generalized Dynamic Factor Model representation this implies that $\tilde{y}_t$ has a Generalized Dynamic Factor Model representation too as $\eta$ is weakly dependent and orthogonal on $y_t$. On the other hand if $\tilde{y}_t$ has a Generalized Dynamic Factor Model representation it has the decomposition $\tilde{y}_t = \tilde{\chi}_t + \tilde{u}_t = \tilde{y}_t|_{\mathscr{A}(y)} + \tilde{u}_t$. By the definition of $\tilde{y}_t$, $\tilde{y}_t|_{\mathscr{A}(y)} = y_t|_{\mathscr{A}(y)}$ and therefore $\tilde{y}_t = \chi_t + \tilde{u}_t = \chi_t + u_t + \eta_t$ holds. As $\tilde{u}_t$ is weakly dependent $u_t$ is weakly dependent as well and consequently $y_t$ has a Generalized Dynamic Factor Model representation.

As the divergence of the eigenvalues in (I) is not uniformly in $[-\pi, \pi]$ the following definition will be helpful for the proof of Theorem 2.3.1.

**Definition 2.3.3.** ($\mathbf{K_D}$). *Let $D \subseteq [-\pi, \pi]$ then we denote by $\mathbf{K_D}$ the set of $q \times q$ matrices $C(\theta)$ which fulfill*

- *the elements $C(\theta)_{i,j}, 1 \leq i, j \leq q$ are essentially bounded functions on $[-\pi, \pi]$*
- *$C(\theta)C^*(\theta) = I_q$ for $\theta \in D$*
- *$C(\theta) = 0_q$ for $\theta \notin D$*

*i.e. $C(\theta)$ is unitary if $\theta \in D$ and zero else.*

*Proof.* ***(Theorem 2.3.1.)*** See also [Forni and Lippi, 2001].
**"only if":** By definition we have $f_{y^N} = f_{\chi^N} + f_{u^N}$ which implies that $f_{y^N} \geq f_{\chi^N}$ holds and therefore $\lambda_i^{y^N} \geq \lambda_i^{\chi^N}, i = 1, ..., N$ follows which establishes (I). Furthermore it is not hard to verify that $\lambda_i^{y^N} \leq \lambda_i^{\chi^N} + \lambda_1^{u^N}, i = 1, ..., N$ holds (see Theorem 1 page 301 in [Lancaster and Tismenetsky, 1985]) and therefore

$$\lambda_{q+1}^{y^N} \leq \lambda_{q+1}^{\chi^N} + \lambda_1^{u^N} = \lambda_1^{u^N} \tag{2.9}$$

holds, which proves (II).

**"if":**

We define $\Pi$ as the subset of $[-\pi,\pi]$ such that $\mathscr{L}([-\pi,\pi]\backslash\Pi)=0$, $\lambda_{q+1}^{y^N}(\theta)\leq W<\infty, \forall N\in\mathbb{N}$ and $\forall\theta\in\Pi$ and $\lambda_q^y(\theta)=\infty, \forall\theta\in\Pi$.

**Step 1 ($\mathscr{A}(\mathbf{y})$ contains a $q$ dimensional white noise process):**

Let us denote by $p_t^N$ the $q$ normalized dynamic principal components of $y^N$ corresponding to the $q$ largest eigenvalues of $f_{y^N}$, i.e. $p_{j,t}^N=(\lambda_j^{y^N})^{-1}(z)o_j^{N*}(z)y_t^N, j=1,...,q$ or equivalently

$$p_t^N=\Lambda_N^{-1/2}(z)O_N^*(z)y_t^N \tag{2.10}$$

where

$$\Lambda_N(z)=diag[\lambda_1^{y^N}(z),...,\lambda_q^{y^N}(z)] \tag{2.11}$$

and

$$O_N(z)=[o_1^N(z),...,o_q^N(z)] \tag{2.12}$$

In the following products of vectors with matrices and matrices with matrices will be used which should be indexed with their dimensions. As this notation would be awful we will understand a product with different columns and rows as the product with the same matrices by adding zero columns or zero rows to the smaller dimensioned vector or matrix. An example would be $O_N^*(z)f_{y^{N+1}}O_N(z)=\Lambda_N$.

Let $P_N(z)=[o_{q+1}^N(z),...,o_N^N(z)]$ be the matrix with the remaining dynamic eigenvectors of $f_{y^N}$, then

$$\begin{aligned} y_t^N &= [O_N(z)O_N^*(z)+P_N(z)P_N^*(z)]y_t^N \\ &= O_N(z)O_N^*(z)y_t^N+P_N(z)P_N^*(z)y_t^N=O_N(z)\Lambda_N^{1/2}p_t^N+P_N(z)P_N^*(z)y_t^N \end{aligned} \tag{2.13}$$

Obviously the two summands on the right hand side in (2.13) are orthogonal and thus the first summand denotes the orthogonal projection of $y_t^N$ on the space spanned by $p_t^N$. The idea is to project $p_t^N$ on the space spanned by $p_t^M$ with $M>N$, call the projected $p_t^N$ $p_t^{N,M}$, and to show that the distance between $p_t^M$ and $p_t^{N,M}$ tends to zero for $N,M\to\infty$. The limit of the just described sequence will be the desired white noise process. The projection of $p_t^N$ on the space spanned by $p_t^M$ is obtained by pre-multiplying (2.13) by $C(z)\Lambda_N^{-1/2}(z)O_N^*(z)$, with $C\in K_D$ and observing that

$$C(z)\Lambda_N^{-1/2}(z)O_N^*(z)y_t^M=C(z)\Lambda_N^{-1/2}(z)O_N^*(z)y_t^N$$

holds, such that we have

$$\begin{aligned} &C(z)\Lambda_N^{-1/2}(z)O_N^*(z)y_t^M = C(z)\Lambda_N^{-1/2}(z)O_N^*(z)y_t^N \\ &= C(z)p_t^N \\ &= C(z)\Lambda_N^{-1/2}(z)O_N^*(z)O_M(z)\Lambda_M^{1/2}p_t^M + C(z)\Lambda_N^{-1/2}(z)O_N^*(z)P_M(z)P_M^*(z)y_t^M \\ &= E(z)p_t^M + F(z)y_t^M \end{aligned} \tag{2.14}$$

The largest eigenvalue of the spectral density of the remainder $F(z)y_t^M$ of this projection is bounded from above by

$$\lambda_1^{F(z)y_t^M} \leq \lambda_{q+1}^{y^M}/\lambda_q^{y^N} \tag{2.15}$$

which follows from the fact that $[I_M - P_M P_M^*]$ and $[\lambda_{q+1}^{y^M} P_M P_M^* - P_M \tilde{\Lambda}_M P_M^*]$ are positive semi definite, where $\tilde{\Lambda}_M = diag[\lambda_{q+1}^{y^M}, ..., \lambda_M^{y^M}]$, and thus $C\Lambda_N^{-1/2}O_N^*[\lambda_{q+1}^{y^M} P_M P_M^* - P_M \tilde{\Lambda}_M P_M^*]O_N\Lambda_N^{-1/2}C^* = \lambda_{q+1}^{y^M} C\Lambda_N^{-1}C^* - Ff_{y^M}F^*$ is positive semi definite too.

Let $D$ be a subset of $[-\pi, \pi]$ with positive measure such that $\Lambda_q^{y^M}(\theta) \geq \alpha_M, \forall\theta \in D$, where $(\alpha_M)_{M\in\mathbb{N}}$ is a monotonically increasing real sequence with $\lim_{M\to\infty}\alpha_M = \infty$ and $\Lambda_{q+1}^{y^M}(\theta) < W < \infty, \forall\theta \in D$. Furthermore let $C \in K_D$ then regarding the spectral densities of (2.14) for $\theta \in D$

$$I_q = E(\theta)E^*(\theta) + F(\theta)f_{y^M}(\theta)F^*(\theta) \tag{2.16}$$

holds, and thus, using (2.15) gives

$$\lambda_1^{F(z)y_t^M} \leq \lambda_{q+1}^{y^N}/\lambda_q^{y^M} \leq W/\alpha_M \tag{2.17}$$

As a consequence, $1 - W/\alpha_M \leq \lambda_j^{EE^*}(\theta) \leq 1, j = 1, ..., q$, holds, where $E(\theta)E^*(\theta) = H(\theta)\Lambda_{EE^*}H(\theta)^*$, $\Lambda_{EE^*} = diag[\lambda_1^{EE^*}(\theta), ..., \lambda_q^{EE^*}(\theta)]$ and $\lambda_j^{EE^*}(\theta)$ denotes the $j$-th largest eigenvalue of $E(\theta)E^*(\theta)$. Thus the filter

$$G(\theta) = \begin{cases} H(\theta)\Lambda_{EE^*}^{-1/2}H^*(\theta)E(\theta) \text{ for } \theta \in D \\ 0_q \text{ for } \theta \notin D \end{cases} \tag{2.18}$$

is well defined, as there always exists an integer $M_0$ such that $0 < 1 - W/\alpha_{M_0} \leq \lambda_q^{EE^*}(\theta)$, and it is of course an element of $K_D$. Summarizing the results we have the following Lemma.

**Lemma 2.3.4.** *Assume that (I) and (II) hold, let $D$ be a subset of $\Pi$ with positive measure and $(\alpha_M)_{M\in\mathbb{N}}$ a monotonically increasing real sequence with $\lim_{M\to\infty}\alpha_M = \infty$. Furthermore assume that $C \in K_D$ and $\Lambda_q^{y^M}(\theta) \geq \alpha_M, \forall\theta \in D$. Then, given $\tau$ with $0 < \tau < 2$, there exists an integer $M_\tau$ such that*

*$W/M_\tau < 1$ and for $N > M \geq M_\tau$ the largest eigenvalue of the spectral density matrix of*

$$C(z)p_t^M - G(z)p_t^N$$

*is smaller than $\tau$ $\forall\theta \in \Pi$, where $G(\theta)$ is defined by* (2.18).

*Proof.* ***(Lemma 2.3.4.)*** See also [Forni and Lippi, 2001].
Using (2.14) we have

$$C(z)p_t^M - G(z)p_t^N = E(z)p_t^N + F(z)y_t^N - G(z)p_t^N = F(z)y_t^N + [E(z) - G(z)]p_t^N \tag{2.19}$$

As $F(z)y_t^N$ is orthogonal on $p_t^N$ by definition the spectral density of the right hand side of (2.19) is the sum of the spectral densities of its summands. Denoting $S$ as the spectral density of the left hand side of (2.19) and using (2.16) we have for $\theta \in D$

$$\begin{aligned} S &= I_q - E(\theta)E^*(\theta) - E(\theta)E^*(\theta) - E(\theta)G^*(\theta) - G(\theta)E^*(\theta) + \underbrace{G(\theta)G^*(\theta)}_{=I_q} \\ &= 2I_q - E(\theta)G^*(\theta) - G(\theta)E^*(\theta) = 2I_q - 2H\Lambda_{EE^*}^{1/2}H^* = 2H[I_q - \Lambda_{EE^*}^{1/2}]H^* \end{aligned} \tag{2.20}$$

The largest eigenvalue of the right hand side of (2.20) is $2(1 - \lambda_q^{EE^*}) \leq 2W/\alpha_M < 2W/\alpha_{M_\tau} < \tau$ for a suitable $M_\tau$. □

The following Lemma is well known (see [Apostol, 1974]).

**Lemma 2.3.5.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of functions which are elements of $L_2([-\pi,\pi])$ which is convergent in the $L_2$ norm. Then there exists a subsequence $(a_{n_m})_{m\in\mathbb{N}}$ such that $\lim_m a_{n_m}(\theta) = a(\theta)$, a.e. in $[-\pi,\pi]$.*

**Definition 2.3.6. (costationary).** *Let $x_t$ be an $n$ dimensional stationary process, then we say that an $m$ dimensional stationary process $y_t$ is* ***costationary*** *with $x_t$ if the crosscovariance $\mathbb{E}x_{i,t}y_{j,t-k}$ (we only consider centered processes) does not depend on $t$ for all $i, j, k$.*

**Lemma 2.3.7.** *Let the sequences of processes $((a_{n,t})_{t\in\mathbb{Z}})_{n\in\mathbb{N}}$ and $((b_{n,t})_{t\in\mathbb{Z}})_{n\in\mathbb{N}}$, whose components belong to $H_y \forall N \in \mathbb{N}$, be costationary with $(y_t)_{t\in\mathbb{Z}}$ and $\lim_n a_{n,t} = a_t$ and $\lim_n b_{n,t} = b_t$ respectively. Then there exists a sequence of integers $(n_m)_{m\in\mathbb{N}}$ such that*

$$\lim_m S(a_{n_m,t}, b_{n_m,t}) = S(a_t, b_t), \text{ a.e. in } [-\pi,\pi]$$

*where $S(a,b)$ denotes the cross spectral density of $a, b$.*

*Proof.* ***(Lemma 2.3.7.)*** See also [Forni and Lippi, 2001].
Note that $\mathbb{E}a_{n_m,t}b_{n_m,t} = \langle a_{n_m,t}, b_{n_m,t}\rangle = \int_{-\pi}^{\pi} S(a_{n_m,t}, b_{n_m,t})$ holds (see Lemma 2 in [Forni and Lippi, 2001]).

As the inner product is continuous we have that $\lim_m \int_{-\pi}^{\pi} |S(a_{n_m,t}, b_{n_m,t}) - S(a_t, b_t)| = 0$. Therefore $S(a_{n_m,t}, b_{n_m,t})$ converges in $L_1$ norm to $S(a_t, b_t)$ and therefore in $L_2$ norm too, and by Lemma 2.3.5 this implies the a.e. convergence. □

We are now ready to establish the desired result. Lemma 2.3.8 shows that there exists a $q$ dimensional vector process which has a spectral density equal to $I_q$ almost everywhere on a subset $D$ of $\Pi$ and whose components are in the aggregation space $\mathscr{A}(y)$. Lemma 2.3.9 will extend this result to almost everywhere in $[-\pi, \pi]$.

**Lemma 2.3.8.** *Assume that (I) and (II) hold and let $D$ and $(\alpha_M)_{M\in\mathbb{N}}$ be as in Lemma 2.3.4, then there exists a $q$ dimensional vector process $v_t$ with*

- *$v_{j,t}$ is an aggregate for $j = 1, ..., q$*
- *$v_t$ has a spectral density which equals $I_q$ a.e. in $D$ and $0_q$ else*

*Proof.* ***(Lemma 2.3.8.)*** See also [Forni and Lippi, 2001].
The proof is done by defining a sequence recursively which will turn out to be a Cauchy Sequence and thus has a limit. Start with letting $G_1$ be any element of $K_D$, setting $\tau_1 = (1/2)^2$ and $M_{\tau_1}$ such that $2W/\alpha_{M_{\tau_1}} < \tau_1$ holds. After that define $v_t^1 = G_1 p_t^{M_{\tau_1}}$ which has of course a spectral density which equals $I_q$ for $\theta \in D$ and $0_q$ else.

Then set $\tau_2 = (1/2)^4$ and $M_{\tau_2}$ such that $2W/\alpha_{M_{\tau_2}} < \tau_2$ holds. Next determine $E_2$ as in (2.14) by replacing $C$ with $G_1$, $N$ with $M_{\tau_1}$ and $M$ with $M_{\tau_2}$. After that $G_2$ is defined as in (2.18) and $v_t^2 = G_2 p_t^{M_{\tau_2}}$. Again the spectral density of $v_t^2$ equals $I_q$ for $\theta \in D$ and $0_q$ else. Furthermore, by Lemma 2.3.4, we have that the largest eigenvalue of the spectral density of $[v_t^1 - v_t^2]$ is bounded by $\tau_1$ a.e. in $[-\pi, \pi]$, which implies that $\|v_{j,t}^1 - v_{j,t}^2\|^2 < \int_{-\pi}^{\pi} \tau_1 d\theta = 2\pi(1/2)^2, j = 1, ..., q$.

By recursion set $\tau_k = (1/2)^{2k}$ and $M_{\tau_k}$ such that $2W/\alpha_{M_{\tau_k}} < \tau_k$ holds. Determine $E_k$ as in (2.14) by replacing $C$ with $G_{k-1}$, $N$ with $M_{\tau_{k-1}}$ and $M$ with $M_{\tau_k}$. After that $G_k$ is defined as in (2.18) and $v_t^k = G_k p_t^{M_{\tau_k}}$. Again the spectral density of $v_t^k$ equals $I_q$ for $\theta \in D$ and $0_q$ else. Furthermore, by Lemma 2.3.4, we have that the largest eigenvalue of the spectral density of $[v_t^{k-1} - v_t^k]$ is bounded by $\tau_{k-1}$ a.e. in $[-\pi, \pi]$, which implies that $\|v_{j,t}^{k-1} - v_{j,t}^k\|^2 < \int_{-\pi}^{\pi} \tau_k d\theta = 2\pi(1/2)^{k-1}, j = 1, ..., q$.

As

$$\|v_{j,t}^k - v_{j,t}^{k+h}\| \leq \underbrace{\|v_{j,t}^k - v_{j,t}^{k+1}\|}_{<\sqrt{2\pi}\frac{1}{2^k}} + \underbrace{\|v_{j,t}^{k+1} - v_{j,t}^{k+2}\|}_{<\sqrt{2\pi}\frac{1}{2^{k+1}}} + ... + \underbrace{\|v_{j,t}^{k+h-1} - v_{j,t}^{k+h}\|}_{<\sqrt{2\pi}\frac{1}{2^{k+h-1}}} < \sqrt{2\pi}\frac{1}{2^{k-1}}$$

holds, $(v_{j,t}^k)_{k\in\mathbb{N}}, j = 1, ..., q$ are Cauchy sequences. Therefore we can define $v_t$ as the vector of the limits. $v_t$ has a spectral density which equals $I_q$ for $\theta \in D$ and $0_q$ else, as it is the limit of the sequence $(v_t^k)_{k\in\mathbb{N}}$ whose elements have the desired spectral density and Lemma 2.3.7 can be applied.

To see that the components $v_{j,t}, j = 1, ..., q$ are aggregates we have to show that $(\tilde{G}_k)_{k\in\mathbb{N}}$ , with $\tilde{G}_k = G_k \Lambda_{M_{\tau_k}}^{-1/2} O_{M_{\tau_k}}^*$ is a DAS. By the definition of $\tilde{G}_k$ we have

$$\tilde{G}_k \tilde{G}_k^* = G_k \Lambda_{M_{\tau_k}}^{-1} G_k^*$$

and thus

$$\lim_{k\to\infty} \int_{-\pi}^{\pi} [\tilde{G}_k(\theta)]_{[j,]} [\tilde{G}_k(\theta)]_{[j,]}^* d\theta = \lim_{k\to\infty} \int_{-\pi}^{\pi} [G_k(\theta)]_{[j,]} \Lambda_{M_{\tau_k}}^{-1} [G_k(\theta)]_{[j,]}^* d\theta \leq \lim_{k\to\infty} \int_{-\pi}^{\pi} (\lambda_q^{y^{M_{\tau_k}}}(\theta))^{-1} d\theta = 0$$

where $(\tilde{G}_k)_{[j,]}$ denotes the $j$-th row of $\tilde{G}_k$, and the last equality follows from the Lebesgue Convergence Theorem. □

**Lemma 2.3.9.** *Assume that (I) and (II) hold, then there exists a q dimensional white noise vector process $z_t$ with spectral density $I_q$ a.e. in $[-\pi, \pi]$ such that its components $z_{j,t}, j = 1, ..., q$ are aggregates.*

*Proof.* ***(Lemma 2.3.9.)*** See also [Forni and Lippi, 2001].
We start with defining $D_0 = \Pi$. Then a sequence of integers $v_m$ is defined by

$$v_m = \min_N \mathscr{L}(\{\theta \in D_{m-1}, \lambda_q^{y^N}(\theta) > m\}) > \pi$$

and

$$D_m = \{\theta \in D_{m-1}, \lambda_q^{y^{v_m}} > m\}$$

for $m = 1, 2, ....$ Note that $v_m$ is a non-decreasing sequence.
In the next step define $N_1 = \bigcap_{m=1}^{\infty} D_m$. The whole procedure is repeated by setting $D_0$ equal to $\Pi \backslash N_1$ and using $\mathscr{L}(\Pi \backslash N_1)/2$ instead of $\pi$.
For $n > 2$ start the whole procedure again by setting $D_0$ equal to $\Pi \backslash \bigcup_{m=1}^{n} N_m$ and using $\mathscr{L}(\Pi \backslash \bigcup_{m=1}^{n} N_m)/2$ as the boundary. As the $N_m$s have no intersection

$$\mathscr{L}(\bigcup_{m=1}^{\infty} N_m) = \sum_{m=1}^{\infty} \mathscr{L}(N_m) = 2\pi$$

Now Lemma 2.3.8 can be applied by using each $N_m$ together with $\alpha_M = m$ and we obtain a $q$ dimensional vector $z_t^m$ whose components are aggregates and its spectral density matrix equals $I_q$ a.e. in $N_m$ and $0_q$ else. Defining $z_t = \sum_{m=1}^{\infty} z_t^m$, and noting again that the sets $N_m$ have no intersection but the union over all of them has Lebesgue measure $2\pi$, we see that $z_t$ has indeed a spectral density $I_q$ a.e. in

$[-\pi, \pi]$. □

Thus step 1 is proofed.

**Step 2 ($\mathscr{A}(\mathbf{y}) = \mathbf{H_z}$):**
Step 1 already implies $\mathscr{A}(y) \supseteq H_z$, thus we have to show $\mathscr{A}(y) \subseteq H_z$. In order to do this let $w_t$ be an aggregate. If the remainder $r_t$ of the projection

$$w_t = w_t|_{H_z} + r_t \tag{2.21}$$

is zero the result follows. Assume that $r_t$ is not equal to 0, then $r_t$ must be necessarily an aggregate by definition. Next define $\tilde{z}_t = [z_t', r_t]'$ with spectral density matrix $f_{\tilde{z}}$ which has an $I_q$ matrix in the upper left $q \times q$ corner which implies that

$$\det(f_{\tilde{z}}) = f_r \tag{2.22}$$

As all $q+1$ components of $\tilde{z}_t$ are aggregates there exist $q+1$ dynamic averaging sequences $(\mathbf{a}_{j,N})_{N\in\mathbb{N}}, j = 1, ..., q+1$ such that

$$\tilde{z}_{j,t} = \lim_{N\to\infty} a_{j,N}(z)y_t^N, j = 1, ..., q+1 \tag{2.23}$$

holds. Furthermore $\lim_{N\to\infty} \|\mathbf{a}_{j,N}\| = \lim_{N\to\infty} \int_{-\pi}^{\pi} |\mathbf{a}_{j,N}(\theta)|^2 d\theta = 0$ for $j = 1, ..., q+1$, such that Lemma 2.3.5 implies that there exist subsequences which converge to zero almost everywhere in $[-\pi, \pi]$. Analogously there exists a subsequence of the spectral densities of $\tilde{z}_t^N = [a_{1,N}(z)y_t^N, ..., a_{q+1,N}(z)y_t^N]'$, which converges a.e. in $[-\pi, \pi]$ to $f_{\tilde{z}}$ (Lemma 2.3.7).
Next define for $j = 1, ..., q+1$, $\mathbf{b}_{j,N} = \mathbf{a}_{j,N}O_N$ and $\mathbf{c}_{j,N} = \mathbf{a}_{j,N} - \mathbf{b}_{j,N}O_N^*$, such that

$$\mathbf{a}_{j,N} = \mathbf{b}_{j,N}O_N^* + \mathbf{c}_{j,N} \tag{2.24}$$

Note that $O_N$ is the $N \times q$ matrix consisting of the eigenvectors of $f_{y^N}$ belonging to the $q$ largest eigenvalues (see also (2.12)). As

$$[\mathbf{a}_{j,N}O_NO_N^*][\mathbf{a}_{j,N} - \mathbf{a}_{j,N}O_NO_N^*]^* = 0$$

we have

$$|\mathbf{a}_{j,N}|^2 = |\mathbf{b}_{j,N}O_N^* + \mathbf{c}_{j,N}|^2 = \mathbf{b}_{j,N}\underbrace{O_N^*O_N}_{=I_q}\mathbf{b}_{j,N}^* + |\mathbf{c}_{j,N}|^2 = |\mathbf{b}_{j,N}|^2 + |\mathbf{c}_{j,N}|^2 \tag{2.25}$$

The Definition of $\mathbf{b}_{j,N}$ and $\mathbf{c}_{j,N}$ imply that

$$a_{j,N}(z)y_t^N = b_{j,N}(z)O_N^*(z)y_t^N + c_{j,N}(z)y_t^N \tag{2.26}$$

and thus the first summand denotes the orthogonal projection of $a_{j,N}(z)y_t^N$ on the space spanned by the not normalized $q$ dynamic principal components (compare to (2.10)) $O_N^*(z)y_t^N$. Therefore the spectral density matrix of $\tilde{z}^N$, $f_{\tilde{z}^N}$, is the sum of the spectral density of

$$[b_{1,N}(z)O_N^*(z)y_t^N, ..., b_{q+1,N}(z)O_N^*(z)y_t^N]'$$

$f_{\tilde{z}^N}^1$ say, and the spectral density matrix of

$$[c_{1,N}(z)y_t^N, ..., c_{q+1,N}(z)y_t^N]'$$

$f_{\tilde{z}^N}^2$ say.

$$f_{\tilde{z}^N} = f_{\tilde{z}^N}^1 + f_{\tilde{z}^N}^2 \tag{2.27}$$

As $|\mathbf{a}_{j,N}|^2$ converges to zero a.e. in $[-\pi, \pi]$, (2.25) implies that $|\mathbf{c}_{j,N}|^2$ converges to zero a.e. in $[-\pi, \pi]$ too. As a consequence and realizing the fact that $c_{j,N}$ is orthogonal on $o_{j,N}, j = 1, ..., q$ we have that the diagonal elements of $f_{\tilde{z}^N}^2 = \mathbf{c}_{q,N} f_{y^N} \mathbf{c}_{q,N}^*, q = 1, ..., q+1$ are bounded from above by $\lambda_{q+1}^{y^N} |\mathbf{c}_{j,N}|^2$ which converges to zero a.e. in $[-\pi, \pi]$ as $\lambda_{q+1}^y$ is essentially bounded. Thus $f_{\tilde{z}^N}^2$ converges to zero a.e. in $[-\pi, \pi]$.
Furthermore $f_{\tilde{z}^N}^1$ is singular by construction. Consequently $\det(f_{\tilde{z}}) = f_r = 0$ a.e. in $[-\pi, \pi]$ which implies that $r_t = 0$ a.e. in $[-\pi, \pi]$.
Thus step 2 is proofed.

**Step 3 ($\tilde{\mathbf{u}}_\mathbf{t}$ is weakly dependent):**

**Definition 2.3.10. (Cauchy sequence of spaces).** *Assume that for all $N \in \mathbb{N}$ the following holds: Let $(z_t^N)_{t\in\mathbb{Z}}$ be a q dimensional white noise process belonging to $H_y$ and costationary with $y_t$, which implies that $z_t^N$ is costationary with $z_t^M, \forall N, M \in \mathbb{N}$. Consider the orthogonal projection*

$$z_t^M = A_{M,N}(z)z_t^N + \delta_{M,N} \tag{2.28}$$

*then the sequence $((z_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ generates a **Cauchy sequence of spaces** if for a given $\varepsilon > 0$, for $\theta$ a.e. in $[-\pi, \pi]$ there exists an integer $M_\varepsilon(\theta)$ such that $\forall N, M > M_\varepsilon(\theta)$, $tr(f_{\delta_{M,N}}) < \varepsilon$.*

The outline of the proof of step 3 is as follows. First, Lemma 2.3.11 shows that $((p_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ generates a Cauchy sequence of spaces. Second, Lemma 2.3.12 states that the sequence of projections of $y_t$ on $(p_t^N)_{N\in\mathbb{N}}$ converges in $H_y$. Third, Theorem 2.3.13 establishes that the limit of the sequence just described, converges to the latent variables, which are the components $y_{i,t}$ projected on the aggregation space $\mathscr{A}(y)$. And finally, Lemma 2.3.14 concludes the proof of step 3 and thus of Theorem 2.3.1 by showing that $\tilde{u}_t$ is idiosyncratic.

**Lemma 2.3.11.** *The sequence of dynamic principal components $((p_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$ generates a Cauchy sequence of spaces.*

*Proof.* ***(Lemma 2.3.11.)*** See also [Forni and Lippi, 2001].
Recall formula (2.14) where $M > N$ holds

$$C(z)p_t^N = E(z)p_t^M + F(z)y_t^M$$

Set $C(z) = I_q$ and let $\delta_{N,M} = F(z)y_t^M$, then we have

$$p_t^N = E(z)p_t^M + \delta_{N,M} \tag{2.29}$$

Furthermore recall formula (2.15)

$$\lambda_1^{F(z)y_t^M} = \lambda_1^{\delta_{N,M}} \leq \lambda_{q+1}^{y^M}/\lambda_q^{y^N}$$

Letting $M$ and $N$ go to infinity, $\lambda_1^{\delta_{N,M}}$ goes to zero a.e. in $[-\pi,\pi]$ and thus the trace of the $q\times q$ spectral density of $\delta_{N,M}$, which is the sum of its $q$ eigenvalues, goes to zero too a.e. in $[-\pi,\pi]$.
For $N > M$ we have

$$p_t^M = E(z)^* p_t^N + \delta_{M,N} \tag{2.30}$$

taking the spectral densities of (2.30) and (2.29) we have

$$f_{p^M} = E(z)E(z)^* + f_{\delta_{N,M}} = E(z)^*E(z) + f_{\delta_{M,N}},\ \text{a.e. in } [-\pi,\pi] \tag{2.31}$$

Consequently we have

$$tr(f_{\delta_{N,M}}) = tr(f_{\delta_{M,N}})$$

and the result follows. □

**Lemma 2.3.12.** *Let $\hat{y}_{i,t}^N$ denote the orthogonal projection of $y_{i,t}$ on the space spanned by $(p_t^N)_{t\in\mathbb{Z}}$, i.e. $\hat{y}_{i,t}^N = y_{i,t}|_{\overline{span}\{p_{i,t}^N | 1\leq i\leq q, t\in\mathbb{Z}\}}$, then the sequence $(\hat{y}_{i,t}^N)_{N\in\mathbb{N}}$ converges in $H_y$.*

*Proof.* ***(Lemma 2.3.12.)*** See also [Forni and Lippi, 2001].
Define

$$r_{i,t}^N = y_{i,t} - \hat{y}_{i,t}^N = y_{i,t} - A_{i,N}(z)p_t^N$$

then we have

$$\hat{y}_{i,t}^N - \hat{y}_{i,t}^M = A_{i,N}(z)p_t^N - A_{i,M}(z)p_t^M = r_{i,t}^M - r_{i,t}^N$$

The spectral density of $\hat{y}_{i,t}^N - \hat{y}_{i,t}^M$ is

$$f_{\hat{y}_i^N - \hat{y}_i^M} = S(\hat{y}_{i,t}^N - \hat{y}_{i,t}^M, r_{i,t}^M - r_{i,t}^N) = S(\hat{y}_{i,t}^N, r_{i,t}^M) + S(\hat{y}_{i,t}^M, r_{i,t}^N) \tag{2.32}$$

where $S(x_t, y_t)$ denotes the cross spectral density between $x_t$ and $y_t$. Using (2.30) we can write

$$\begin{aligned} S(\hat{y}_{i,t}^M, r_{i,t}^N) &= S(A_{i,M}(z)p_t^M, r_{i,t}^N) = S(A_{i,M}(z)[E(z)^* p_t^N + \delta_{M,N}], r_{i,t}^N) \\ &= S(A_{i,M}(z)E(z)^* p_t^N, r_{i,t}^N) + S(A_{i,M}(z)\delta_{M,N}, r_{i,t}^N) = S(A_{i,M}(z)\delta_{M,N}, r_{i,t}^N) \end{aligned} \tag{2.33}$$

Now as the squared entries of $A_{i,M}(z)$ and the spectral density of $r_{i,t}^N$ are bounded in modulus by the spectral density of $y_{i,t}$ and $tr(f_{\delta_{M,N}}) \to 0$ for $M, N \to \infty$ holds, as $p_t^N$ generates a Cauchy sequence of spaces, $S(A_{i,M}(z)\delta_{M,N}, r_{i,t}^N)$ converges to 0 a.e. in $[-\pi, \pi]$. The same holds for $S(\hat{y}_{i,t}^N, r_{i,t}^M)$ and we have shown that $f_{\hat{y}_i^N - \hat{y}_i^M} \to 0$ for $M, N \to \infty$.
Since both $f_{\hat{y}_i^N}$ and $f_{\hat{y}_i^N}$ are dominated by $f_{y_i}$, the integral of the spectral density $f_{\hat{y}_i^N - \hat{y}_i^M}$ also converges to zero by the Lebesgue Convergence Theorem. Hence $\hat{y}_{i,t}^N - \hat{y}_{i,t}^M$ is a Cauchy sequence and thus converges. □

We have already shown that the projection of the components of $y_t$ onto the spaces spanned by the dynamic principal components converge in $H_y$. The next Theorem shows that the limit is actually what we expect, namely the projection of $y_t$ onto the aggregation space $\mathscr{A}(y)$.

**Theorem 2.3.13.** *For all $i \in \mathbb{N}$*

$$\lim_{N\to\infty} \hat{y}_{i,t}^N = \lim_{N\to\infty} O_{i,N}(z)O_N^*(z)y_t^N = y_{i,t}|_{\mathscr{A}(y)}$$

*in mean square holds.*

*Proof.* ***(Theorem 2.3.13.)*** See also [Forni and Lippi, 2001].
We have already proofed that $\lim_{N\to\infty} \hat{y}_{i,t}^N = \lim_{N\to\infty} O_{i,N}(z)O_N^*(z)y_t^N =: \tilde{\chi}_{i,t} \in H_y$, in mean square holds (Lemma 2.3.11 and Lemma 2.3.12), where $O_{i,N}(z)$ denotes that $i$-th row of $O_N(z)$.
Note that $(O_{i,N}(z)O_N^*(z))_{N\in\mathbb{N}}$ is a DAS, as

$$\begin{aligned} \lim_{N\to\infty} \|O_{i,N}(z)O_N^*(z)\| &= \lim_{N\to\infty} \int_{-\pi}^{\pi} O_{i,N}(\theta)O_N^*(\theta)O_N(\theta)O_{i,N}^*(\theta)d\theta \\ &= \lim_{N\to\infty} \int_{-\pi}^{\pi} O_{i,N}O_{i,N}^*(\theta)d\theta \leq \lim_{N\to\infty} \int_{-\pi}^{\pi} f_{y_i}(\theta)/\lambda_q^{y^N}(\theta)d\theta = 0 \end{aligned} \tag{2.34}$$

the inequality in (2.34) holds because of

$$O_{i,N}(z)\Lambda_N(z)O_{i,N}^*(z) \geq \lambda_q^{y^N}(z)O_{i,N}(z)O_{i,N}^*(z) \tag{2.35}$$

$$O_{i,N}(z)\Lambda_N(z)O_{i,N}^*(z) = O_{i,N}(z)\Lambda_N^{1/2}(z)\underbrace{f_{p^N}}_{=I_q}\Lambda_N^{1/2}(z)O_{i,N}^*(z) = f_{\hat{y}_i^N} \leq f_{y_i} \tag{2.36}$$

and the last equality in (2.34) holds because of the Lebesgue Convergence Theorem. As $\tilde{\chi}_{i,t} \in H_y$ holds, the residuals $\tilde{u}_{i,t} = \lim_{N\to\infty} \tilde{u}_{i,t}^N = [y_{i,t}^N - \lim_{N\to\infty} \hat{y}_{i,t}^N]$ are also elements of $H_y$. As $\tilde{u}_{i,t}^N$ is orthogonal on $(p_t^N)_{t\in\mathbb{Z}}$ by construction, the continuity of the inner product ensures that $\tilde{u}_{i,t}$ is orthogonal on $\mathscr{A}(y)$. As the orthogonal projection is unique the result follows. □

**Lemma 2.3.14.** *$\tilde{u}_t = [\tilde{u}_{1,t}, ..., \tilde{u}_{2,t}, ...]'$ is weakly dependent, where $\tilde{u}_{i,t} = y_{i,t} - y_{i,t}|_{\mathscr{A}(y)}$.*

*Proof.* ***(Lemma 2.3.14.)*** See also [Forni and Lippi, 2001].
Let $f_{\tilde{u}_N^M}$ for $N > M$ be the $M \times M$ spectral density matrix of $\tilde{u}_{t,N}^M = [\tilde{u}_{1,t}^N, ..., \tilde{u}_{M,t}^N]'$ and $\lambda_1^{N,M}$ be its largest eigenvalue. As $\tilde{u}_{i,t}^N$ converges to $\tilde{u}_{i,t}$ in mean square $\forall i \in \mathbb{N}$, by Theorem 2.3.13, we have that by Lemma 2.3.7 a subsequence of $(f_{\tilde{u}_N^M})_{N\in\mathbb{N}}$ converges to $f_{\tilde{u}^M}$ a.e. in $[-\pi, \pi]$, where $\tilde{u}_t^M = [\tilde{u}_{1,t}, ..., \tilde{u}_{M,t}]'$. Without loss of generality we can assume that $\lim_{N\to\infty} f_{\tilde{u}_N^M} = f_{\tilde{u}^M}$ holds a.e. in $[-\pi, \pi]$. As the eigenvalues are continuous functions of the matrices we have that

$$\lim_{N\to\infty} \lambda_1^{N,M}(\theta) = \lambda_1^{\tilde{u}^M}(\theta) \text{ a.e. in } [-\pi, \pi] \tag{2.37}$$

Furthermore we have that

$$\lambda_1^{N,M}(\theta) \leq \lambda_1^{N,N}(\theta) = \lambda_{q+1}^{y^N}(\theta), \forall\theta \in [-\pi, \pi] \tag{2.38}$$

where the inequality holds because the eigenvalues increase if additional rows and columns appear, and the equality holds by the definition of $\tilde{u}_{t,N}^N = y_t^N - O_N(z)O_N^*(z)y_t^N$. Hence we have by (2.37) and (2.38)

$$\lim_{N\to\infty} \lambda_1^{N,M}(\theta) \leq \lim_{N\to\infty} \lambda_{q+1}^{y^N}(\theta) \tag{2.39}$$

$$\Rightarrow \lambda_1^{\tilde{u}^M}(\theta) \leq \lambda_{q+1}^{y}(\theta) \tag{2.40}$$

$$\Rightarrow \lambda_1^{\tilde{u}}(\theta) \leq \lambda_{q+1}^{y}(\theta) \text{ a.e. in } [-\pi, \pi] \tag{2.41}$$

Thus $\lambda_1^{\tilde{u}}(\theta)$ is essentially bounded and therefore $\tilde{u}_t$ is weakly dependent. □

Summarizing we have that (I) and (II) imply that

$$y_{i,t} = y_{i,t}|_{\mathscr{A}(y)} + \tilde{u}_{i,t}$$
$$y_{i,t} = \tilde{\chi}_{i,t} + \tilde{u}_{i,t}$$

holds, where $\tilde{u}_t$ is weakly dependent. Furthermore $\lambda_q^{\tilde{\chi}} = \lim_{N\to\infty} \lambda_q^{\tilde{\chi}^N} \geq \lim_{N\to\infty} \lambda_q^{y^N} - \lim_{N\to\infty} \lambda_1^{\tilde{u}^N} = \infty$ holds, as the inequality follows by Theorem 1 page 301 in [Lancaster and Tismenetsky, 1985], which finishes the proof of Theorem 2.3.1. □

### 2.3.1 Identification of the Latent Variables

A very important consequence of Theorem 2.3.1 is, that the components $\chi_{i,t}$, and thus $u_{i,t}$, are uniquely determined which will be summarized in the following Theorem.

**Theorem 2.3.15.** *Let $y_t = (y_{i,t})_{i\in\mathbb{N},t\in\mathbb{Z}}$ have a Generalized Dynamic Factor Model representation fulfilling Definition 2.2.5, then*

$$\chi_{i,t} = y_{i,t}|_{\mathscr{A}(y)}$$

*Proof.* ***(Theorem 2.3.15.)*** See also [Forni and Lippi, 2001].
On the one hand we have by Definition 2.2.5

$$y_{i,t} = \chi_{i,t} + u_{i,t} = w_i(z)\varepsilon_t + u_{i,t} \tag{2.42}$$

on the other hand Theorem 2.3.1 shows that

$$y_{i,t} = y_{i,t}|_{\mathscr{A}(y)} + \tilde{u}_{i,t} = \tilde{w}(z)z_t + \tilde{u}_{i,t} \tag{2.43}$$

with $z_t$ defined in Lemma 2.3.9, holds. As $u_t$ is weakly dependent we have $\mathscr{A}(y) \subseteq H_\chi$. Additionally $H_\chi \subseteq H_\varepsilon$ holds by construction and therefore we have

$$H_z = \mathscr{A}(y) \subseteq H_\chi \subseteq H_\varepsilon$$

as $z_t$ and $\varepsilon_t$ are both $q$ dimensional white noise processes we have the equality $H_z = H_\varepsilon$ which implies $\chi_{i,t} \in \mathscr{A}(y)$. As $u_{i,t}$ is orthogonal on $\mathscr{A}(y)$ the result follows. □

# Chapter 3

# Structure Theory of GDFMs

In the previous chapter we discussed the result that the observations have a representation as a Generalized Dynamic Factor Model if and only if the largest eigenvalues of their spectral density diverge and the remaining eigenvalues are bounded. Furthermore the latent variables are then uniquely determined by the projection of the observations on its aggregation space.
Here we want to analyze what can be said about the transfer function corresponding to Wold decomposition of the latent variables in the case that the spectral density of the latent variables is rational. The general idea behind that theory, which uses the spectral density of an unobserved process, is, that useful insights concerning the "true" system are used afterwards for estimation procedures.

From now on, in this section, we will deal with the latent variables $\chi_t$ of $y_t$ exclusively.

## 3.1 Spectral Factorization

Recall Assumption 2 which states that the spectral density of the latent variables $f_\chi(\lambda)$ is rational and has rank $q$ $\forall\lambda \in [-\pi, \pi]$. That means that there exists an $N_0$ such that for all $N \geq N_0$ the spectral density $f_{\chi^N}$ is a rational matrix of rank $q$. The question is how we factorize $f_{\chi^N}$ in order to get a "good" factorization. We will use the result of [Rozanov, 1967], which has been slightly corrected by [Hannan, 1970], that we can always factorize such a spectral density in spectral factors with no zeros and poles inside and on the unit circle, which are unique up to post-multiplication by constant orthogonal matrices. For the sake of simplicity we will omit the superscript $N$ until the contrary is stated explicitly.

**Theorem 3.1.1.** *An $(N \times N)$ dimensional rational spectral density matrix $f_\chi$ of rank $q \leq N$ can be factorized as*

$$f_\chi(\lambda) = \frac{1}{2\pi} w(e^{-i\lambda}) w^*(e^{-i\lambda}) \tag{3.1}$$

*where the $(N \times q)$ dimensional spectral factor $w(z)$ $z \in \mathbb{C}$ is a rational full column rank matrix which*

*has no poles and zeros for* $|z| \leq 1$.

*Proof.* ***(Theorem 3.1.1.)***
See e.g. [Rozanov, 1967] Theorem 10.1. page 47 (or [Hannan, 1970]). □

To select a unique spectral factor which fulfills the Theorem above (two spectral factors fulfilling Theorem (3.1.1) are related by post-multiplication by a constant orthogonal matrix) one can use the following proposition.

**Proposition 3.1.2.** *Let* $A \in \mathbb{R}^{p\times m}$ *with* $p \geq m$ *and* $\mathrm{rk}(A) = m$*, then there exists a unique factorization of* $A = RQ$ *where* $Q \in \mathbb{R}^{m\times m}$ *is orthogonal and* $R \in \mathbb{R}^{p\times m}$ *is a quasi lower triangular matrix, i.e. if the first row of* $A$ *is not zero then* $r_{11} \neq 0$*, and* $r_{1j} = 0, j > 1$*, where* $r_{ij}$ *is the* $[i, j]$ *element of* $R$*, otherwise* $r_{1j} = 0, j \geq 1$*. If the second row of* $A$ *is linearly independent of the first, then* $r_{22} \neq 0$ *and* $r_{2j} = 0, j > 2$ *and otherwise* $r_{2j} = 0, j \geq 2$ *etc.*

*Proof.* ***(Proposition 3.1.2.)***
The proof is straight forward from the following observation: Note that there exists an $(m \times m)$ sub matrix of $A$ which is regular and therefore has a unique QR description satisfying $sign(r_{ii}) = 1$. To be more precise one has to pick the first regular sub matrix of $A$ to achieve the result. □

Up to now we have discussed how we get a unique $q$ dimensional full column rank spectral factor $w(z)$ of the spectral density $f_\chi$. Note that $w(z)$ satisfies all criteria for a Wold representation

$$\chi_t = w(z)\varepsilon_t = \sum_{j=0}^{\infty} w_j \varepsilon_{t-j} \tag{3.2}$$

namely

- $\sum_{j=0}^{\infty} \|w_j\|^2 < \infty$ for some matrix norm $\|\cdot\|$, which is fulfilled if $\|\cdot\|$ denotes the Frobenius Norm as $f_\chi = \frac{1}{2\pi} w(e^{-i\lambda}) w^*(e^{-i\lambda})$ is integrable.
- $H_\chi = H_\varepsilon$. $H_\chi \subseteq H_\varepsilon$ is obvious and $H_\chi \supseteq H_\varepsilon$ holds because

  $$w^-(z) := v^{-1}(z)[l'(z)l(z)]^{-1}l'(z)u^{-1}(z)$$

  is causal, where $w = ulv$ denotes the Smith-Mc Millan form (compare to (9.16)) of $w$. Note that the white noise process $\varepsilon_t$ is a dynamic factor (compare to equation (2.4)).

**Realizations of the transfer function**

There are three different possibilities to realize an $(N \times q)$ rational transfer function $w(z)$ of full column rank and without zeros and poles within and on the unit circle. The first one is an ***ARMA*** realization

$$w(z) = a(z)^{-1}b(z) \tag{3.3}$$

where $a(z)$ and $b(z)$ are polynomial matrices of dimensions $(N \times N)$ and $(N \times q)$ respectively. Furthermore the matrices $a(z), b(z)$ can be chosen to be left coprime (Definition 9.1.2) and fulfill the stability assumption

$$\det(a(z)) \neq 0, |z| \leq 1 \tag{3.4}$$

as well as the mini-phase assumption

$$\mathrm{rk}(b(z)) = q, |z| \leq 1 \tag{3.5}$$

Another possibility is a ***Right Matrix Fraction Description*** as is used in [Forni et al., 2005]

$$w(z) = c(z)d^{-1}(z) \tag{3.6}$$

where $c(z)$ and $d(z)$ are polynomial matrices of dimensions $(N \times q)$ and $(q \times q)$ respectively which are right coprime (Definition 9.1.2).

Our approach is to use a ***State Space realization***

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{3.7}$$
$$y_t = Hx_t \tag{3.8}$$

with

$$w(z) = H[I - Fz]^{-1}G \tag{3.9}$$

where $F \in \mathbb{R}^{n\times n}, G \in \mathbb{R}^{n\times q}, H \in \mathbb{R}^{N\times n}$, the stability condition

$$\lambda_1(F) < 1 \tag{3.10}$$

is fulfilled and the mini-phase condition (compare to Lemma 9.2.7)

$$\mathrm{rk}\begin{bmatrix} I - Fz & -G \\ H & 0 \end{bmatrix} = n + q, |z| \leq 1 \tag{3.11}$$

holds.

In the following section we will describe how a "good" state space realization can be chosen for the spectral factor $w(z)$.

## 3.2 State Space Realization of the Latent Variables

We commence from a (unique) spectral factor of the spectral density of $f_\chi$. The first step will be to show how a minimal state space system (3.7), (3.8) for the latent variables $\chi_t$ can be achieved. Therefore we will restate the Theorem 2.4.1 in [Hannan and Deistler, 1988] and adopt it to our system.

**Definition 3.2.1. (Hankel matrix).** *Let $w(z) = \sum_{j=0}^{\infty} w_j z^j$ be an $(N \times q)$ transfer function than the infinite dimensional matrix*

$$\mathscr{H} = \begin{bmatrix} w_0 & w_1 & w_2 & \dots \\ w_1 & w_2 & w_3 & \dots \\ w_2 & w_3 & \dots & \\ \vdots & & & \end{bmatrix}$$

*is called the **Hankel matrix** of $w(z)$.*

Furthermore let us define $\tilde{w}(z) := z^{-1}w(z^{-1}) = \sum_{j=0}^{\infty} w_j z^{-j-1} = \tilde{a}^{-1}(z)\tilde{b}(z)$, which will be very useful for the theorem below. It is clear that $w(z)$ and $\tilde{w}(z)$ are in a one-to-one relation. Additionally if the polynomial matrices $\tilde{a}(z), \tilde{b}(z)$ are left coprime, the degree of $det(\tilde{a})$ is called the *order* of $\tilde{w}$.

Let $h[i,j]$ denote the $j$-th row in the $i$-th block of rows of $\mathscr{H}$, i.e. $h[i,j]$ is the $((i-1)N+j)$-th row of $\mathscr{H}$.

**Theorem 3.2.2.** *Let $w(z)$ be an $N \times q$ transfer function than the corresponding Hankel matrix $\mathscr{H}$ has the following properties:*

*(i) If $h[i,j]$ is in the linear span of $\{h[i_1,j_1],...,h[i_k,j_k]\}$, then $h[i+1,j]$ is in the linear span of $\{h[i_1+1,j_1],...,h[i_k+1,j_k]\}$.*

*(ii) The rank of $\mathscr{H}$ is finite if and only if $w$ is rational.*

*(iii) The order $n$ of $\tilde{w}(z)$ is equal to the rank of $\mathscr{H}$.*

*Proof.* ***(Theorem 3.2.2.)***
See Theorem 2.4.1. in [Hannan and Deistler, 1988] page 51. □

The following theorem completes our preparation for establishing a minimal state space system for the latent variables.

**Theorem 3.2.3.** *Let $[F, G, H]$ be a state space system with $F \in \mathbb{R}^{n\times n}$, then $\mathscr{H}$ has rank smaller than or equal to $n$. If $[F, G, H]$ is minimal equality holds.*

*Proof.* ***(Theorem 3.2.3.)***
See Theorem 2.3.2. in [Hannan and Deistler, 1988] page 47. □

Let us summarize where we have ended up. We have started with the spectral density of the latent variables and have got a spectral factor which is rational and has no zeros and poles within and on the unit circle. Now from Theorem 3.2.2 we know that the Hankel matrix $\mathscr{H}$ has finite rank $n$ say. Additionally Theorem 3.2.3 tells us, that if we have an $[F, G, H]$ system, it is minimal if its state dimension is equal to $n$. All we have to show is that there exists an $[F, G, H]$ system for $w(z)$ which has a state dimension $n$ and this will be shown in the following.

**Construction of a (canonical) State Space Model**

The algorithm has been introduced by [Ho and Kalman, 1966] and [Akaike, 1974] and has been introduced in the framework of GDFMs in [Zinner, 2008] and [Deistler et al., 2010a]. A detailed discussion about State Space realizations and of this algorithm in particular can be found in [Hannan and Deistler, 1988] and [Deistler, 2001].

At this point we want to use the index $N$ again, indicating the dependence on the dimension of the cross-sectional dimension of $\chi_t^N$.
From (3.2) we know that $\chi_t^N = w^N(z)\varepsilon_t$ is a Wold representation, i.e. $\varepsilon_t$ is the "driving noise" and $w^N(0)\varepsilon_t$ the one-step-ahead prediction error of $\chi_t$. Note that $\varepsilon_t$ does not depend on $N$. The reason for this is, that by Assumption 2 the rank of $f_{\chi^N}$ is constant from a certain $N$ onwards. It is obvious that the sequence of spectral matrices $(f_{\chi^N})_{N\in\mathbb{N}}$ is nested and thus the sequence of spectral factors $(w^N)_{N\in\mathbb{N}}$ is nested too. Therefore we have

$$\begin{bmatrix} \chi_t^N \\ \chi_{t+1}^N \\ \chi_{t+2}^N \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} w_0^N & w_1^N & w_2^N & \dots \\ w_1^N & w_2^N & w_3^N & \dots \\ w_2^N & w_3^N & \dots & \\ \vdots & & & \end{bmatrix}}_{=\mathscr{H}^N} \underbrace{\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{bmatrix}}_{=:\varepsilon_t^-} + \begin{bmatrix} 0 \\ w^N(0)\varepsilon_{t+1} \\ w^N(1)\varepsilon_{t+1} + w^N(0)\varepsilon_{t+2} \\ \vdots \end{bmatrix} \tag{3.12}$$

and thus the predictor for $[\chi_t^{N'}, \chi_{t+1}^{N'}, \chi_{t+2}^{N'}, \dots]'$ is

$$\begin{bmatrix} \chi_t^N \\ \chi_{t+1|t}^N \\ \chi_{t+2|t}^N \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} w_0^N & w_1^N & w_2^N & \dots \\ w_1^N & w_2^N & w_3^N & \dots \\ w_2^N & w_3^N & \dots & \\ \vdots & & & \end{bmatrix}}_{=\mathscr{H}^N} \underbrace{\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{bmatrix}}_{=:\varepsilon_t^-} = \mathscr{H}^N \varepsilon_t^- \tag{3.13}$$

**Assumption 4.** *The matrix $\mathscr{H}$ belonging to the $\infty \times q$ dimensional transfer function $w(z)$, such that $\chi_t = w(z)\varepsilon_t$ holds, has finite rank $n$.*

This assumption is important in our framework and excludes the case like in [Forni and Lippi, 2009] where the state dimension can be infinity. It means that the dynamics of the transfer function are limited, i.e. additional observations do not change the dynamics from a certain $N$ onwards.

Let $N$ be large enough such that the rank of $f_{\chi^N}$ is already $q$ and the rank of $\mathscr{H}^N$ is already $n$, the latter is also reasonable as the sequence $(n_N)_{N\in\mathbb{N}}$ is bounded by Assumption 4 and non-decreasing.

As the rank of $\mathscr{H}^N$ is $n$ we know that there exist $n$ linearly independent rows of $\mathscr{H}^N$. We choose the first linearly independent rows of $\mathscr{H}^N$ and define a selector matrix $S^N$, i.e. $S^N$ is an $n \times \infty$ dimensional matrix which has in each row one entry equal to one and zeros elsewhere. According to that selection we can define the Kronecker Indices $\{\alpha_1^N, \alpha_2^N, ...\}$ where $\alpha_i^N = j$ means that the $i$-th row of the first $j$ blocks of rows of $\mathscr{H}^N$ are in the basis. So we can define the $n$ dimensional state

$$x_t := S^N \mathscr{H}^N \varepsilon_t^- = \mathscr{H}_\alpha \varepsilon_t^- \tag{3.14}$$

Note that $x_t$ can be chosen independently from $N$ as the same rows which build a basis of $\mathscr{H}^N$ build a basis for $\mathscr{H}^{N+1}$ too. Further note that we can assume without any loss of generality that the Kronecker Indices do not depend on $N$ as by Assumption 4 they do not change from some $N_0$ onwards (which does

not need to be same $N_0$ from which on $n_{N_0} = n$ holds). Thus we can define $\mathscr{H}_\alpha = S^N \mathscr{H}^N$.

$$\begin{aligned} x_{t+1} = S^N \mathscr{H}^N \varepsilon_{t+1}^- &= S^N \begin{bmatrix} w_0^N & w_1^N & w_2^N & \dots \\ w_1^N & w_2^N & w_3^N & \dots \\ w_2^N & w_3^N & \dots & \\ \vdots & & & \end{bmatrix} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \end{bmatrix} \\ &= S^N \begin{bmatrix} w_1^N & w_2^N & \dots \\ w_2^N & w_3^N & \dots \\ w_3^N & \dots & \\ \vdots & & \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \end{bmatrix} + S^N \begin{bmatrix} w_0^N \\ w_1^N \\ w_2^N \\ \vdots \end{bmatrix} \varepsilon_{t+1} \\ &= S^N F^N \mathscr{H}_\alpha \varepsilon_t^- + S^N \begin{bmatrix} w_0^N \\ w_1^N \\ w_2^N \\ \vdots \end{bmatrix} \varepsilon_{t+1} \\ &= F x_t + G \varepsilon_{t+1} \end{aligned} \tag{3.15}$$

where $F^N \mathscr{H}_\alpha = \begin{bmatrix} w_1^N & w_2^N & \dots \\ w_2^N & w_3^N & \dots \\ w_3^N & \dots & \\ \vdots & & \end{bmatrix}$ holds as the latter is a sub matrix of $\mathscr{H}^N$ and can therefore be expressed by the basis rows of $\mathscr{H}^N$. As $S^N$ will always pick the same rows (for growing $N$), $F = S^N F^N$ and $G = S^N \begin{bmatrix} w_0^N \\ w_1^N \\ w_2^N \\ \vdots \end{bmatrix}$ can be chosen independently of $N$. Finally $H^N$ is defined by expressing the first block of rows of $\mathscr{H}^N$ by the basis rows $\mathscr{H}_\alpha$

$$[w_0^N, w_1^N, w_2^N, ...] = H^N \mathscr{H}_\alpha \tag{3.16}$$

So we have established a state space realization for $w^N(z)$ with state dimension equal to $n$ (the rank of the Hankel matrix), which implies that the system is minimal. Next we want to show that the system is indeed stable. Note that it is not sufficient to argue that the poles of $w^N(z) = H^N[I - Fz]^{-1}G$ lie outside the unit circle and therefore the eigenvalues of $F$ have to be smaller than one, as unstable eigenvalues can cancel out. But as the system is minimal, we will see that $F$ is indeed stable:

In the construction of our state space system we use (3.14) to define the state. Obviously $x_t$ is stationary by construction. Thus

$$\mathbb{E}x_t x_t' = \gamma_0^x = F\mathbb{E}x_{t-1}x_{t-1}'F' + G\mathbb{E}\varepsilon_t\varepsilon_t'G' = F\gamma_0^x F' + G\Sigma G' \tag{3.17}$$

Suppose that $y$ is an eigenvector of $F$ with eigenvalue $\lambda, |\lambda| \geq 1$, i.e. $y'F = \lambda y'$, then

$$y'\gamma_0^x\bar{y} - y'F\gamma_0^x F'\bar{y} = y'\gamma_0^x\bar{y} - \lambda y'\gamma_0^x F'\bar{y}\lambda^* = \underbrace{(1-|\lambda|^2)}_{\leq 0}\underbrace{y'\gamma_0^x\bar{y}}_{\geq 0} = \underbrace{y'G\Sigma G'\bar{y}}_{\geq 0} \tag{3.18}$$

As the right hand side must be zero, $y'$ is in the (left) kernel of $G$. Thus $y'[\lambda I - F, G] = 0$ must hold which is equivalent (see e.g. [Kailath, 1980] Theorem 2.4-9. page 136) to $y'[G, FG, ..., F^{n-1}G] = 0$, which is a contradiction to the reachability of the pair $F, G$. Thus $F$ must be stable.

Note that $\gamma_0^x$ is regular as the stability of $F$ implies that $x_t = [I - Fz]^{-1}G\varepsilon_t = \sum_{j=0}^{\infty} F^j G\varepsilon_{t-j}$ holds. Therefore the covariance can be decomposed in

$$\mathbb{E}x_t x_t' = [G, FG, F^2G, ...]\begin{bmatrix}\Sigma & 0 & \dots & \\ 0 & \Sigma & 0 & \dots \\ 0 & 0 & \Sigma & 0 \\ \vdots & \ddots & \ddots & \ddots\end{bmatrix}\begin{bmatrix}G' \\ G'F' \\ G'F'^2 \\ \vdots\end{bmatrix} \tag{3.19}$$

where the first and the last matrix on the right hand side are of full row and column rank (reachability), such that $\gamma_0^x$ is regular.

Summarizing we have shown that an $[F, G, H^N]$ state space system (3.7), (3.8) for $w^N(z)$ which is minimal, stable and mini-phase (as $w^N(z) = H^N[I - Fz]^{-1}G$ the $[F, G, H^N]$ representation must be mini-phase) always exists.
Let us shortly discuss the relation between the dimensions of the observations and the latent variables $N$, of the state $n$ and of the driving noise (dynamic factors) $q$. As $N$ goes to infinity and $n$ is bounded for $N$ going to infinity $N \geq n$ holds. Furthermore $n \geq q$ is fulfilled by construction. Note that the state is constructed by selecting the first linearly independent rows of the Hankel matrix, and as $w(z)$ has no zeros within the unit circle $w(0)$ is of full rank $q$. Obviously this implies that the rank of the Hankel matrix is greater than or equal to $q$.
Summarizing we have proved the following theorem.

**Theorem 3.2.4.** *Let $(w^N(z))_{N\in\mathbb{N}}$ be a nested sequence of rational transfer functions of dimensions $(N \times q)$ which have rank q and have no zeros and poles within and on the unit circle. Let $N$ be large enough such that Assumption 4 is fulfilled. Then there exist minimal stable state space realizations* (3.7)*,*

(3.8) *of* $w^N(z) = H^N[I - Fz]^{-1}G$*, such that* $F$ *and* $G$ *are independent of* $N$ *and the sequence* $(H^N)_{N\in\mathbb{N}}$ *is nested.*

In the following we will omit the superscript $N$ again, for the sake of simplicity, until the contrary is stated explicitly.

Up to now we know, how we can define our desired minimal $[F, G, H]$ system. Unfortunately, we cannot guarantee that the matrix $H$ is of full column rank, which means that not all information of the state is available from $\chi_t$ and therefore, we will define so called static factors.

### 3.2.1 Static Factors

In (2.4) and (3.2) we introduced already dynamic factors which we obtained by a spectral factorization of the spectral density matrix of the latent variables. Analogously we define static factors by a factorization of the zero lag covariance matrix of the latent variables.

$$\gamma_0^\chi = \mathbb{E}\chi_t\chi_t' = O\Lambda O' = O_1\Lambda_1 O_1' \tag{3.20}$$

where $\Lambda$ is the diagonal matrix of the eigenvalues of $\gamma_0^\chi$ and $O$ the matrix consisting of the corresponding eigenvectors (columnwise). As $\gamma_0^\chi$ has finite rank, $r$ say, which is smaller than $N$ (if $N$ is big enough), $r \leq n < N$, $\Lambda_1$ is the $r \times r$ dimensional diagonal matrix consisting only of the non-zero eigenvalues of $\gamma_0^\chi$. $O_1$ is defined in an evident way. Note that a minimal $[F, G, H]$ system does not require a matrix $H$ of full column rank and although $\mathbb{E}x_t x_t'$ has full rank $n$, $\mathrm{rk}(\mathbb{E}\chi_t\chi_t') = \mathrm{rk}(H) = r \leq n$ holds. Thus there exists a regular matrix $T$ such that $HT = [H_1, 0]$ holds, where $H_1$ is of full column rank $r$.

$$\mathbb{E}\chi_t\chi_t' = H\mathbb{E}x_t x_t' H' = H_1\mathbb{E}f_t f_t' H_1' \tag{3.21}$$

where $f_t$ is a so called *minimal static factor* defined by

$$f_t = [H_1'H_1]^{-1}H_1'\chi_t \tag{3.22}$$

Furthermore two equivalent minimal state space systems $[F, G, H]$ and $[\tilde{F}, \tilde{G}, \tilde{H}]$ corresponding to a rational function $w(z)$ are related by a regular matrix $T$ such that

$$\tilde{H} = HT, \tilde{F} = T^{-1}FT, \tilde{G} = T^{-1}G$$

Note that the matrix $H$ of the canonical state space model described in the previous section has already this $[H_1, 0]$ structure. This can be seen as

$$\chi_t = [w_0, w_1, w_2, ...]\varepsilon_t^- \tag{3.23}$$

and (3.21) imply that $[w_0, w_1, w_2, ...]$ has rank $r$, which implies that the first $r$ rows of $\mathscr{H}_\alpha$ are sufficient to describe $[w_0, w_1, w_2, ...]$. Thus, as $[w_0, w_1, w_2, ...] = H\mathscr{H}_\alpha$, $H$ must have this structure. An immediate consequence of this is, that $f_t$ in (3.21) is the process containing the first $r$ components of $x_t$. Also note that a state is always a static factor, but it is a minimal one if and only if all Kronecker Indices are either one or zero (that means that the first $n$ linearly independent rows of $\mathscr{H}$ are in the first block of rows).

From the previous discussion we know $\chi_t = w(z)\varepsilon_t = H_1 f_t$, where $H_1$ is of full column rank. Thus $f_t = k(z)\varepsilon_t$ holds, where $k(z) = [H_1'H_1]^{-1}H_1'w(z)$ is an $r \times q$ rational transfer functions for the static factors $f_t$.

**Theorem 3.2.5.** *Let $\chi_t = w(z)\varepsilon_t = H_1 k(z)\varepsilon_t$ where $w(z)$ and $k(z)$ are rational matrices of dimensions $(N \times q)$ and $(r \times q)$ respectively and $H_1 \in \mathbb{R}^{N\times r}$ of full column rank, then $w(z)$ is zeroless if and only if $k(z)$ is zeroless.*

*Proof.* ***(Theorem 3.2.5.)***
If $z_0$ is a zero of $k(z)$, i.e. $k(z_0)x = 0$ for an $x \in \mathbb{R}^q$ of course $w(z_0)x = 0$ holds too. If $z_0$ is a zero of $w(z)$, i.e. $w(z_0)x = 0$ for an $x \in \mathbb{R}^q$, $k(z_0)x = 0$ must hold too as $H_1$ is of full column rank. □

But not only the zeros of the transfer functions of $w$ and $k$ are the same:

**Theorem 3.2.6.** *Let $\chi_t$ be an $N$ dimensional process fulfilling Assumptions 1, 2 and 4 with an $[F, G, H]$ state space representation of order $n$ of a spectral factor $w$ of $f_\chi$ obtained by Theorem 3.1.1. Let $\mathrm{rk}(H) = r$ and w.l.o.g let $H = [H_1, 0]$, with $H_1 \in \mathbb{R}^{N\times r}$ of full column rank, such that $\chi_t = w(z)\varepsilon_t = H_1 k(z)\varepsilon_t$ holds, where $w(z) = [H_1, 0][I - Fz]^{-1}G$ and $k(z) = [H_1'H_1]^{-1}H_1'[H_1, 0][I - Fz]^{-1}G$ are rational matrices of dimensions $(N \times q)$ and $(r \times q)$ respectively. Then the minimal static factors $f_t = k(z)\varepsilon_t$ have a minimal $[F, G, C]$ state space realization of order $n$*

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{3.24}$$

$$f_t = Cx_t \tag{3.25}$$

*where $C = [I_r, 0]$, if and only if the $[F, G, H]$ state space representation of $\chi_t$ is minimal.*

*Proof.* ***(Theorem 3.2.6.)***
As the matrices $F$ and $G$ in the representation of $k(z)$ are the same as of $w(z)$ and $C = [H_1'H_1]^{-1}H_1'[H_1, 0]$ we just have to show the observability statement. If the matrix $\begin{bmatrix} C \\ CF \\ \vdots \\ CF^{n-1} \end{bmatrix}$ is of full column rank, then

the matrix $\begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} = \begin{bmatrix} H_1 & 0 & \dots & 0 \\ 0 & H_1 & \ddots & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & H_1 \end{bmatrix} \begin{bmatrix} C \\ CF \\ \vdots \\ CF^{n-1} \end{bmatrix}$ is of full column rank. And if the matrix $\begin{bmatrix} C \\ CF \\ \vdots \\ CF^{n-1} \end{bmatrix}$ is not of full column rank the same holds obviously for $\begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix}$. □

As the matrices $w$ and $k$ have the same properties (dynamics) and $k$ is of smaller dimension, of course we want to model $k$ instead of $w$.

## 3.3 Zeroless Transfer Functions

Now we want to discuss the very important and interesting property of zerolessness. We will see in the next section that a rational function with more rows than columns is generically[1] zeroless. The results of this section allows us to emphasize the case where the minimal static factors have an autoregressive representation, although the matrix $H$ has not full column rank. This yields a generalization of the model of [Stock and Watson, 2002a] who have only discussed the full rank case.
Note, that we have always worked with rational functions $k$ which have no zeros and poles within and on the unit circle. We know, that such a transfer function has always a causal left inverse (they can be used for a Wold representation, compare to (3.2)) which can be seen by using the Smith-Mc Millan form

$$k = ulv \tag{3.26}$$

and defining a left inverse

$$k^- = v^{-1}[l'l]^{-1}l'u^{-1} \tag{3.27}$$

which is causal as the matrix $l$ contains all zeros and poles of $k$ in its numerator and denominator polynomials such that $[l'l]^{-1}l'$ has again all its zeros and poles outside the unit circle. Obviously the zeros of $k$ become the poles of $k^-$ and the same holds for the poles of $k$ and the zeros of $k^-$. This implies that if $k$ has no zeros $k^-$ has no poles and is therefore a polynomial matrix instead of a rational one.
Thus we have the following theorem (see for instance [Deistler et al., 2010a])

**Theorem 3.3.1.** *Let $\chi_t$ satisfy Assumptions 1, 2 and 4 and let $f_t$ be an associated minimal static factor,*

[1] Genericity is a property of a set. Thus saying that a function is generically zeroless is very sloppy. A detailed discussion can be found in the next section.

*of dimension $r$; then the following statements for*

$$f_t = k(z)\varepsilon_t \tag{3.28}$$

*are equivalent:*

*(i) The spectral factors $k$ of the spectral density $f_f$ (of $f_t$) satisfying the properties listed in Theorem 3.1.1, are zeroless*

*(ii) There exists a polynomial left inverse $k^-$ corresponding to* (3.27) *and thus the input $\varepsilon_t$ in* (3.28) *is determined from a finite number of outputs $f_t, f_{t-1}, \ldots, f_{t-L}$, for some $L$*

*(iii) $f_t$ is a stationary solution of a stable AR system*

$$f_t = a_1 f_{t-1} + \cdots + a_p f_{t-p} + \nu_t;\ a_i \in \mathbb{R}^{r\times r} \tag{3.29}$$

*where*

$$\det \underbrace{[I - a_1 z - \cdots - a_p z^p]}_{a(z)} \neq 0,\ |z| \leq 1$$

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

*Proof.* ***(Theorem 3.3.1.)***
In order to show (i) $\Rightarrow$ (iii), we commence from an ARMA representation for $f_t$

$$\tilde{a}(z) f_t = b(z)\varepsilon_t \tag{3.30}$$

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

$$[b, g] = \tilde{u} \begin{bmatrix} \tilde{d}, \begin{bmatrix} 0 \\ I \end{bmatrix} \end{bmatrix} \begin{bmatrix} \tilde{v} & 0 \\ 0 & I \end{bmatrix}$$

Then

$$\tilde{a}(z) f_t = u(z) \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

and

$$u^{-1}(z)\tilde{a}(z) f_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \tag{3.31}$$

gives an autoregressive representation, and pre-multiplying (3.31) by $\tilde{a}^{-1}(0)u(0)$ gives the desired form (3.29). The stability of $a(z)$ follows from the stability of $\tilde{a}(z)$.

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

$$\Sigma_\nu = PP', P \in \mathbb{R}^{r\times q}, \operatorname{rk} P = q \tag{3.32}$$

Then pre-multiplying (3.29) by $[P'P]^{-1}P'$ yields a $k^-$ of the desired form. □

In the previous section we have established the relation $N \geq n \geq r \geq q$. That means that the $N$ dimensional vector of latent variables $\chi_t$ is driven by the dynamic factor process of far smaller dimension $q$. Furthermore the dynamics of the system can be described by an AR(1) system for the $n$ dimensional state (where $n$ should also be, and actually will be, far smaller than $N$ for $N$ big enough).

The reason why we are using the $[F, G, H]$ state space representation is, $x_t$ is already a static factor, although not necessarily a minimal one. The logical way to proceed is to get the state from the latent variables by a linear transformation (a left inverse of $H$) and to use the AR(1) system of the state to get a forecast of the state and therefore a forecast of the latent variables. There is only one problem which might arise, namely that the matrix $H$ does not need to have full column rank (that is not needed for minimality) such that the state cannot be reconstructed by the latent variables.

Nevertheless we can reconstruct a minimal static factor from the latent variables, and if the transfer function $w$ of the latent variables is zeroless, we have shown with Theorem 3.2.5 and Theorem 3.3.1 that the minimal static factor $f_t$ defined by (3.22) has an autoregressive representation (3.29). Thus we result in a system for the latent variables:

$$f_t = a_1 f_{t-1} + \cdots + a_p f_{t-p} + \nu_t \tag{3.33}$$

$$\chi_t = H_1 f_t \tag{3.34}$$

with $a_i \in \mathbb{R}^{r\times r}$, $H_1 \in \mathbb{R}^{N\times r}$ and $\nu_t = b\varepsilon_t$ with $b \in \mathbb{R}^{r\times q}$.

In the next section we will see that the assumption of a zeroless transfer function $w$ and thus of $k$ is reasonable.

## 3.4 A Generic Set

Before we present the main Theorem 3.4.3 we need the two following lemmas.

Define

$$M(z) = I - Fz$$

with $F \in \mathbb{R}^{n\times n}$.

**Lemma 3.4.1.** *The set* $\mathscr{F} = \{F | \operatorname{rk}(M(z)) \geq n-1, \forall z \in \mathbb{C}\}$ *is open and dense in* $\mathbb{R}^{n\times n}$

*Proof. (Lemma 3.4.1.)*
dense: Let $F_0$ be an $n \times n$ real matrix with $\exists z_0$ such that $\operatorname{rk}(M(z_0)) = n-2$. (Note only the case where the rank drops to $n-2$ is analyzed as the generalization to $n-j, j \geq 2$ is straightforward.) Therefore there exists a 2-dimensional kernel of $M(z_0)$. Let $[x_1, x_2]$ be the span of two linearly independent vectors $x_1, x_2$ of this kernel. We can always find two linearly independent vectors of this kernel which have a permutation of a $2 \times 2$ identity matrix somewhere in their rows. Consequently we can manipulate one entry of $F_0$ such that one of the vectors is not in the kernel of the new $M(z_0)$ anymore, but the other one still is. If the perturbation, $\varepsilon_0$ say, is small enough, the other zeros of the new $M(z)$ will be in a certain neighborhood of the old zeros. Therefore there is a sequence of matrices $M(z)$ (if $\varepsilon_0 \to 0$) which have ranks $n-1$ whose corresponding $F$s converge to $F_0$.
open: $\mathscr{F}$ is open is equivalent to $\mathscr{F}^C$ (= the complement of $\mathscr{F}$) is closed. Assume that $\mathscr{F}^C$ is not closed, i.e. there exists a sequence $(F_m)_{m\in\mathbb{N}}$ with $F_m \to F_0$ where $F_m \in \mathscr{F}^C$ and $F_0 \in \mathscr{F}$. $F_m \in \mathscr{F}^C$ means that there exists a $0 \neq z_m \in \mathbb{C}$ with $\operatorname{rk}(M(z_m)) \leq n-2$. Consequently $\operatorname{rk}(\underbrace{Iz_m^{-1} - F_m}_{=:B_m}) \leq n-2 \Leftrightarrow$ $\lambda_{n-1}(B_m) = \lambda_n B_m = 0$ where $\lambda_i(A)$ denotes the i-th largest eigenvalue of $A$. As $B_m \to B_0 := Iz_0^{-1} - F_0$ holds as $F_m \to F_0$ holds and $z_m^{-1}$ is an eigenvalue of $F_m$ and consequently $z_m^{-1} \to z_0^{-1}$ holds, $\lambda_{n-1}(B_m) \to \lambda_{n-1}(B_0)$ must hold. But as $\lambda_{n-1}(B_0) > 0$ holds by assumption, we have a contradiction to $\lambda_{n-1}(B_m) \to 0$ and the result follows. □

Define

$$N(z) = \begin{pmatrix} I - Fz & -G \\ C & 0 \end{pmatrix} = \begin{pmatrix} I & -G \\ C & 0 \end{pmatrix} - \begin{pmatrix} F & 0 \\ 0 & 0 \end{pmatrix} z = \begin{pmatrix} I_r & 0 & -G_1 \\ 0 & I_{n-r} & -G_2 \\ C_1 & 0 & 0 \end{pmatrix} - \begin{pmatrix} F_{11} & F_{12} & 0 \\ F_{21} & F_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} z$$

with $F \in \mathbb{R}^{n\times n}, G \in \mathbb{R}^{n\times q}, C \in \mathbb{R}^{r\times n}, C_1 \in \mathbb{R}^{r\times r}, G_1 \in \mathbb{R}^{r\times q}, G_2 \in \mathbb{R}^{(n-r)\times q}$, with $q < r \leq n$. Moreover let

$$N_2(z) := \begin{pmatrix} 0 & -G_{11} \\ 0 & -G_{12} \\ I_{n-r} & -G_2 \end{pmatrix} - \begin{pmatrix} F_{12,1} & 0 \\ F_{12,2} & 0 \\ F_{22} & 0 \end{pmatrix} z$$

with $F_{12,1} \in \mathbb{R}^{(r-q)\times(n-r)}, F_{12,2} \in \mathbb{R}^{q\times(n-r)}, F_{22} \in \mathbb{R}^{(n-r)\times(n-r)}, G_{11} \in \mathbb{R}^{(r-q)\times q}, G_{12} \in \mathbb{R}^{q\times q}$ and $G_2 \in \mathbb{R}^{(n-r)\times q}$, and

$$N_1(z) := \begin{pmatrix} 0 & -G_{12} \\ I_{n-r} & -G_2 \end{pmatrix} - \begin{pmatrix} F_{12,2} & 0 \\ F_{22} & 0 \end{pmatrix} z$$

with $G_{12} \in \mathbb{R}^{q\times q}$ and $F_{12,2} \in \mathbb{R}^{q\times(n-r)}$.

**Lemma 3.4.2.** *Let $F_{12,2} \in \mathbb{R}^{q\times(n-r)}, F_{22} \in \mathbb{R}^{(n-r)\times(n-r)}, G_{12} \in \mathbb{R}^{q\times q}$ and $G_2 \in \mathbb{R}^{(n-r)\times q}$, then the set $\mathscr{N}_1 = \{[F_{12,2}, F_{22}, G_{12}, G_2] | \operatorname{rk}(N_1(z)) \geq n-r+q-1, \operatorname{rk}(G_{12}) = q\}$ is open and dense in the set $\{[F_{12,2}, F_{22}, G_{12}, G_2] | \operatorname{rk}(G_{12}) = q\}$.*

*Proof. (Lemma 3.4.2)*
dense: Define $T = \begin{pmatrix} 0 & -G_{12} \\ I_{n-r} & -G_2 \end{pmatrix}$ and let $z_k$ be a zero of $N_1(z)$ such that $\operatorname{rk} N_1(z_k) = n-r+q-2$ holds. Furthermore let $x_1, x_2$ be two linearly independent vectors which span the corresponding kernel. Therefore $N_1(z_k)x_i = 0, i = 1,2$ holds. Let $x_i = [x_{i1}', x_{i2}']'$ with $x_{i1} \in \mathbb{R}^{n-r}$, then $x_{11}, x_{21}$ must be linearly independent because otherwise there would exist scalars $a$ and $b$ such that $ax_1 - bx_2 = [0, ax_{12}' - bx_{22}']'$ with $G_{12}[ax_{12} - bx_{22}] = 0 \Rightarrow ax_{12} = bx_{22} \Rightarrow ay_1 - by_2 = 0$. Consequently we can add to a certain entry of $F_{22}$ a small $\varepsilon$ such that the kernel of the new $N_1(z_k)$ is only one dimensional (see Lemma 3.4.1)
open: It is equivalent to show that the complement, $\mathscr{N}_1^C$ say, is closed. Assume that the complement is not closed, i.e. there exists a sequence $[F_{12,2,m}, F_{22,m}, G_{12,m}, G_{2,m}]_{m\in\mathbb{N}} \to [F_{12,2,0}, F_{22,0}, G_{12,0}, G_{2,0}]$ where $[F_{12,2,m}, F_{22,m}, G_{12,m}, G_{2,m}]$ is in the complement of $\mathscr{N}_1$ for all $m$ and $[F_{12,2,0}, F_{22,0}, G_{12,0}, G_{2,0}] \in \mathscr{N}_1$. $[F_{12,2,m}, F_{22,m}, G_{12,m}, G_{2,m}] \in \mathscr{N}_1^C \Leftrightarrow \exists z_m \in \mathbb{C} : \operatorname{rk}(Iz_m^{-1} - \underbrace{T_m^{-1}\begin{pmatrix} F_{12,2,m} & 0 \\ F_{22,m} & 0 \end{pmatrix}}_{C_m}) \leq n+q-2$. As $C_m \to C_0 := T^{-1}\begin{pmatrix} F_{12,2,0} & 0 \\ F_{22,0} & 0 \end{pmatrix}$ holds, $\lambda_i(C_m) \to \lambda_i(C_0)$ holds too. Furthermore $B_m = [Iz_m^{-1} - C_m] \to [Iz_0^{-1} - C_0] =: B_0$ holds as $z_m \to z_0$, it follows that $\lambda_{n-1}(B_m) \to \lambda_{n-1}(B_0) = 0$ which is a contradiction to $\lambda_{n-1}(B_0) > 0$ and the result follows. □

**Theorem 3.4.3.** *Let $k(z)$ be an $r \times q$ dimensional rational matrix function with a minimal $[F, G, C]$ state space realization of dimension $n$, such that $q < r \leq n$ holds, then the set $\mathscr{N}_0 = \{[F, G, C] | \operatorname{rk}(C) = r, \lambda_1(F) < 1, \operatorname{rk}(G) = q, N_1(z) \in \mathscr{N}_1, \operatorname{rk}(C[I - Fz]^{-1}G) = q \forall z \in \mathbb{C}\}$ is generic (i.e. open and dense) in $\mathscr{N} = \{[F, G, C] | \operatorname{rk}(C) = r, \lambda_1(F) < 1, N_1(z) \in \mathscr{N}_1, \operatorname{rk}(G) = q\}$.*

Note that the restrictions $\lambda_1(F) < 1$ and $\operatorname{rk}(G) = q$ are mandatory for the transfer functions we are interested in. Furthermore Lemma 3.4.2 shows that the restriction of $N_1(z) \in \mathscr{N}_1$ is negliable.

*Proof.* ***(Theorem 3.4.3.)***
Note: $C$ must have already the special form $C = [C_1, 0]$ which can be achieved without any loss of generality.
A zero of $N(z)$ means $\exists z_0 \in \mathbb{C}, x = [x_1', x_2', x_3']', x_1 \in \mathbb{R}^r, x_2 \in \mathbb{R}^{n-r}, x_3 = \mathbb{R}^q$ such that $N(z_0)x = 0$ holds. Consequently $x_1 = 0$, and $N(z_0)x = 0 \Leftrightarrow \begin{pmatrix} F_{12}z_0 & -G_1 \\ I_{n-r} - F_{22}z_0 & -G_2 \end{pmatrix}\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = 0 = N_2(z_0)\begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$.
Note that $\operatorname{rk}(G) = q$ implies that there must exist a square sub matrix of $G$ which has full rank. Without loss of generality assume that this sub matrix is $G_{12}$.
dense: Using Lemma 3.4.2 the set of matrices corresponding to the square sub matrix $N_1(z)$ of $N_2(z)$

which have only zeros with one dimensional kernels is dense. It is clear that the set of matrices $[F_{12,1}z, -G_{11}]$ which is not orthogonal to any of the finite one dimensional kernels, is dense.
open: Assume that ${\mathscr{N}_0}^C$ is not closed. Then there exists a sequence of matrices $[F_{12,1}, F_{12,2}, F_{22}, G_{11}, G_{12}, G_2]_m \in {\mathscr{N}_0}^C$ which converges to $[F_{12,1}, F_{12,2}, F_{22}, G_{11}, G_{12}, G_2]_0 \in \mathscr{N}_0$ with: $\forall m \exists z_m$ and $x_m$ with $N_1(z_m)x_m = 0$ and $[F_{12,1}z_m, -G_{11}]_m x_m = 0$. $N_1(z_m)x_m = 0$ is possible $\forall m$ and occurs if and only if $z_m$ is an eigenvalue and $x_m$ the corresponding eigenvector (note that $N_1(z) \in \mathscr{N}_1$ implies that all eigenvalues are different) of $T^{-1}\begin{pmatrix} F_{12,2,m} & 0 \\ F_{22,m} & 0 \end{pmatrix}$ (compare to Lemma 3.4.2). But $\underbrace{[F_{12,1}z_m, -G_{11}]_m}_{a_m} x_m = 0 =< a_m, x_m >$ implies $lim_{m\to\infty} < a_m, x_m >=< a_0, x_0 >= 0$ which is a contradiction to $[F_{12,1}, F_{12,2}, F_{22}, G_{11}, G_{12}, G_2]_0 \in \mathscr{N}_0$. □

**Remark 3.4.4.** *Theorem 3.4.3 uses the relation $n \geq r > q$. If $n < r$ the result follows immediately as the matrices $C$ and $G$ are generically of full column rank.*

**Remark 3.4.5.** *The relation $r > q$ is crucial as otherwise the "tall" argument does not exist anymore and we have generically zeros.*

Tall transfer functions have a lot of interesting properties. More discussions can be found for example in [Anderson and Deistler, 2008a], [Anderson and Deistler, 2008b] and [Anderson and Deistler, 2009].

# Chapter 4

# Singular AR Systems

Autoregressive systems might be the most important models in time series analysis. Normally an AR system has a driving white noise (input) process $\varepsilon_t$, which has the same dimension as the output process and a regular covariance matrix $\Sigma = \mathbb{E}\varepsilon_t\varepsilon_t'$. In our structure theory we are facing the fact that the minimal static factors $f_t$ are described by

$$f_t = a_1 f_{t-1} + \cdots + a_p f_{t-p} + \nu_t \tag{4.1}$$

$a_i \in \mathbb{R}^{r\times r}$, where the driving white noise process $\nu_t$, of dimension $r$ has a covariance matrix which is singular $\Sigma_\nu = \mathbb{E}\nu_t\nu_t' = b\Sigma b'$ with $b \in \mathbb{R}^{r\times q}, q < r$, such that $\nu_t$ can be described by $\nu_t = b\varepsilon_t$ where $\varepsilon_t$ is a white noise process of dimension $q$ and has a regular covariance matrix $\mathbb{E}\varepsilon_t\varepsilon_t' = \Sigma$. As the regularity of the covariance matrix of the errors $\nu_t$ is of central importance for the existing theory of AR systems many properties change in the singular case. Here will present some of the most important facts and theorems of [Inouye, 1983], [Filler et al., 2009], [Deistler et al., 2010a] and our most recent paper [Chen et al., 2010].

If we are talking about autoregressive models of the form (4.1) we always want the error $\nu_t$ to be the one-step-ahead prediction error of $f_t$. That means that $\nu_t$ is orthogonal on all past values of $f_t$, i.e. $\mathbb{E}\nu_t f_s' = 0, s < t$. Furthermore we are interested in stationary solutions $f_t$ of (4.1) only and we know that any solution $f_t$ of (4.1) is a sum of a particular solution $f_t^p$ and a homogeneous solution $f_t^h$, i.e. $a(z)f_t^h = 0$, such that

$$f_t = f_t^p + f_t^h \tag{4.2}$$

$$a(z)f_t^p = \nu_t = b\varepsilon_t \tag{4.3}$$

$$a(z)f_t^h = 0 \tag{4.4}$$

As $f_t$ has to be stationary we know that $f_t^h \not\equiv 0$ if and only if $f_t^h$ belongs to a zero $z_0$ of unit modulus of the polynomial matrix $a(z)$, i.e. $a(z_0)f_t^h = 0, |z_0| = 1$. From Wold's Decomposition we know that $f_t$ can also be represented as

$$f_t = f_t^r + f_t^s \tag{4.5}$$

where $f_t^r$ is a linearly regular process that admits a causal representation $f_t^r = \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}$ and $f_t^s$ a linearly singular process with $\mathbb{E} f_t^r f_t^{s'} = 0$. We will see below that $f_t^r = a(z)^{-1} b \varepsilon_t$ is always a causal linearly regular solution of (4.1). Note that we could also include degenerated systems such that $\Sigma_\nu = 0$ holds which would imply $b = 0$ ("the" particular solution is equal to 0). Thus the linearly singular component is a homogeneous solution of (4.4) and is therefore a harmonic process.

## 4.1 Yule-Walker Equations

Let us commence from a stationary process $f_t$ which is a solution of an AR(p) system (4.1). Here we are interested in the question whether the coefficients $a_1, ..., a_p$ of the AR(p) system are unique and if not if we can find a canonical representative.

As $f_t$ is stationary and has an AR(p) representation the Yule-Walker equations

$$[a_1, ..., a_p]\Gamma_p = [\gamma_1, ..., \gamma_p] \tag{4.6}$$

$$\Sigma_\nu = \gamma_0 - [a_1, ..., a_p]\Gamma_p[a_1, ..., a_p]' \tag{4.7}$$

hold, where $\gamma_j = \mathbb{E} f_t f_{t-j}$ and $\Gamma_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1' & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{p-1}' & \cdots & \gamma_1' & \gamma_0 \end{bmatrix}$. Contrary to the regular case the matrix $\Gamma_p$ does not need to be regular such that the solution of (4.6) does not need to be unique. Also note, that $\Gamma_{p+1}$ is always singular as $b^\perp[I, -a_1, ..., -a_p]$ with $b^\perp b = 0$, is always in the (left) kernel of $\Gamma_{p+1}$. However there always exists a solution of (4.6) as the (right) kernels of $\Gamma_p$ and $[\gamma_1, ..., \gamma_p]$ are the same. Furthermore $\Sigma_\nu$ in (4.7) is always unique as the difference of two solutions of (4.6) is in the (left) kernel of $\Gamma_p$. Nevertheless the next lemma shows that if the end matrix $a_p$ of $a(z)$ has generic entries[1] the coefficient matrices $[a_1, ..., a_p]$ are unique, which is equivalent to the condition that $\Gamma_p$ be regular.

**Definition 4.1.1. (observationally equivalent)** *Two ARMA systems $[a(z), b(z)]$ and $[\bar{a}(z), \bar{b}(z)]$ with $a_i, \bar{a}_i \in \mathbb{R}^{r\times r}$ and $b_i, \bar{b}_i \in \mathbb{R}^{r\times q}, q \leq r$, are called **observationally equivalent** if $\forall \Sigma > 0$ the spectral densities $a^{-1}b\Sigma b^* a^{-1*}$ and $\bar{a}^{-1}\bar{b}\Sigma\bar{b}^*\bar{a}^{-1*}$ are the same.*

[1] The wording: "If the matrix has generic entries" is a bit clumsy, but it means that the matrix fulfills the property on a generic set.

**Lemma 4.1.2.** *Let $a(z) = I - a_1 z - ... - a_p z^p$ and $[a(z), b]$ be left coprime, then $a(z)$ is unique among all observationally equivalent matrix polynomials $a(z)$ of maximal order $p$, with $a(0) = I$ if and only if $[a_p, b]$ has full rank.*

*Proof.* ***(Lemma 4.1.2.)*** See also [Zinner, 2008].
As $[a(z), b]$ is left coprime all observationally equivalent AR representations $[\bar{a}(z), \bar{b}]$, with $\bar{a}(0) = I$, fulfill $[\bar{a}(z), \bar{b}] = u(z)[a(z), b]$ with $\bar{u}(0) = I$ and thus $[\bar{a}(z), \bar{b}] = [\bar{a}(z), b]$.If $[a_p, b]$ has not full rank then there exists a matrix $u_1 \neq 0$ with $u_1[a_p, b] = 0$. Thus $\bar{a}(z) := [I + u_1 z]a(z)$ is of order $p$ and thus $a(z)$ is not unique.
Conversely if $a(z)$ is not unique then there exists a $u(z) = I - u_1 z - ... - u_m z^m$ of order $m$ with $\bar{a}(z) = u(z)a(z)$ where the order of $\bar{a}(z)$ is smaller or equal to $p$. Therefore $u_m a_p = 0$ holds, together with $u_m b = 0$ this completes the proof. □

So we know that the Yule Walker equations have a unique solution if $\Gamma_p$ is regular and this result is generic if the end matrix $a_p$ has generic entries. Of course the result even holds if sufficiently many columns (if the matrix $b$ has generic entries) of $a_p$ are generic.

Next we are concerned with a singular $\Gamma_p$. It is obvious that if $[a_1, ..., a_p]$ is a solution of (4.7) then $[\bar{a}_1, ..., \bar{a}_p] = [a_1, ..., a_p] + [l_1, ..., l_p]$ is also a solution for all matrices $[l_1, ..., l_p]$ which are in the left kernel of $\Gamma_p$. Nevertheless we can always choose a canonical representative by taking the row wise minimum norm ($l_2$ norm) solution. It is well known that we get the minimum norm solution by using the pseudo-inverse of $\Gamma_p$. Let

$$\Gamma_p = \begin{bmatrix} O_1, O_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} O_1' \\ O_2' \end{bmatrix} \tag{4.8}$$

be the eigenvalue decomposition where the diagonal matrix $\Lambda_1$ consists of the non zero eigenvalues of $\Gamma_p$ and $O_1$ is the matrix consisting of the corresponding (normalized) eigenvectors. Hence

$$[a_1, ..., a_p] = [\gamma_1, ..., \gamma_p]\Gamma_p^{\#} \tag{4.9}$$

is the minimum norm solution, where $\Gamma_p^{\#} = O_1 \Lambda_1^{-1} O_1'$.

### 4.1.1 Discrete Ljapunov Equation

Suppose $[a_1, ..., a_p]$ is an arbitrary solution of the Yule-Walker equations (4.6), (4.7), then the AR(p) system for the process $f_t$ can be tranformed into an AR(1) system for the stacked process $[f_t', f_{t-1}', \ldots, f_{t-p+1}']'$

$$\begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix} = A \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-p} \end{bmatrix} + \begin{bmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4.10}$$

where $A$ is the block companion matrix of the polynomial $a(z) = I - a_1 z - ... - a_p z^p$

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & & a_p \\ I & 0 & \ldots & & 0 \\ 0 & I & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & I & 0 \end{bmatrix} \tag{4.11}$$

As $f_t$ is stationary by assumption, the stacked process $[f_t', f_{t-1}', \ldots, f_{t-p+1}']'$ is stationary too. Therefore, the discrete Ljapunov equation

$$\begin{aligned} \Gamma_p &= \mathbb{E} \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix} [f_t', f_{t-1}', \ldots, f_{t-p+1}']' \\ &= A\, \mathbb{E} \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \vdots \\ f_{t-p} \end{bmatrix} [f_{t-1}', \ldots, f_{t-p}']'\, A' + \mathbb{E} \begin{bmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} [\nu_t', 0, \ldots, 0]' \\ &= A\Gamma_p A' + \underbrace{\begin{bmatrix} \Sigma_\nu & 0 & \ldots \\ 0 & 0 & \\ \vdots & & \ddots \end{bmatrix}}_{=:Q} = A\Gamma_p A' + Q = A\Gamma_p A' + BB' \end{aligned} \tag{4.12}$$

holds, where $B = [\Sigma_\nu^{1/2'}, 0, ..., 0]'$. It is well known that (4.12) for given $A$ and $Q$ does not necessarily have a unique solution $\Gamma_p$ (see Theorem 2.1, page 16 in [Tummeltshammer, 2009]). This fact leads us to

the following section.

## 4.2 (Unique) Solutions of Singular AR Systems

We know already that there are two difficulties in the theory of singular AR systems. First, the Yule-Walker equations do not need to have a unique solution. Second, given a solution of the Yule-Walker equations, the corresponding Ljapunov equation does not need to have a unique solution (for example if the autoregressive polynomial has a zero on the unit circle). Nevertheless even in the worst case some nice results are available.

The following lemma shows that a unique solution of (4.12) always exists, and that this solution is even unique for different solutions of the Yule-Walker equations.

**Lemma 4.2.1.** *Suppose equation* (4.12) *has a non negative solution* $\Gamma_p$ *for fixed* $A$ *and* $Q$. *Define* $\Gamma_p^{min} = \sum_{j=0}^{\infty} A^j Q A^{j'}$. *Then* $\Gamma_p^{min}$ *exists and satisfies* (4.12)*, and any nonnegative solution* $\Gamma_p$ *of* (4.12) *satisfies* $\Gamma_p^{min} \leq \Gamma_p$. *Furthermore suppose that* $A_1$ *and* $A_2$ *are the block companion matrices of two solutions of the Yule-Walker equations, then* $\Gamma_p^{min,1} = \sum_{j=0}^{\infty} A_1^j Q A_1^{j'} = \sum_{j=0}^{\infty} A_2^j Q A_2^{j'} = \Gamma_p^{min,2}$ *holds.*

*Proof.* ***(Lemma 4.2.1.)***
Define $\Gamma_{(i)} = \sum_{j=0}^{i} A^j Q A^{j'}$, such that $\Gamma_{(0)} = Q$ and $\Gamma_{(1)} = A\Gamma_{(0)}A' + Q$. As $\Gamma_p$ is a solution of (4.12) we have $\Gamma_p = A\Gamma_p A' + Q \geq Q = \Gamma_{(0)}$. For induction assume that $\Gamma_p \geq \Gamma_{(i)}$ holds. Consequently $\Gamma_p = A\Gamma_p A' + Q \geq A\Gamma_{(i)}A' + Q = \Gamma_{(i+1)}$. Therefore $\Gamma_p^{min} = \lim_{i\to\infty} \Gamma_{(i)}$ is well defined and obviously $\Gamma_p \geq \Gamma_p^{min}$ holds for each solution $\Gamma_p$ of (4.12).
Independency of the solution of the Yule-Walker equations:
It is trivial that $[A_1 - A_2]\Gamma_p$=0. It follows that $[A_1 - A_2]\Gamma_p[A_1' - A_2'] = 0$ and then, because $\Gamma_p \geq \Gamma_p^{min,1}, \Gamma_p^{min,2}$ there holds $[A_1 - A_2]\Gamma_p^{min,1}[A_1 - A_2]' = 0$ and $[A_1 - A_2]\Gamma_p^{min,2}[A_1 - A_2]' = 0$. Hence $[A_1 - A_2][\Gamma_p^{min,1}]^{1/2}$=0 or $A_1[\Gamma_p^{min,1}]^{1/2} = A_2[\Gamma_p^{min,1}]^{1/2}$; likewise, $A_1[\Gamma_p^{min,2}]^{1/2} = A_2[\Gamma_p^{min,2}]^{1/2}$.Now recall that

$$\Gamma_p^{min,1} - A_1\Gamma_p^{min,1}A_1' = Q \tag{4.13}$$

It follows that

$$\Gamma_p^{min,1} - A_2\Gamma_p^{min,1}A_2' = Q \tag{4.14}$$

and likewise $\Gamma_p^{min,2}$ satisfies both of these equations. Because $\Gamma_p^{min,1}$ is the minimum solution of the first equation, there holds $\Gamma_p^{min,1} \leq \Gamma_p^{min,2}$ and the reverse equality holds because $\Gamma_p^{min,2}$ is the minimum solution of the second equation. Hence $\Gamma_p^{min,1} = \Gamma_p^{min,2}$, as required. □

It will be shown that $\Gamma_p^{min}$ is of central importance as it is the covariance of the linearly regular part of the underlying stationary process $f_t$. This is at first glance surprising as no stability requirement on the

solution of the Yule-Walker equations is required, and therefore it is not clear whether the representation $f_t^r = a(z)^{-1}b\varepsilon_t$ is indeed a causal linear transformation of the $\varepsilon_t$. Fortunately it is causal, as the zeros which are not stable, i.e. not outside the unit circle, cancel out in $a(z)^{-1}b$.

**Theorem 4.2.2.** *Let $[a_1, ..., a_p]$ be a solution of the Yule-Walker equations* (4.6) *and let $\Sigma_\nu = b\Sigma b'$, then $a(z)^{-1}b$ is a causal linear filter and is unique for all solutions $[a_1, ..., a_p]$ of* (4.6).

*Proof.* ***(Theorem 4.2.2.)***
Causality:
If $a(z)$ is stable the causality follows immediately. Now assume that $a(z)$ is not stable. Let $z_0 = \lambda^{-1} \in \mathbb{C}, |\lambda| \geq 1$ be a zero of $a(z)$ inside or on the unit circle. It is well known that a zero $z_0$ of $a(z)$ corresponds to an eigenvalue $z_0^{-1} = \lambda$ of the corresponding block companion matrix $A$, i.e. there exists an $x = [x_1', ..., x_p']'$ such that $A'x = \lambda x$. This implies

$$\begin{bmatrix} a_1' & I & 0 & \\ a_2' & 0 & I & \\ \vdots & & & \\ a_{p-1}' & & & I \\ a_p' & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} a_1'x_1 + x_2 \\ \vdots \\ a_{p-1}'x_1 + x_p \\ a_p'x_1 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \tag{4.15}$$

Expressing in each block of equations the $x_j$ with the highest index, starting in the last block, and replacing $x_j$ in the other blocks of equations with this expression, gives $x_1'[I - a_1\lambda^{-1} - ... - a_p\lambda^{-p}] = 0$. Additionally from the Ljapunov equation (4.12) we see that

$$\bar{x}'\Gamma_p x - \bar{x}' A\Gamma_p A' x = \underbrace{\bar{x}'\Gamma_p x}_{\geq 0} \underbrace{(1 - |\lambda|^2)}_{\leq 0} = \underbrace{\bar{x}' Q x}_{\geq 0} \tag{4.16}$$

holds and therefore $\bar{x}'Qx = \bar{x}_1'\Sigma_\nu x_1 = 0$ follows which is equivalent to $b'x_1 = 0$. Summarizing we have $\bar{x}_1'[a(\lambda^{-1}), b] = 0$ which is equivalent to (compare to Lemma 9.1.3) $[a(z), b]$ is not left coprime and therefore we can eliminate all unpleasant zeros until we result in a stable coprime pair such that the causality is ensured.
Uniqueness: We know already that $f_t^r = a(z)^{-1}b\varepsilon_t$ is well defined.

$$F_t^r = \begin{bmatrix} f_t^r \\ f_{t-1}^r \\ \vdots \\ f_{t-p+1}^r \end{bmatrix} = \begin{bmatrix} a(z)^{-1}b\varepsilon_t \\ a(z)^{-1}b\varepsilon_{t-1} \\ \vdots \\ a(z)^{-1}b\varepsilon_{t-p+1} \end{bmatrix} = \begin{bmatrix} a(z)^{-1}b\varepsilon_t \\ za(z)^{-1}b\varepsilon_t \\ \vdots \\ z^{p-1}a(z)^{-1}b\varepsilon_t \end{bmatrix} = [I - Az]^{-1} \begin{bmatrix} b\varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4.17}$$

because

$$
\begin{aligned}
[I - Az]^{-1} &= \left[\begin{array}{c|cccc} I - a_1 z & -a_2 z & \dots & -a_{p-1} z & -a_p z \\ \hline -Iz & I & 0 & \dots & 0 \\ 0 & -Iz & I & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & -Iz & I \end{array}\right]^{-1} \\
&= \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} A^{11} & A^{12} \\ \hline A^{21} & A^{22} \end{array}\right] = \\
&= \left[\begin{array}{c|c} [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & A^{12} \\ \hline -A_{22}^{-1}A_{21}[A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & A^{22} \end{array}\right] = \left[\begin{array}{c|ccc} a(z)^{-1} & x & x & x \\ \hline z a(z)^{-1} & x & x & x \\ \vdots & \vdots & & \vdots \\ z^{p-1} a(z)^{-1} & x & \dots & x \end{array}\right]
\end{aligned}
\tag{4.18}
$$

where $A_{22}^{-1} = \begin{bmatrix} I & 0 & \dots & 0 \\ Iz & I & 0 & \dots \\ \vdots & & \ddots & \\ Iz^{p-2} & \dots & Iz & I \end{bmatrix}$ holds. Obviously $\mathbb{E}F_t^r F_t^{r'} = \Gamma_p^{min} = \sum_{j=0}^{\infty} A^j Q A^{j'}$ is well defined and is the same for different $A$. From (4.17) it follows that $a(z)^{-1}b$ must be the same for all solutions $a(z)$ of the Yule-Walker equations as otherwise $\Gamma_p^{min}$ cannot be unique. □

So we know that no matter which solution of the Yule-Walker equations we choose, we will always get the linearly regular part of the underlying stationary process and its covariance by taking the unique solution $\Gamma_p^{min}$ of the Ljapunov equation which we denote from now on as $\Gamma_p^r$. Now it is also clear what the difference between the given covariance $\Gamma_p$ and the covariance of the linearly regular part $\Gamma_p^r$ is. It is the covariance of the linearly singular part, which of course cannot be reconstructed by the transfer function $a(z)^{-1}b$.

**Theorem 4.2.3.** *Suppose an* $(rp \times rp)$ *matrix* $\Gamma_p^s$ *satisfies*

$$\Gamma_p^s = A\Gamma_p^s A'$$

*then the random process $F_t^s$ generated by*

$$F_{t+1}^s = AF_t^s, t = 0, 1, ...$$
$$F_{t-1}^s = \Gamma_p^s A' \Gamma_p^{s\#} F_t^s, t = 0, 1, ...$$

*with initial conditions $\mathbb{E}F_0^s = 0$ and $\mathbb{E}F_0^s F_0^{s'} = \Gamma_p^s$, where # denotes the pseudo inverse, is stationary.*

*Proof.* ***(Theorem 4.2.3.)***
See Lemma 6 on page 49 in [Inouye, 1983]. □

Of course the difference $\Gamma_p - \Gamma_p^r$ fulfills the requirement of Theorem 4.2.3 as both $\Gamma_p$ and $\Gamma_p^r$ are solutions of the Ljapunov equation.
Finally we can decompose the underlying stationary process $f_t$ in its linearly regular part $f_t^r = a(z)^{-1}b\varepsilon_t$ and its orthogonal linearly singular part $f_t - f_t^r$.

## 4.3 Properties of the Minimum Norm Solution

At the beginning of the previous section we stated that we commence from a stationary process $f_t$ which has an AR(p) representation (4.1). Furthermore we have already mentioned in the introduction of this chapter, that stationary solutions of an autoregressive system (4.1) may consist of a linearly regular and a linearly singular component which is orthogonal on the regular one. Moreover we know already that such a linearly singular component corresponds to a solution of the homogeneous solutions fulfilling (4.4). Now, the questions arise as to when the minimum norm solution

$$[a_1, ..., a_p] = [\gamma_1, ..., \gamma_p]\Gamma_p^{\#} \tag{4.19}$$

is stable and as to when a linearly singular component is present.

### 4.3.1 There is an underlying stable system

Let us assume that we know that the stationary process $f_t$ has a stable AR(p) representation. Consequently, there is at least one solution of the Yule-Walker equations which is stable. In the following, we will show that in this case the minimum norm solution always yields a stable autoregressive polynomial.

**Lemma 4.3.1.** *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.* ***(Lemma 4.3.1.)***
See Lemma 3 page 218 in [Deistler et al., 2010a]. □

**Theorem 4.3.2.** *Let $\Gamma_p$ be as in* (4.6)*, corresponding to an $r$ dimensional stationary process $f_t$ which has a stable AR(p) representation*

$$f_t = \bar{a}_1 f_{t-1} + ... + \bar{a}_p f_{t-p} + \nu_t \tag{4.20}$$

*and let $a_i, i = 1, \ldots, p$ denote the minimum norm solution of the Yule-Walker equation* (4.6) *defined by* (4.19)*. Then the system defined by the $a_i$s is stable, and there are $s$ (the rank of $\Gamma_p$) eigenvalues of $A$, defined in* (4.11)*, which are identical to the eigenvalues of $\bar{A}$, the block companion matrix associated with the $\bar{a}_i$s.*

*Proof.* ***(Theorem 4.3.2.)*** See also [Deistler et al., 2010a].
Let $O_p$ be the orthogonal matrix $[O_1, O_2]$ as in (4.8) such that

$$O_p' \Gamma_p O_p = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

with $\Lambda_1$ 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

$$[\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_p] = [\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p] O_p$$
$$[f_1, f_2, \ldots, f_p] = [a_1, a_2, \ldots, a_p] O_p$$

Since

$$\begin{aligned} [f_1, f_2, \ldots, f_p] O_p' \Gamma_p O_p &= [a_1, a_2, \ldots, a_p] \Gamma_p O_p \\ &= [\gamma_1, \gamma_2, \ldots, \gamma_p] O_p \\ &= [\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_p] \Gamma_p O_p \\ &= [\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_p] O_p' \Gamma_p O_p \end{aligned}$$

or equivalently

$$[f_1, f_2, \ldots, f_p] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} = [\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_p] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}$$

we see that the first $s$ columns of the block row matrix $[f_1, f_2, \ldots, f_p]$ are identical with those of

$[\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_p]$. Also, we can argue that the last $rp - s$ columns are zero:

$$\begin{aligned}[f_1, f_2, \ldots, f_p] &= [\gamma_1, \gamma_2, \ldots, \gamma_p]\Gamma_p^{\#} O_p \\ &= [\gamma_1, \gamma_2, \ldots, \gamma_p] O_p \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}\end{aligned}$$

Now consider the block companion matrices $\bar{A}, A$ defined by the $(\bar{a}_i)_{i=1,\ldots,p}$ and $(a_i)_{i=1,\ldots,p}$ respectively, together with their transforms

$$F = O_p' A O_p$$
$$\bar{F} = O_p' \bar{A} O_p$$

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

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

Now the matrix $\bar{F}$ satisfies the following transformed version of (4.12):

$$\begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} - \bar{F} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \bar{F}' = O_p' Q O_p \tag{4.21}$$

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

upper triangular. Now consider

$$\begin{aligned}
\bar{F} - F &= O_p'[\bar{A} - A]O_p \\
&= O_p' \begin{bmatrix} \bar{a}_1 - a_1 & \bar{a}_2 - a_2 & \dots & \bar{a}_{p-1} - a_{p-1} & \bar{a}_p - a_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} O_p \\
&= O_p' \begin{bmatrix} \bar{f}_1 - f_1 & \bar{f}_2 - f_2 & \dots & \bar{f}_{p-1} - f_{p-1} & \bar{f}_p - f_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}
\end{aligned}$$

We have shown above that the first $s$ columns of $[\bar{f}_1, \bar{f}_2, \dots, \bar{f}_p]$ are identical with those of $[f_1, f_2, \dots, f_p]$, 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 $\bar{A}$ and therefore $\bar{F}$ 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 $F$, and in particular the last $rp - s$ rows of these columns; this is because $F$ is block triangular, and it has to be proved yet that the lower triangular block has all eigenvalues of modulus smaller than 1. Observe that

$$\begin{aligned}
F = O_p'AO_p &= O_p' \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} O_p \\
&= O_p' \begin{bmatrix} f_1 & f_2 & \dots & f_{p-1} & f_p \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} + O_p' \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} O_p
\end{aligned}$$

Now our interest is in the last $rp-s$ columns, and the entries of the last $rp-s$ columns of $[f_1\, f_2, \dots, f_{p-1}, f_p]$ have all been shown to be zero. Since what is in the first $s$ columns is irrelevant, we can say that the last

$rp - s$ columns of $F$ are actually identical with the last $rp - s$ columns of the matrix

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

Thus

$$F_{22} = [0_{(rp-s)\times s}, I_{rp-s}] O_p' \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} O_p \begin{bmatrix} 0_{s\times(rp-s)} \\ I_{rp-s} \end{bmatrix}$$

By Lemma 4.3.1, we know that $|\lambda_1(F_{22})| < 1$. In summary,

$$F = \begin{bmatrix} \bar{F}_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix} \tag{4.22}$$

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

Therefore, we know that if the underlying process is the solution of a stable AR system, the minimum norm solution yields a stable AR polynomial.

### 4.3.2 There is not an underlying stable system

Now we want to analyze the situation where we do not have an underlying stable system, i.e. the Yule-Walker equations do not have a stable solution. That means we have a stationary process $f_t$ which is a solution of a system (4.1) which is not stable. That is the point where the difference $\Gamma_p - \Gamma_p^r$ is not zero. Note that even in the case where $\Gamma_p$ is regular the corresponding unique solution of the Yule-Walker equations is not stable.

Although there does not exist a stable solution of the Yule-Walker equations, the minimum norm solution still has some nice properties.

**Theorem 4.3.3.** *Let $f_t$ be a stationary process of* (4.1) *and suppose that its covariance matrix $\Gamma_p$ is singular. Let $[a_1, ..., a_p]$ be the corresponding minimum norm solution and $A$ its block companion matrix. If the AR system* (4.1) *is not stable, i.e. the process $f_t$ has a linearly singular part, then the characteristic polynomial of the minimum norm solution has the following properties*

*(i) It cannot have zeros inside the unit circle and must have zeros both outside and on the unit circle.*

*(ii) The number of unit circle zeros is equal to* $\mathrm{rk}(\Gamma_p) - \mathrm{rk}([B, AB, ..., A^{rp-1}B])$*, where* $B$ *is as in* (4.12)*.*

*(iii) It has the least number of unit circle zeros among all solutions of the Yule Walker equations.*

*(iv) It has the largest number of stable zeros among all solutions of the Yule Walker equations.*

*Proof.* ***(Theorem 4.3.3.)*** See also [Chen et al., 2010].
(i): Let $\Gamma_p = O_p \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} O_p'$, then as in (4.21) we have that $F = O_p' A O_p$ fulfills

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

and is therefore upper diagonal (as in (4.22))

$$F = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix}$$

Since $\Gamma_p$ is singular $F_{22}$ exists (all eigenvalues of $F_{22}$ are inside the unit circle, compare to the proof of Theorem 4.3.2) and therefore the minimum norm solution has at least one eigenvalue inside the unit circle. According to the definition of $F$ define

$$G = O_p' B = \begin{bmatrix} G_1 \\ 0 \end{bmatrix} \tag{4.23}$$

such that the number of rows of $G_1$ is equal to the number of rows of $F_{11}$. We know that the pair $[F_{11}, G_1]$ is not reachable as otherwise all eigenvalues of $F_{11}$ are of modulus smaller than one (compare to (3.18))which is a contradiction to the assumption that the minimum norm solution is not stable. As the pair $[F_{11}, G_1]$ is not reachable it is well known (see for instance [Kailath, 1980]) that there exists a regular matrix $T$ such that

$$TF_{11}T^{-1} = \bar{F} = \begin{bmatrix} \bar{F}_{11} & \bar{F}_{12} \\ 0 & \bar{F}_{22} \end{bmatrix}, \quad \bar{G} = TG_1 = \begin{bmatrix} \bar{G}_1 \\ 0 \end{bmatrix} \tag{4.24}$$

where the number of rows of $\bar{G}_1$ and $\bar{F}_{11}$ are the same, and the pair $[\bar{F}_{11}, \bar{G}_1]$ is reachable. By construction the reachable modes of $[A, B]$, which are the eigenvalues $\lambda$ of $A$ such that $[\lambda I - A, B]$ has full rank (Definition 9.2.4), are the same as of $\left[ \begin{bmatrix} \bar{F} & TF_{12} \\ 0 & F_{22} \end{bmatrix}, \begin{bmatrix} \bar{G} \\ 0 \end{bmatrix} \right]$. Now the eigenvalues of $\bar{F}_{11}$ are exactly the reachable modes of $F_{11}$ and therefore they must be of modulus smaller than one as

$$\Lambda_1 - F_{11}\Lambda_1 F_{11}' = G_1 G_1' \tag{4.25}$$

holds (compare (3.18)). Furthermore pre-multiplying $T$ and post-multiplying $T'$ on (4.25) we obtain

$$\bar{\Lambda}_1 - TF_{11}\bar{\Lambda}_1 F_{11}'T' = \bar{G}\bar{G}'$$

where $\bar{\Lambda}_1 = T\Lambda_1 T' = \begin{bmatrix} \bar{\Lambda}_{1,11} & \bar{\Lambda}_{1,12} \\ \bar{\Lambda}_{1,21} & \bar{\Lambda}_{1,22} \end{bmatrix}$, which implies

$$\bar{\Lambda}_{1,22} - \bar{F}_{22}\bar{\Lambda}_{1,22}\bar{F}_{22}' = 0$$

and therefore all eigenvalues of $\bar{F}_{22}$ lie on the unit circle. As the set of eigenvalues of $A$ is the union of the sets of eigenvalues of $F_{22}, \bar{F}_{22}$ and $\bar{F}_{11}$ we have shown (i).
(ii): We know already that the number of zeros of modulus one of the minimum norm solution equals the rank of $\bar{F}_{22}$, as all eigenvalues of $\bar{F}_{22}$ are of modulus one. The rank of $\bar{F}_{22}$ is equal to $\mathrm{rk}(\Lambda_1) - \mathrm{rk}([G_1, F_{11}G_1, ..., F_{11}^{rp-1}G_1])$. As $\mathrm{rk}(\Gamma_p) = \mathrm{rk}(\Lambda_1)$ and
$\mathrm{rk}([B, AB, ..., A^{rp-1}B]) = \mathrm{rk}([G_1, F_{11}G_1, ..., F_{11}^{rp-1}G_1])$ holds, (ii) is shown.
(iii) and (iv): Recall from proof of Theorem 4.3.2 that the first $s$ columns of $\tilde{F} = O_p\bar{A}O_p'$, where $\bar{A}$ is a block companion matrix corresponding to any solution of the Yule-Walker equations, equal the first $s$ columns of $F = O_pAO_p'$ such that

$$\tilde{F} = \begin{bmatrix} F_{11} & F_{12} + V_{12} \\ 0 & F_{22} + V_{22} \end{bmatrix} \tag{4.26}$$

holds. As $F_{22}$ has only eigenvalues of modulus smaller than one the result follows. □

**Remark 4.3.4.** *From* (4.26) *it is clear that all solutions of the Yule-Walker equations have a common set of stable and unit modulus zeros corresponding to the eigenvalues of the matrix $F_{11}$. The others, corresponding to $F_{22}$, vary. Of course this result also holds in the case where the minimum norm solution is stable.*

Let us give a simple example to show that the minimum norm solution does not need to be stable.

**Example 4.3.5.** *Given the covariances*

$$\gamma_0 = \begin{bmatrix} 1.1061728 & 0 \\ 0 & 1 \end{bmatrix}, \gamma_1 = \begin{bmatrix} 0.3160494 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \gamma_2 = \begin{bmatrix} 0.217284 & 0 \\ 0 & 1 \end{bmatrix}$$

*we see that*

$$\Gamma_2 = \begin{bmatrix} 1.106173 & 0 & 0.3160494 & 0 \\ 0 & 1 & 0 & 1 \\ 0.3160494 & 0 & 1.106173 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

*is singular. Computing the minimum norm solution gives*

$$[\hat{a}_1, \hat{a}_2] = \begin{bmatrix} 0.25 & 0 & 0.1250000 & 0 \\ 0 & 0.5 & 0 & 0.5 \end{bmatrix}$$

*which is obviously not stable. Of course this is a very trivial example as we actually started with a system*

$$f_t = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.5 \end{bmatrix} f_{t-1} + \begin{bmatrix} 0.1250000 & 0 \\ 0 & 0.5 \end{bmatrix} f_{t-2} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t$$

*where we could have modeled the regular and the singular component seperately.*

A far more complicated model would be the following

**Example 4.3.6.** *Given the covariances*

$$\Gamma_2 = \begin{bmatrix} 1.402890 & 1.089555 & 1.145263 & 1.147196 & 0.6925516 & 0.6395243 & 0.6479484 & 0.6489144 \\ 1.089555 & 2.029612 & 1.048229 & 1.051525 & 0.192671 & -0.8649423 & 0.1412201 & 0.1399458 \\ 1.145263 & 1.048229 & 1.079348 & 1.084143 & 0.3199951 & 0.2208047 & 0.2336144 & 0.2323454 \\ 1.147196 & 1.051525 & 1.084143 & 1.090505 & 0.3290454 & 0.2193521 & 0.2308703 & 0.2288655 \\ 0.6925516 & 0.192671 & 0.3199951 & 0.3290454 & 1.402890 & 1.089555 & 1.145263 & 1.147196 \\ 0.6395243 & -0.8649423 & 0.2208047 & 0.2193521 & 1.089555 & 2.029612 & 1.048229 & 1.051525 \\ 0.6479484 & 0.1412201 & 0.2336144 & 0.2308703 & 1.145263 & 1.048229 & 1.079348 & 1.084143 \\ 0.6489144 & 0.1399458 & 0.2323454 & 0.2288655 & 1.147196 & 1.051525 & 1.084143 & 1.090505 \end{bmatrix}$$

*and*

$$[\gamma_1, \gamma_2] = \begin{bmatrix} 0.6925516 & 0.6395243 & 0.6479484 & 0.6489144 & 0.1082618 & 0.1047109 & 0.1031836 & 0.1032337 \\ 0.192671 & -0.8649423 & 0.1412201 & 0.1399458 & 0.09678233 & 1.11157 & 0.1110588 & 0.1118255 \\ 0.3199951 & 0.2208047 & 0.2336144 & 0.2323454 & 0.1830316 & 0.1952257 & 0.1953497 & 0.1963567 \\ 0.3290454 & 0.2193521 & 0.2308703 & 0.2288655 & 0.1820901 & 0.2113735 & 0.2090648 & 0.2102938 \end{bmatrix}$$

*we get the minimum norm solution*

$$[\hat{a}_1, \hat{a}_2] = \begin{bmatrix} 0.5 & -0.1312222 & 0.1302779 & 0.1259442 & -0.1072777 & -0.1312222 & -0.3025917 & 0.3 \\ 0.25 & -1.117207 & 0.8736488 & 0.1185584 & -0.2669369 & -0.1172071 & 0.2816035 & 1.025054e-13 \\ 0.4 & -0.09908457 & 0.01895525 & -0.1198707 & -0.08287562 & -0.09908457 & 0.1595288 & 2.795523e-13 \\ 0.5 & -0.0721236 & -0.06364235 & -0.1642340 & -0.1193951 & -0.0721236 & 0.1486095 & 6.326255e-13 \end{bmatrix}$$

*whose corresponding block companion matrix has eigenvalues of modulus*

$$\{1, 4.823482 \cdot 10^{-1}, 4.617895 \cdot 10^{-1}, 4.617895 \cdot 10^{-1}, 2.106091 \cdot 10^{-1}, 2.106091 \cdot 10^{-1}, 1.124859 \cdot 10^{-1}, 5.050252 \cdot 10^{-12}\}$$

*and thus $\hat{\mathrm{a}}(z)$ is not stable. The system was originated by modeling the linearly regular part as*

$$f_t^r = \begin{bmatrix} 0.5 & -0.25 & 0.25 & 0.125 \\ 0.25 & 0.25 & 0.125 & -0.5 \\ 0.4 & 0.2 & 0.1 & -0.5 \\ 0.5 & 0.5 & -0.3 & -0.5 \end{bmatrix} f_{t-1}^r + \begin{bmatrix} -0.125 & -0.125 & -0.3 & 0.3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} f_{t-2}^r + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \varepsilon_t$$

*and the linearly singular part as*

$$f_t^s = \begin{bmatrix} 0 \\ e^{i\pi t}r + e^{-i\pi t}\bar{r} \\ 0 \\ 0 \end{bmatrix}$$

*where* $\mathbb{E}r = \mathbb{E}\bar{r} = 0$ *and* $\mathbb{E}r\bar{r} = 1/2$

An immediate consequence of Theorem 4.3.3 is the following corollary which shows a possibility to check whether the minimum norm solution is stable or not.

**Corollary 4.3.7.** *Let* $f_t$ *be a stationary solution of* (4.1) *and suppose its corresponding* $\Gamma_p$ *is singular. Then the minimum norm solution* $[a_1, ..., a_p]$ *is stable if and only if* $\mathrm{rk}(\Gamma_p) = \mathrm{rk}([B, AB, ..., A^{rp-1}B])$*, where* $A$ *is the block companion matrix and* $B$ *is defined by* (4.12)*.*

*Proof.* ***(Corollary 4.3.7.)*** See also [Chen et al., 2010].
The proof of Theorem 4.3.3 (ii) points out that the unreachable modes of $[A, B]$ correspond to the eigenvalues of $\bar{F}_{22}$, and that the rank of $\bar{F}_{22}$ is equal to $\mathrm{rk}(\Gamma_p) - \mathrm{rk}([B, AB, ..., A^{rp-1}B])$. □

The last theorem in this chapter shows, that in the case of multiple solutions of the Yule-Walker equations, no matter if there is an underlying stable system or not, there always exists a solution with at least one unstable zero.

**Theorem 4.3.8.** *Let* $f_t$ *be a stationary solution of* (4.1) *and suppose its corresponding* $\Gamma_p$ *is singular. Then there always exists a solution of the Yule-Walker equations which has at least one unstable zero.*

*Proof.* ***(Theorem 4.3.8.)*** See also [Chen et al., 2010].
From (4.26) we know how two different solutions of the Yule-Walker equations are linked together. Our interest is in $V_{22}$, as we know that this part is responsible for the zeros which are not fixed. Let $O_p$ be the matrix consisting of the eigenvectors of $\Gamma_p$, i.e. $\Gamma_p = O_p \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} O_p'$, and let $O_p = \begin{bmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{bmatrix}$ then it is easy to verify that $V_{22} = O_{21}'\Phi$ for some $\Phi$. $O_{21}$ is of course not zero as otherwise $\Gamma_p$ would have zero rows and columns (as $O_{12}$ would be zero too). Therefore the pair $[F_{22}, O_{21}']$ has at least one reachable mode and thus there always exists a $\Phi$ such that the matrix $F_{22} + O_{21}'\Phi$ has an eigenvalue outside the unit circle. □

## 4.4 Canonical Representatives

As the Yule-Walker equations do not have to have a unique solution one is interested in a canonical representative. Here we will present three possibilities. The first is the already intensively discussed minimum norm solution of the Yule-Walker equations. The second is achieved by using row polynomial-echelon forms of certain rows of the autoregressive polynomial. And the third is inducted from a state space

construction where the least number of past observations is used and therefore a set of columns in the AR polynomial are zero.

In this section we will deal with linearly regular processes exclusively. This means we observe a stationary process $f_t$ which has a stable autoregressive representation

$$f_t = a_1 f_{t-1} + \cdots + a_p f_{t-p} + \nu_t \tag{4.27}$$

### 4.4.1 Canonical Representative I

We have already discussed many properties of the minimum norm solution. Here we want to point out that the minimum norm solution is of course always unique and therefore yields a canonical representative in the class of all singular AR(p) systems. Nevertheless it has some drawbacks. First it is only unique for fixed $p$ (not surprisingly), i.e. one has to know the minimal lag of the underlying AR process from which one only knows the sequence of covariances.

**Example 4.4.1.** *Let*

$$f_t = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{bmatrix} f_{t-1} + \begin{bmatrix} 0.25 & 0 \\ 0 & 0 \end{bmatrix} f_{t-2} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t$$

*be a stable stationary singular AR(2) process. Solving the discrete Ljapunov equation with*

$$A = \begin{bmatrix} 0.5 & 0 & 0.25 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

*gives*

$$\Gamma_2 = \begin{bmatrix} 1.920000 & 1.105455 & 1.280000 & 0.930909 \\ 1.105455 & 1.376970 & 1.512727 & 1.241212 \\ 1.280000 & 1.512727 & 1.920000 & 1.105455 \\ 0.930909 & 1.241212 & 1.105455 & 1.376970 \end{bmatrix}$$

*and*

$$[\gamma_1, \gamma_2] = \begin{bmatrix} 0.5 & 0 & 0.25 & 0 \\ 0.5 & 0.5 & 0 & 0 \end{bmatrix} \Gamma_2 = \begin{bmatrix} 1.280000 & 0.930909 & 1.120000 & 0.7418182 \\ 1.512727 & 1.241212 & 1.396364 & 1.0860606 \end{bmatrix}$$

*such that*

$$\Gamma_3 = \begin{bmatrix} 1.920000 & 1.105455 & 1.280000 & 0.930909 & 1.120000 & 0.7418182 \\ 1.105455 & 1.376970 & 1.512727 & 1.241212 & 1.396364 & 1.0860606 \\ 1.280000 & 1.512727 & 1.920000 & 1.105455 & 1.280000 & 0.930909 \\ 0.930909 & 1.241212 & 1.105455 & 1.376970 & 1.512727 & 1.241212 \\ 1.120000 & 1.396364 & 1.280000 & 1.512727 & 1.920000 & 1.105455 \\ 0.7418182 & 1.0860606 & 0.930909 & 1.241212 & 1.105455 & 1.376970 \end{bmatrix}$$

*and*

$$\begin{aligned}[\gamma_1, \gamma_2, \gamma_3] &= \begin{bmatrix} 0.5 & 0 & 0.25 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \end{bmatrix} \Gamma_3 \\ &= \begin{bmatrix} 1.280000 & 0.930909 & 1.120000 & 0.7418182 & 0.880000 & 0.6036364 \\ 1.512727 & 1.241212 & 1.396364 & 1.0860606 & 1.258182 & 0.9139394 \end{bmatrix}\end{aligned}$$

*Calculating the minimum norm solutions for* $p = 2, 3$ *gives*

$$[\hat{a}_1, \hat{a}_2] = \begin{bmatrix} 0.5 & 0.08333333 & 0.2083333 & -0.04166667 \\ 0.5 & 0.16666667 & 0.1666667 & 0.16666667 \end{bmatrix}$$

$$[\hat{a}_1, \hat{a}_2, \hat{a}_3] = \begin{bmatrix} 0.5 & 0.09375 & 0.203125 & -0.015625 & -0.015625 & -0.015625 \\ 0.5 & 0.12500 & 0.187500 & 0.062500 & 0.062500 & 0.062500 \end{bmatrix}$$

*which shows the dependence on the lag* $p$*.*

Second the minimum norm solution does not necessarily result in a coprime pair $[a(z), b]$.

**Example 4.4.2.** *Take the same process as in Example 4.4.1 where the block companion matrices of the original system and of the minimum norm solution are*

$$A = \begin{bmatrix} 0.5 & 0 & 0.25 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \hat{A} = \begin{bmatrix} 0.5 & 0.08333333 & 0.2083333 & -0.04166667 \\ 0.5 & 0.16666667 & 0.1666667 & 0.16666667 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

*with corresponding eigenvalues*

$$EV_A = \{0.809017, 0.500000, -0.309017, 0\}, EV_{\hat{A}} = \{0.8090170, 0.5000000, -0.3090170, -0.3333333\}$$

*The smallest eigenvalues of both sets correspond to unreachable modes of the pair* $[A, B]$ *and* $[\hat{A}, B]$ *respectively, where* $B = [1, 0, 0, 0]'$*. But as the eigenvalue* $-1/3$ *of* $\hat{A}$ *corresponds to a zero* $3$ *of the*

*corresponding polynomial $\hat{a}(z)$, the polynomial $a(z)$ has a zero at infinity and thus $[a(z), b]$ is left coprime whereas $[\hat{a}(z), b]$ is not. Additionally we see that 3 eigenvalues are the same ($\Gamma_2$ has rank 3) which affirms the results of Theorem 4.3.2 and Theorem 4.3.3.*

Especially the latter example shows that the minimum norm solution is not the ultimate solution of singular AR systems, as coprimeness is something desirable.

### 4.4.2 Canonical Representative II

In our conference paper [Anderson et al., 2009] and our current working paper [Anderson et al., 2010] we present a canonical representative of singular autoregressive models using row polynomial-echelon or Popov forms for a certain set of rows of the AR polynomial. Contrary to the previous sections we will not deal with an AR representation $[a(z), b]$ with $a(z) = I - a_1 z - ... - a_p z^p$ but with representations of the form $\left[d(z), \begin{bmatrix} I \\ 0 \end{bmatrix}\right]$ where $d(z) = d_0 - d_1 z - ... - d_p z^p$ is a stable polynomial and $d(0) \neq I$ may hold. Note that every representation $[d(z), [I, 0]']$ has a unique $[a(z), b] = d(0)^{-1}[d(z), [I, 0]']$ but not vice versa as any regular matrix $A$ with $Ab = [I, 0]'$ yields a pair $[d(z), [I, 0]']$. Of course such a matrix $A$ always exists, one example would be $A = \begin{bmatrix} (b'b)^{-1}b' \\ b^{\perp} \end{bmatrix}$ where $b^{\perp}$ is a basis for the left kernel of $b$.
Thus we start with a pair

$$d(z) = \begin{bmatrix} d_1(z) \\ d_2(z) \end{bmatrix}, c = \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{4.28}$$

where $d_1(z)$ and $d_2(z)$ are defined in an evident way. The first lemma will show that we can assume without loss of generality that this pair is left coprime.

**Lemma 4.4.3.** *Let $[d(z), c]$ be an arbitrary pair as in* (4.28)*, then there exists an observationally equivalent coprime pair $[\bar{d}(z), c]$ with $v(\bar{d}(z)) \leq v(d(z))$, where $v(d(z))$ denotes the degree of $d(z)$.*

*Proof.* ***(Lemma 4.4.3.)*** See also [Anderson et al., 2010].
It is clear that $[d(z), c]$ is not left coprime if and only if the rank of $d_2(z)$ drops for some $z_0 \in \mathbb{C}$. Using e.g. the Smith-McMillan form of $d_2$ we find always a decomposition $d_2 = e\bar{d}_2$ where $\bar{d}_2$ is of full rank $\forall z \in \mathbb{C}$ and with the same dimensions as $d_2$. Thus $[d_2(z), c] = \left[\begin{bmatrix} d_1(z) \\ \bar{d}_2(z) \end{bmatrix}, c\right]$ is a left coprime pair and it is observationally equivalent with $[d(z), c]$ as $[d(z), c] = \begin{bmatrix} I & 0 \\ 0 & e(z) \end{bmatrix} \left[\begin{bmatrix} d_1(z) \\ \bar{d}_2(z) \end{bmatrix}, c\right]$. It remains to show that the degree of $\bar{d}$ is lower or equal to the degree of $d$.
Let us denote with $d_{2,[i,]}$ the $i$-th row of $d_2$ and with $e_{i,j}$ the $i, j$ element of $e$. As $d_2$ can be assumed to be row reduced without loss of generality (pre-multiplying with unimodular matrices will give us that)

then by Theorem 6.3-13. on page 387 in [Kailath, 1980] there holds

$$v(d_{2[i,]}) = \max_j(v(e_{i,j}) + v(\bar{d}_{2[j,]}))$$

where $v(a) = -\infty$ if all elements of $a$ are zero. Let $p_{max} = \max_j(v(\bar{d}_{2[j,]}))$ then for each $i$ the corresponding row of $e$ cannot have only zeros and thus $v(d_{2[i,]}) \geq p_{max}$ holds for all $i$. □

The members of the equivalence class of autoregressive polynomials of the form $[d(z), c]$ are linked by matrices of the form

$$v = \begin{bmatrix} I & v_{12} \\ 0 & v_{22} \end{bmatrix} \tag{4.29}$$

If $[d(z), c]$ is left coprime, all other observationally equivalent left coprime representations are linked via a matrix $v$ (4.29) where $v_{22}$ is unimodular.

### Choice of $v_{22}$

**Definition 4.4.4. (row polynomial-echelon or Popov form).** *An $(m \times n)$ polynomial matrix $X(z)$ of rank$m$ is said to be in **Popov form or row polynomial-echelon form** if the following properties hold:*

*(i) It is row reduced and the row degrees are in descending order, say $k_1 \geq k_2 \geq ... \geq k_m$.*

*(ii) For row $i$ with $1 \leq i \leq m$, there is a pivot index $p_i$ such that $X_{i,p_i}$ is monic and has degree $k_i$, and $v(X_{i,j}) < k_i$ for all $j > p_i$*

*(iii) If $k_i = k_j$ and $i < j$, then $p_i < p_j$, i.e. the pivot indices corresponding to the same row degree are increasing*

*(iv) $X_{i,p_j}$ has degree less than $k_j$ if $i \neq j$.*

**Theorem 4.4.5.** *Let $X(z)$ be an $(m \times n)$ row reduced polynomial matrix, then $X(z)$ can be transformed by pre-multiplying unimodular matrices to Popov form. Moreover it is canonical, i.e. any other polynomial matrix which is linked to $X(z)$ by pre-multiplying $X$ with a polynomial matrix, has the same Popov form.*

*Proof.* ***(Theorem 4.4.5.)***
See Theorem 2 on page 500 in [Forney jr., 1975]. □

Therefore $v_{22}$ is chosen to bring $d_2(z)$ into Popov form.

### Choice of $v_{12}$

Note, that the choice of (the polynomial matrix) $v_{12}$ is not affected by the choice of $v_{22}$ as $v_{12}d_2 = v_{12}v_{22}^{-1}v_{22}d_2$, and as we started with a left coprime pair $[d(z), c]$, $v_{22}$ is unimodular and thus $v_{12}v_{22}^{-1}$ is polynomial. So we start with a pair $[d(z), c]$, where $d_2$ is already in Popov form.
Note, that the Popov form also gives for each row a pivot index. For the ease of simplicity assume that all pivot indices are in the first $m$ columns such that

$$d = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \tag{4.30}$$

where $d_{21}$ contains all pivot indices of $d_2$. Obviously $d_{21}$ is column reduced. Then applying the Division Theorem for Polynomial Matrices (Theorem 6.3-15. on page 389 in [Kailath, 1980]) we get unique polynomial matrices $\tilde{v}_{12}$ and $r_1$

$$d_{11} = \tilde{v}_{12}d_{21} + r_1 \tag{4.31}$$

where $r_1 d_{21}^{-1}$ is strictly proper. This implies, as $d_{21}$ is column reduced (Lemma 6.3-11. on page 385 in [Kailath, 1980]), that each column of $r_1$ has degree less than the degree of the corresponding column of $d_{21}$. Thus choosing $v_{12} = -\tilde{v}_{12}$ yields a unique $\bar{d}_1 = r_1$.

### Summary

We started with an arbitrary pair $[a(z), b]$ defining a singular AR system (4.27). Through pre-multiplying by a regular matrix we yield in an observationally equivalent pair $[d(z), c]$, where $c = [I, 0]'$. If $[d(z), c]$ is not left coprime one can always find an observationally equivalent pair $[\bar{d}(z), c]$ which is left coprime. By pre-multipling this pair with a unimodular matrix $v$ as in (4.29) gives the canonical representative, if $v_{22}$ and $v_{12}$ are choosen to bring $\bar{d}_2$ to Popov form first and forcing the column degrees of the sub matrix $\bar{d}_{11}$ to be bounded by the corresponding column degrees of the new $\bar{d}_{21}$ afterwards.

Let $[d(z), c]$ be the canonical representative for the systems with $c = [I, 0]'$ then by pre-multiplying $[d(z), c]$ with $d(0)^{-1}$ yields a canonical $[a(z), b]$ which is coprime and has degree lower than or equal to the degree of the starting polynomial.

### 4.4.3 Canonical Representative III

In our most recent paper [Deistler et al., 2010c] we present an additional way to choose a canonical representative by defining a suitable minimal state containing past values of $f_t$. Note that we can always write an AR system

$$f_t = a_1 f_{t-1} + \cdots + a_p f_{t-p} + \nu_t$$

where $\nu_t = b\varepsilon_t$, as a state space system

$$x_t = \begin{bmatrix} f_t \\ \vdots \\ \vdots \\ \vdots \\ f_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_p \\ I & 0 & \dots & 0 \\ 0 & I & 0 & \dots \\ \vdots & \ddots & \ddots & \\ & & I & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ \vdots \\ \vdots \\ \vdots \\ f_{t-p} \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \nu_t = Ax_{t-1} + B\nu_t \tag{4.32}$$

$$f_t = [a_1, ..., a_p]x_{t-1} + \nu_t \tag{4.33}$$

where $x_t = [f_t', ..., f_{t-p+1}']$ is an $rp$ dimensional state if $f_t$ is $r$ dimensional. Of course $x_t$ does not need to be a minimal state. Thus we perform two steps.

**Step 1**

Choose the first independent components of $x_{t-1}$ that build a basis for the linear span of $x_{t-1}$. There are indices $0 \leq m_i, i = 1, ..., r$ corresponding to the $r$ components of $f_t$ such that $m_1 + ... + m_r = n_1$ where $n_1$ is the dimension of the basis. An $m_i = 0$ for some $i$ means that the $i$-th component does not appear in the state at all. The basis components are $\{f_{t-1,1}, ..., f_{t-m_1,1}, f_{t-1,2}, ..., f_{t-m_2,2}, ..., f_{t-1,r}, ..., f_{t-m_r,r}\}$, where $f_{t-i,j}$ denotes the $j$-th component of $f_{t-i}$. This can be seen as follows. Suppose that $f_{t-i,j}$ is not linearly independent from the preceding components, i.e. $f_{t-i,j} = [f_{t,1}, ..., f_{t-i,j-1}]c$ for some $c$, then $f_{t-i-1,j}$ cannot be linearly independent from its preceding components as $f_{t-i-1,j} = zf_{t-i,j} = z[f_{t,1}, ..., f_{t-i,j-1}]c = [f_{t-1,1}, ..., f_{t-i-1,j-1}]c$.

Of course $\tilde{x}_{t-1} = [f_{t-1,1}, ..., f_{t-m_1,1}, f_{t-1,2}, ..., f_{t-m_2,2}, ..., f_{t-1,r}, ..., f_{t-m_r,r}]' = S_1 x_{t-1}$, where $S_1$ is a selector matrix, is a state and the corresponding AR polynomial $[\tilde{a}_1, ..., \tilde{a}_p]$ has zero columns corresponding to non basis components such that $f_t = [\tilde{a}_1, ..., \tilde{a}_p]x_{t-1}$ holds. This polynomial is of course unique as the corresponding Yule-Walker equations have a unique solution ($\mathbb{E}\tilde{x}_t\tilde{x}_t'$ is regular).

**Step 2**

Although the components of $\tilde{x}_{t-1}$ build a basis of the linear span of $x_{t-1}$, and $\tilde{x}_{t-1}$ is a state, $\tilde{x}_{t-1}$ does not need to be a minimal state as some components might not be necessary for any predictor. Thus we pick only the components of $\tilde{x}_{t-1}$ which are really needed and denote it as $\bar{x}_{t-1}$. Again there are indices $0 \leq \bar{m}_i, i = 1, ..., r$ corresponding to the $r$ components of $f_t$ such that
$\bar{x}_{t-1} = [f_{t-1,1}, ..., f_{t-\bar{m}_1,1}, f_{t-1,2}, ..., f_{t-\bar{m}_2,2}, ..., f_{t-1,r}, ..., f_{t-\bar{m}_r,r}]'$. To see this note that if $f_{t-i,j}$ is never used for a predictor $f_{t+j|t-1}, j \geq 0$ we get from $f_{t|t-1}$ that the $j$-th column of $a_i$ is zero. As $f_{t+1|t-1} = a_1^2 f_{t-1} + [a_1a_2 + a_2]f_{t-2} + ... + [a_1a_{p-1} + a_p]x_{t-p+1} + a_1a_p x_{t-p}$ we have that the $j$-th

column of $[a_1 a_i + a_{i+1}]$ and thus the $j$-th column of $a_{i+1}$ must be zero. By induction the result follows.

This construction yields a state which has the minimal number of components of past $f_t$ and we have

$$\bar{x}_t = \bar{A}\bar{x}_{t-1} + \bar{B}\varepsilon_t \tag{4.34}$$

$$f_t = C\bar{x}_{t-1} + D\varepsilon_t \tag{4.35}$$

where $\bar{B} = (\bar{b}', 0, ..., 0)'$ and $\bar{b}$ are the rows of $b$ corresponding to the actually appearing elements of $f_{t-1}$ in $\bar{x}_{t-1}$. Furthermore $\bar{A}$ is in a "quasi companion form" which means that $\bar{A}$ has almost the structure of a block companion form, in particular the last rows have a structure $[I, 0]S$ where $S$ shuffles the columns of $[I, 0]$. This can be seen as follows: The first block of rows of $\bar{A}$ corresponds to the one-step-ahead predictor of $f_{t|t-1}$. The remaining rows have the special structure $[I, 0]S$ as the components of $\bar{x}_t$ corresponding to time $t-1$, i.e. the components of $f_{t-i}$ which are in the state, are also in the shifted state $\bar{x}_{t-1}$ as the same components of $f_{t-i+1}$ must occur in $\bar{x}_t$.

**Lemma 4.4.6.** *The system* (4.34) *is reachable if and only if it is stable.*

*Proof.* ***(Lemma 4.4.6.)***
Let $y$ be an eigenvector of $A$ with eigenvalue $\lambda$, i.e. $y'A = \lambda y'$, then by (4.34) we have

$$(1 - |\lambda|^2) \underbrace{y'\Gamma_{\bar{x}}\bar{y}}_{>0} = \underbrace{y'\bar{B}\bar{B}'\bar{y}}_{\geq 0} \tag{4.36}$$

where $\Gamma_{\bar{x}} = \mathbb{E}\bar{x}_t\bar{x}_t'$. Obviously the left hand side of (4.36) is zero if and only if $\lambda$ is of modulus one and thus $y'\bar{B}\bar{B}'\bar{y}$ is zero if and only if $\lambda$ is of modulus one. Therefore $[I\lambda - \bar{A}, \bar{B}]$ is not of full rank for all $\lambda \in \mathbb{C}$ if and only if $\bar{A}$ has an eigenvalue of modulus one. □

Fill the quasi companion matrix $\bar{A}$ (with zeros or with coefficients corresponding to the one-step-ahead predictor, if some elements of $f_{t-1}$ are not in the state $\bar{x}_{t-1}$) such that $A$ be the block companion matrix corresponding to the state $x_t$. Then $A$ has the same eigenvalues as $\bar{A}$ plus some zero eigenvalues corresponding to the components which cancel out through Step 1 and 2.

**Example 4.4.7.** *Let $f_t$ be a stationary solution of an AR(3) process*

$$f_t = \begin{bmatrix} f_{t,1} \\ f_{t,2} \end{bmatrix} = \begin{bmatrix} x & 0 \\ x & x \end{bmatrix} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \end{bmatrix} + \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \begin{bmatrix} f_{t-2,1} \\ f_{t-2,2} \end{bmatrix} + \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \begin{bmatrix} f_{t-3,1} \\ f_{t-3,2} \end{bmatrix} + \begin{bmatrix} \nu_{t,1} \\ \nu_{t,2} \end{bmatrix} \tag{4.37}$$

*where the $x$ entries are arbitrary subject to the polynomial is stable. Then*

$\bar{x}_{t-1} = [f_{t-1,1}, f_{t-1,2}, f_{t-2,1}, f_{t-3,1}]'$ *is a minimal state if its components build a basis. Thus*

$$\begin{bmatrix} f_{t,1} \\ f_{t,2} \\ f_{t-1,1} \\ f_{t-2,1} \end{bmatrix} = \underbrace{\begin{bmatrix} x & 0 & x & x \\ x & x & x & x \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\bar{A}} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \\ f_{t-2,1} \\ f_{t-3,1} \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\bar{B}} \begin{bmatrix} \nu_{t,1} \\ \nu_{t,2} \end{bmatrix} \tag{4.38}$$

*is the stable AR(1) polynomial for the "minimal" state, where the coefficient matrix $\bar{A}$ has almost the form of a block companion matrix, and the corresponding block companion matrix $A$ for the state $x_t$ is*

$$\begin{bmatrix} f_{t,1} \\ f_{t,2} \\ f_{t-1,1} \\ f_{t-1,2} \\ f_{t-2,1} \\ f_{t-2,2} \end{bmatrix} = \underbrace{\begin{bmatrix} x & 0 & x & 0 & x & 0 \\ x & x & x & 0 & x & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}}_{A} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \\ f_{t-2,1} \\ f_{t-2,2} \\ f_{t-3,1} \\ f_{t-3,2} \end{bmatrix} + \begin{bmatrix} \nu_{t,1} \\ \nu_{t,2} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.39}$$

*By the Laplace expansion of* $\det[\lambda I - A]$ *it is obvious that there are two and only two additional zero eigenvalues of $A$ compared to $\bar{A}$.*

Note, that if it happens that a component of $f_{t-1}$, $j$ say, is not in the minimal state $\bar{x}_{t-1}$, the first block of rows in $\bar{A}$ has fewer rows than the first block of rows in $A$. In such a case the $j$-th column of $A$ has all entries equal to zero, except the one corresponding to the shift. This implies that the $j$-th column of $\lambda I - A$ has only two entries namely the $j$-th entry, which is equal to $\lambda$, and the one corresponding to the shift, which is equal to one. The latter cancels out by the Laplace expansion of the determinant of $\lambda I - A$ and thus an additional zero eigenvalue occurs.

**Example 4.4.8.** *Use the same example as above with one exception*

$$f_t = \begin{bmatrix} f_{t,1} \\ f_{t,2} \end{bmatrix} = \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \end{bmatrix} + \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \begin{bmatrix} f_{t-2,1} \\ f_{t-2,2} \end{bmatrix} + \begin{bmatrix} x & 0 \\ x & 0 \end{bmatrix} \begin{bmatrix} f_{t-3,1} \\ f_{t-3,2} \end{bmatrix} + \begin{bmatrix} \nu_{t,1} \\ \nu_{t,2} \end{bmatrix} \tag{4.40}$$

*Then the "minimal" state is $\bar{x}_t = [f_{t-1,1}, f_{t-2,1}, f_{t-3,1}]'$ and the corresponding $\bar{A} = \begin{bmatrix} x & x & x \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. The*

*block companion matrix $A$ of $x_t$ is then*

$$\begin{bmatrix} f_{t,1} \\ f_{t,2} \\ f_{t-1,1} \\ f_{t-1,2} \\ f_{t-2,1} \\ f_{t-2,2} \end{bmatrix} = \underbrace{\begin{bmatrix} x & 0 & x & 0 & x & 0 \\ x & 0 & x & 0 & x & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}}_{A} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \\ f_{t-2,1} \\ f_{t-2,2} \\ f_{t-3,1} \\ f_{t-3,2} \end{bmatrix} + \begin{bmatrix} \nu_{t,1} \\ \nu_{t,2} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.41}$$

In the two examples above we always started with a stable $A$ (which corresponded already to an extended version of a $\bar{A}$ matrix) which implied that $\bar{A}$ was stable and thus by Lemma 4.4.6 the system (4.34), (4.35) was reachable.

**Theorem 4.4.9.** *The system* (4.34)*,* (4.35) *is reachable and stable if and only if there is an underlying stable system.*

*Proof.* ***(Theorem 4.4.9.)***
One direction is evident. If there is no underlying stable system, $\bar{A}$ and thus the corresponding $A$ (with zero columns in the coefficient matrices) are not stable and thus not reachable. The other direction can be seen as follows. If there is an underlying stable system the corresponding AR process $f_t$ is linearly regular which implies that also $\bar{x}_t$ is linearly regular. Thus

$$\Gamma_{\bar{x}} = [\bar{B}, \bar{A}\bar{B}, ...][\bar{B}, \bar{A}\bar{B}, ...]'$$

holds and as $\Gamma_{\bar{x}}$ is regular, $[\bar{B}, \bar{A}\bar{B}, ...]$ must have full rank which implies reachability and by Lemma 4.4.6 stability. □

**Corollary 4.4.10.** *Let $[a(z), b]$ be the pair where $a(z)$ corresponds to the polynomial with block companion matrix $A$ which is the extended version of $\bar{A}$ from the system* (4.34) *and* (4.35)*, then $[a(z), b]$ is left coprime and stable if and only if there is an underlying stable system.*

*Proof.* ***(Corollary 4.4.10.)***
The stability of $a(z)$ is an immediate consequence of Theorem 4.4.9. Coprimeness of $[a(z), b]$, i.e. $[a(z), b]$ has full rank for all $z \in \mathbb{C}$, is equivalent to $[I\lambda - A, B]$ has full rank for all $0 \neq \lambda \in \mathbb{C}$ (can be seen easily by (4.15) and the following paragraph), as a singularity for $\lambda = 0$ corresponds to a zero of $a(z)$ at infinity and a singularity for $\lambda = \infty$ is equivalent to $a(0)$ is singular. As $[I\lambda - A, B]$ has full rank for all $0 \neq \lambda \in \mathbb{C}$ (consequence of Theorem 4.4.9) the result follows. □

# Chapter 5

# Estimation of Singular AR Systems

Here we want to present estimation procedures for singular autoregressive system

$$f_t = a_1 f_{t-1} + ... + a_p f_{t-p} + \nu_t \tag{5.1}$$

discussed in the previous section. First, it is shown, that given the integer parameters, such as $p$ the minimal order of the AR system and $s$ the rank of the covariance matrix $\Gamma_p$, the minimum norm solution can be estimated consistently. Second, procedures are presented to estimate the minimal order $p$.

## 5.1 Estimating the Minimum Norm Solution

Denote by

$$\hat{\gamma}_j^T = \sum_{t=1+j}^{T} f_t f_{t-j} \tag{5.2}$$

and for $j \leq h$

$$\tilde{\gamma}_j^{T,h} = \sum_{t=1+h}^{T} f_t f_{t-j} \tag{5.3}$$

two estimates of the covariances $\gamma_j = \mathbb{E} f_t f'_{t-j}$. Consequently

$$\hat{\Gamma}_p^T = \begin{bmatrix} \hat{\gamma}_0^T & \hat{\gamma}_1^T & \cdots & \hat{\gamma}_{p-1}^T \\ \hat{\gamma}_1^{T\prime} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_1^T \\ \hat{\gamma}_{p-1}^{T\prime} & \cdots & \hat{\gamma}_1^{T\prime} & \hat{\gamma}_0^T \end{bmatrix} \tag{5.4}$$

and if $p \leq h$

$$\tilde{\Gamma}_p^{T,h} = \begin{bmatrix} \tilde{\gamma}_0^{T,h} & \tilde{\gamma}_1^{T,h} & \cdots & \tilde{\gamma}_{p-1}^{T,h} \\ \tilde{\gamma}_1^{T,h'} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tilde{\gamma}_1^{T,h} \\ \tilde{\gamma}_{p-1}^{T,h'} & \cdots & \tilde{\gamma}_1^{T,h'} & \tilde{\gamma}_0^{T,h} \end{bmatrix} \tag{5.5}$$

We will use the following well known theorem to establish consistency of the estimation of the minimum norm solution.

**Theorem 5.1.1.** *Let $f_t$ be an AR(p) process generated by* (5.1)*. If $\operatorname{rk}\Gamma_p = s \leq pr$ and if all nonzero eigenvalues of $\Gamma_p$ are distinct, then*

$$[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] = [\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T] O_p^s [\Lambda_p^s]^{-1} O_p^{s'} \tag{5.6}$$

*where $\hat{\Gamma}_p^T = [O_p^s, O_p^x] \begin{bmatrix} \Lambda_p^s & 0 \\ 0 & \Lambda_p^x \end{bmatrix} \begin{bmatrix} O_p^{s'} \\ O_p^{x'} \end{bmatrix}$, defines a function of $\hat{\gamma}_0^T, ..., \hat{\gamma}_p^T$ which is continuous at $\gamma_0, ..., \gamma_p$ (here the $\hat{\gamma}_j^T$ as well as the $\hat{\hat{a}}_j$ are considered to be matrices with real entries).*

*Proof.* ***(Theorem 5.1.1.)*** See also [Deistler et al., 2010a].
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 (5.6) is obviously a continuous function at $\gamma_0, ..., \gamma_p$. □

**Theorem 5.1.2.** *Let $f_t$ be an AR(p) process generated by* (5.1)*. If $\operatorname{rk}\Gamma_p = s < pr$ and if all nonzero eigenvalues of $\Gamma_p$ are distinct, then*

$$[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] = [\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T] O_p^s [\Lambda_p^s]^{-1} O_p^{s'}$$

*as in Theorem 5.1.1, is a consistent estimator of the minimum norm solution of the Yule-Walker equations.*

*Proof.* ***(Theorem 5.1.2.)*** See also [Deistler et al., 2010a].
As $\hat{\gamma}_j^T$ yields a (strong) consistent estimator of $\gamma_j$ for $T \to \infty$, $[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p]$ is (by the continuity result of Theorem 5.1.1) a (strong) consistent estimator of $[\hat{a}_1, ..., \hat{a}_p]$. □

**Remark 5.1.3.** *Of course Theorem 5.1.2 also holds if the covariance estimators $\tilde{\gamma}_j^{T,h}$ are used. We will call* (5.6) *the "Yule-Walker" estimator of the minimum norm solution and the estimator using $\tilde{\gamma}_j^{T,h}$ instead of $\hat{\gamma}_j^T$ the "OLS" estimator.*

The weak point of this estimating procedure is the need of the rank of the covariance matrix $\Gamma_p$.

Fortunately there is a possibility to determine this rank by using $\tilde{\Gamma}_p^{T,h}$. Note that

$$\tilde{\Gamma}_p^{T,h} = 1/T \begin{bmatrix} f_p & f_{p+1} & \dots & f_T \\ f_{p-1} & f_p & \dots & f_{T-1} \\ \vdots & \vdots & & \vdots \\ f_1 & f_2 & \dots & f_{T-p+1} \end{bmatrix} \begin{bmatrix} f'_p & \dots & f'_1 \\ \vdots & & \vdots \\ f'_t & \dots & f'_{T-p+1} \end{bmatrix}$$

has the same kernel as $\Gamma_p$ and therefore they have the same rank. Thus we can "estimate" $s$ by counting the eigenvalues larger than a certain boundary near zero. Also note that

$$\hat{\Gamma}_p^T = 1/T \begin{bmatrix} f_1 & \dots & f_T & 0 & \dots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & \ddots & \vdots \\ 0 & \dots & f_1 & f_2 & \dots & f_T \end{bmatrix} \begin{bmatrix} f'_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ f'_p & \dots & & f'_1 \\ \vdots & & & \vdots \\ f'_T & \dots & & f'_{T-p+1} \\ 0 & \ddots & \ddots & f'_{t-p+2} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & & f'_T \end{bmatrix}$$

has not exactly the same rank as $\Gamma_p$ and will therefore have some eigenvalues not equal but near to zero. The problem of estimating the rank will of course be even more complicated if the autoregressive process is not observed but has to be estimated, as it is the case in factor models. A test on the eigenvalues might be of interest. A possibility might be the test described in [Robin and Smith, 2000].

The next theorem shows, that a slightly different (we only remarked that slight difference in our paper [Deistler et al., 2010a] in the proof) estimation of the minimum norm solution than (5.6) always yields a stable polynomial, irrespective if the underlying system is stable or has a unit circle zero. Define the

$(T+p)\times rp$ matrices

$$Z = \begin{bmatrix} 0 & \cdots & & 0 \\ f_1' & 0 & \cdots & 0 \\ f_2' & \ddots & & \vdots \\ \vdots & \vdots & \vdots & f_1' \\ \vdots & \vdots & \vdots & \vdots \\ f_T' & f_{T-1}' & \cdots & f_{T-p+1}' \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & f_T' & f_{T-1}' \\ 0 & \cdots & 0 & f_T' \end{bmatrix} \tag{5.7}$$

and

$$Y = \begin{bmatrix} f_1' & 0 & \cdots & 0 \\ f_2' & \ddots & & \vdots \\ \vdots & \vdots & \vdots & f_1' \\ \vdots & \vdots & \vdots & \vdots \\ f_T' & f_{T-1}' & \cdots & f_{T-p+1}' \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & f_T' & f_{T-1}' \\ \vdots & \cdots & 0 & f_T' \\ 0 & \cdots & 0 & 0 \end{bmatrix} \tag{5.8}$$

and let

$$Z = U_1\Omega_1 V_1' = [U_{1s}, U_{1x}]\begin{bmatrix}\Omega_{1s} & 0 \\ 0 & \Omega_{1x}\end{bmatrix}\begin{bmatrix}V_{1s}' \\ V_{1x}'\end{bmatrix},\ Y = U_2\Omega_2 V_2' = [U_{2s}, U_{2x}]\begin{bmatrix}\Omega_{2s} & 0 \\ 0 & \Omega_{2x}\end{bmatrix}\begin{bmatrix}V_{2s}' \\ V_{2x}'\end{bmatrix} \tag{5.9}$$

be their "thin" Singular Value Decompositions[1], and the submatrices indexed with $s$ correspond to the $s$ largest singular values. It is easy to see that

$$V_{1s}\Omega_{1s}^2 V_{1s}' = O_p^s \Lambda_p^s O_p^{s'}$$

**Theorem 5.1.4.** *(i) If* $\operatorname{rk}\Gamma_p = pr$ *holds, then the "Yule-Walker" estimator* (5.6) *of the minimum norm solution yields a stable autoregression*

[1]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 $U \in \mathbb{R}^{m\times h}, \Sigma \in \mathbb{R}^{h\times h}, V \in \mathbb{R}^{h\times h}$.

*(ii) For* $\operatorname{rk}\Gamma_p = s < pr$*, the solution*

$$[\bar{a}_1, ..., \bar{a}_p] = [\bar{\gamma}_1^T, ..., \bar{\gamma}_p^T] O_p^s [\Lambda_p^s]^{-1} O_p^{s'} \tag{5.10}$$

*corresponds to a stable autoregression, where* $[\bar{\gamma}_1^T, ..., \bar{\gamma}_p^T] = [V_2]_{[1:r,]} \Omega_{2s} U_{2s}' U_{1s} \Omega_{1s} V_{1s}'$ *and the matrices* $V_2, \Omega_{2s}, U_{2s}, U_{1s}, \Omega_{1s}$ *and* $V_{1s}$ *are matrices of the Singular Value Decompositions of* $Y$ *and* $Z$*, and* $[A]_{[i:j,]}$ *denote the rows* $i$ *to* $j$ *of a matrix* $A$*.*

*Proof.* ***(Theorem 5.1.4.)*** See also [Deistler et al., 2010a].
(i) As we need to show $\det \hat{\hat{a}}(z) \neq 0,\ |z| \leq 1$ we proceed as follows: Let

$$\hat{\hat{A}} = \begin{bmatrix} \hat{\hat{a}}_1 & \cdots & \cdots & \hat{\hat{a}}_{p-1} & \hat{\hat{a}}_p \\ I_r & 0 & \cdots & 0 & 0 \\ 0 & I_r & \ddots & \vdots & 0 \\ \vdots & \cdots & I_r & 0 & 0 \\ 0 & \cdots & 0 & I_r & 0 \end{bmatrix}$$

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

$$\det[\hat{\hat{A}} - zI_{rp}] \tag{5.11}$$

are within the unit circle. Note that $\Omega_1 = \Omega_2 =: \Omega$ 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

$$\hat{\hat{A}} = Y'Z[Z'Z]^{-1}$$

and therefore

$$\hat{\hat{A}} = V\Omega U_2' U_1 \Omega^{-1} V'$$

holds. The roots of (5.11) are the same as the roots of

$$\det[\Omega^{-1}V'[\hat{\hat{A}} - zI_{rp}]V\Omega] = \det[U_2'U_1 - zI_{rp}]$$

Now we have

$$|\lambda_{max}[U_2'U_1]| \leq max_{\|x\|=1}|x^*U_2'U_1x| < max_{\|x\|=1}|x^*U_1'U_1x| = 1$$

where the first inequality can be shown very easily and the strict inequality is valid because of the form of $U_1 = [0, U']'$ and $U_2 = [U', 0]'$ (strict Cauchy-Schwarz inequality).

(ii) First we want to repeat that, as is well known, the sample covariance matrix $\hat{\Gamma}_p$ is "typically" nonsingular. As we omit the smallest eigenvalues of $\hat{\Gamma}_p$ we have to think about the consequences on the matrices $Y$ and $Z$ defined above. We define by using the Singular Value Decompositions (5.9)

$$\bar{Z} = [U_{1s}, U_{1x}] \begin{bmatrix} \Omega_{1s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V'_{1s} \\ V'_{1x} \end{bmatrix} = U_{1s}\Omega_{1s}V'_{1s} \tag{5.12}$$

and

$$\bar{Y} = [U_{2s}, U_{2x}] \begin{bmatrix} \Omega_{2s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V'_{2s} \\ V'_{2x} \end{bmatrix} = U_{2s}\Omega_{2s}V'_{2s} \tag{5.13}$$

Following the argument in (i) it is clear that the last row of $U_{2s}$ is zero and we have $U_{2s} = [U'_s, 0]'$, where $U_{1s} = [0, U'_s]'$. Furthermore, as in (i), we can choose $\Omega_1 = \Omega_2 =: \Omega$, $V_1 = V_2 =: V_s$ and thus $\Omega_{1s} = \Omega_{2s} =: \Omega_s$ and $V_{1s} = V_{2s} =: V_s$. As can be easily seen $V_{[1:r(p-1),]}V' = \left[I_{r(p-1)}, 0\right]$, where $V_{[1:r(p-1),]}$ is $V$ from (5.12) without its last $r$ rows. Now observe that

$$[\bar{Y}']_{[1:r,]}\bar{Z} = [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s V'_s \tag{5.14}$$

holds, which is an estimate of $T[\gamma_1, \ldots, \gamma_p]'$ (note that if $\Omega_x \neq 0$ holds, then (5.14) does not correspond to the sample covariances $T[\hat{\gamma}_1, \ldots, \hat{\gamma}_p]'$). It follows that

$$\begin{aligned}[\bar{a}_1, ..., \bar{a}_p] = [\bar{Y}]'_{[1:r,]}\bar{Z}[\bar{Z}'\bar{Z}]^{\#} = [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s V'_s V_s \Omega_s^{-2} V'_s = \\ = [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s\Omega_s^{-2}V'_s = [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s^{-1}V'_s\end{aligned}$$

Therefore

$$\begin{aligned}\bar{A} = \begin{bmatrix} \bar{a}_1 & \cdots & \cdots & \bar{a}_{p-1} & \bar{a}_p \\ I_r & 0 & \cdots & 0 & 0 \\ 0 & I_r & \ddots & \vdots & 0 \\ \vdots & \cdots & I_r & 0 & 0 \\ 0 & \cdots & 0 & I_r & 0 \end{bmatrix} &= \begin{bmatrix} [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s^{-1}V'_s \\ V_{[1:r(p-1),]}V' \end{bmatrix} = \begin{bmatrix} [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s^{-1}V'_s \\ [[V_s]_{[1:r(p-1),]}, [V_x]_{[1:r(p-1),]}] \begin{bmatrix} V'_s \\ V'_x \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s^{-1}V'_s \\ [V_s]_{[1:r(p-1),]}V'_s + [V_x]_{[1:r(p-1),]}]V'_x \end{bmatrix} \\ &= \begin{bmatrix} [V_s]_{[1:r,]}\Omega_s U'_{2s}U_{1s}\Omega_s^{-1}V'_s \\ [V_s]_{[1:r(p-1),]}V'_s \end{bmatrix} + \begin{bmatrix} 0 \\ [V_x]_{[1:r(p-1),]}V'_x \end{bmatrix}\end{aligned}$$

holds. Call the first summand on the right hand side in the last equality above $B$ and the second summand $\tilde{V}$. Recall that we want to show that the roots of $det[\hat{A} - zI_{rp}]$ are within the unit circle. Now we have

(with the penultimate equality following using (5.15) and (5.16) below)

$$\begin{aligned} det[\bar{A} - zI_{rp}] &= det[B + \tilde{V} - zI_{rp}] = det[V'[B + \tilde{V} - zI_{rp}]V] \\ &= det\left[\overbrace{\begin{bmatrix}\Omega_s^{-1} & 0\\ 0 & I_{rp-s}\end{bmatrix}}^{\tilde{\Omega}^{-1}} V'[B + \tilde{V} - zI_{rp}]V \begin{bmatrix}\Omega_s & 0\\ 0 & I_{rp-s}\end{bmatrix}\right] \\ &= det\left[\tilde{\Omega}^{-1}V'[B + \tilde{V}]V\tilde{\Omega} - zI_{rp}\right] = det\left[\begin{bmatrix}U_{2s}'U_{1s} & \Omega_s^{-1}V_s'\bar{V}\\ 0 & V_x'\bar{V}\end{bmatrix} - zI_{rp}\right] \\ &= det[U_{2s}'U_{1s} - zI_s]det[V_x'\bar{V} - zI_{rp-s}] \end{aligned}$$

with $\bar{V} := \begin{bmatrix}0\\ [V_x]_{[1:r(p-1),]}\end{bmatrix}$. As the matrices $V_x'\bar{V} = V_x'\begin{bmatrix}0\\ [V_x]_{[1:r(p-1),]}\end{bmatrix} = [V_x', 0]\begin{bmatrix}0\\ V_x\end{bmatrix}$ and $U_{2s}'U_{1s}$ (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 $\bar{Z}$ and $\bar{Y}$ to justify the third equality)

$$\begin{aligned} B &= \begin{bmatrix}[V_s]_{[1:r,]}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s'\\ [V_s]_{[1:r(p-1),]}V_s'\end{bmatrix} = \begin{bmatrix}[V_s]_{[1:r,]}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s'\\ [V_s]_{[1:r(p-1),]}\Omega_s U_{1s}'U_{1s}\Omega_s^{-1}V_s'\end{bmatrix} \\ &= \begin{bmatrix}[V_s]_{[1:r,]}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s'\\ [V_s]_{[r+1:rp,]}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s'\end{bmatrix} = V_s\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s' \end{aligned}$$

where $[V_s]_{[r+1:rp,]}$ is the matrix consisting of the last $r(p-1)$ rows of $V_s$. Therefore

$$\begin{aligned} &\tilde{\Omega}^{-1}V'BV\tilde{\Omega} = \tilde{\Omega}^{-1}V'V_s\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}V_s'V\tilde{\Omega} \\ =&\tilde{\Omega}^{-1}\begin{bmatrix}I_s\\ 0\end{bmatrix}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}[I_s, 0]\tilde{\Omega} = \begin{bmatrix}\Omega_s^{-1}\Omega_s U_{2s}'U_{1s}\Omega_s^{-1}\Omega_s & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}U_{2s}'U_{1s} & 0\\ 0 & 0\end{bmatrix} \end{aligned} \tag{5.15}$$

and

$$
\begin{aligned}
&\tilde{\Omega}^{-1}V'\tilde{V}V\tilde{\Omega} \\
&= \begin{bmatrix} \Omega_s^{-1} & 0 \\ 0 & I_{rp-s} \end{bmatrix} V'\tilde{V}V \begin{bmatrix} \Omega_s & 0 \\ 0 & I_{rp-s} \end{bmatrix} = \begin{bmatrix} \Omega_s^{-1} & 0 \\ 0 & I_{rp-s} \end{bmatrix} V' \begin{bmatrix} 0 \\ [V_x]_{[1:r(p-1),]}V_x' \end{bmatrix} V \begin{bmatrix} \Omega_s & 0 \\ 0 & I_{rp-s} \end{bmatrix} \\
&= \begin{bmatrix} \Omega_s^{-1} & 0 \\ 0 & I_{rp-s} \end{bmatrix} V' \underbrace{\begin{bmatrix} 0 \\ [V_x]_{[1:r(p-1),]} \end{bmatrix}}_{\bar{V}} V_x'V \begin{bmatrix} \Omega_s & 0 \\ 0 & I_{rp-s} \end{bmatrix} = \begin{bmatrix} \Omega_s^{-1}V_s'\bar{V} \\ V_x'\bar{V} \end{bmatrix} [0, I_{rp-s}] \\
&= \begin{bmatrix} 0 & \Omega_s^{-1}V_s'\bar{V} \\ 0 & V_x'\bar{V} \end{bmatrix}
\end{aligned} \tag{5.16}
$$

□

**Remark 5.1.5.** *Note that* (5.10) *is of course a consistent estimator of the minimum norm solution as* $1/T[\bar{Y}'\bar{Z}]$ *is asymptotically equivalent to* $1/T[Y'Z]$*, and the latter estimates the corresponding covariances consistently.*

## 5.2 Comparison of Solutions of the Sample Yule-Walker Equations

We pointed already out, that two estimation procedures exist which consistently estimate the minimum norm solution. Nowadays the computer power is that tremendous that "almost" singular matrices, i.e. very badly conditioned matrices, can still be inverted. In our case this means that $\hat{\Gamma}_p$ can be inverted in many cases. The question arises if the "truncation" of $\hat{\Gamma}_p$, i.e. setting the smallest eigenvalues to zero, is reasonable. The answer is yes. The truncation ensures that the two estimation procedures are consistent estimates. Furthermore it is shown that one of them always gives a stable polynomial (the second one most likely gives a stable one too). Nevertheless, inverting $\hat{\Gamma}_p$, if possible, makes somehow sense too. Note that

$$
\begin{aligned}
[\check{a}_1, ..., \check{a}_p] =& [\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T]\Gamma_p^{-1} = [\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T][O_p^s[\Lambda_p^s]^{-1}O_p^{s'} + O_p^x[\Lambda_p^x]^{-1}O_p^{x'}] \\
=& [\hat{\check{a}}_1, ..., \hat{\check{a}}_p] + [\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T]O_p^x[\Lambda_p^x]^{-1}O_p^{x'}]
\end{aligned} \tag{5.17}
$$

and that $\lim_{T\to\infty}[\hat{\gamma}_1^T, ..., \hat{\gamma}_p^T]O_p^x = 0$ which implies that the second term on the right hand side of the equation above has a "$0 \cdot \infty$" behavior for $T$ going to infinity. So it is not clear if it has a limit but for forecasting purposes (and of course finite $T$) $[\check{a}_1, ..., \check{a}_p]$ can be an appropriate tool. Especially as it is somehow (the smallest eigenvalues are net equal to zero) another solution of the sample Yule-Walker equations and thus it is not that important to result in a stable polynomial, as the unstable zeros correspond to unreachable modes anyway (compare to section 4.3.2).

We compare the three estimation methods

1. $[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p]$ from (5.6)
2. $[\bar{a}_1, ..., \bar{a}_p]$ from (5.10)
3. $[\check{a}_1, ..., \check{a}_p]$ from (5.17)

where the third is calculated with the maximal number of eigenvalues of $\hat{\Gamma}_p$ such that the (pseudo) inverse can be computed. The dimension of the process is set equal to $r = 3$, the driving white noise has a dimension of $q = 1$ and the minimal order is set to $p = 2$. Furthermore the processes are simulated for $T = T_{in} + T_{out} = 150 + 100$ time points, where $T_{in}$ denotes the insample size and $T_{out}$ denotes the out-of-sample size, with coefficient matrices $[a_1, a_2] = \left[ \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}, \begin{bmatrix} x & x & x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right]$, where the $x$ denotes an arbitrary entry, such that $a(z) = I + a_1 z + a_2 z^2$ is stable (the $x$ entries are chosen arbitrarily and are multiplied by 0.9 until the polynomial is stable). The special structure of $a_2$ ensures that the true covariance matrix $\Gamma_2$ has at max rank 5, and can be of rank 4 if $b = [1, 0, 0]'$ (this is very unlikely to happen as the entries of $b$ are chosen arbitrarily too).
The three estimators are compared by simulating 100 different processes (with different coefficient matrices) and calculating the mean of

- The Frobenius norm of the insample covariance matrix of the errors: $\|\hat{\Sigma}_{\nu,in}\|_F$, where $\hat{\Sigma}_{\nu,in} = 1/T_{in} \sum_{t=3}^{T_{in}} \hat{\nu}_t \hat{\nu}_t'$ and $T_{in}$ denotes the insample size.
- The Frobenius norm of the out-of-sample covariance matrix of the errors: $\|\hat{\Sigma}_{\nu,out}\|_F$, where $\hat{\Sigma}_{\nu,out} = 1/T_{out} \sum_{t=1}^{T_{out}} \hat{\nu}_t \hat{\nu}_t'$ and $T_{out}$ denotes the out-of-sample size.

Furthermore

$$\|\hat{\hat{a}} - \check{a}\|_F = \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] - [\check{a}_1, ..., \check{a}_p]\|_F, \quad \|\hat{\hat{a}} - \bar{a}\|_F = \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] - [\bar{a}_1, ..., \bar{a}_p]\|_F$$

and

$$\|\hat{\hat{a}} - \check{a}\|_F / \|\hat{\hat{a}}\|_F = \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] - [\check{a}_1, ..., \check{a}_p]\|_F / \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p]\|_F$$

$$\|\hat{\hat{a}} - \bar{a}\|_F / \|\hat{\hat{a}}\|_F = \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p] - [\bar{a}_1, ..., \bar{a}_p]\|_F / \|[\hat{\hat{a}}_1, ..., \hat{\hat{a}}_p]\|_F$$

are calculated.

$r = 3, q = 1, p = 2, T_{in} = 150, T_{out} = 100$, using the "Yule-Walker" estimates for the covariances:

| | min. norm $(\hat{\hat{a}})$ | | | "inverting" $(\check{a})$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | stable models: 100 | | | stable models: 100 | | | | |
| # | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\|\hat{\hat{a}} - \check{a}\|_F$ | $\|\hat{\hat{a}} - \check{a}\|_F / \|\hat{\hat{a}}\|_F$ |
| 24 | 5 | 1.0840906 | 1.2171754 | 6 | 1.0833791 | 1.2174736 | 2.634527 | 1.730764 |
| 76 | 6 | 0.8119854 | 0.8681686 | 6 | 0.8119854 | 0.8681686 | 0.00000 | 0.00000 |

$r = 3, q = 1, p = 2, T_{in} = 150, T_{out} = 100$, using the "Yule-Walker" estimates for the covariances:

| | min. norm $(\hat{\hat{a}})$ | | | min. norm modified $(\bar{a})$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | stable models: 100 | | | stable models: 100 | | | | |
| # | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\|\hat{\hat{a}} - \bar{a}\|_F$ | $\|\hat{\hat{a}} - \bar{a}\|_F / \|\hat{\hat{a}}\|_F$ |
| 24 | 5 | 1.0840906 | 1.2171754 | 5 | 1.0840916 | 1.2171754 | 3.021762e-02 | 2.120281e-02 |
| 76 | 6 | 0.8119854 | 0.8681686 | 6 | 0.8119854 | 0.8681686 | 2.888728e-12 | 4.983646e-13 |

All calculations are repeated with the "OLS" estimates of the covariances as this method is often used in practice.

$r = 3, q = 1, p = 2, T_{in} = 150, T_{out} = 100$, using the "OLS" estimates for the covariances:

| | min. norm $(\hat{\hat{a}})$ | | | "inverting" $(\check{a})$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | stable models: 100 | | | stable models: 100 | | | | |
| # | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\|\hat{\hat{a}} - \check{a}\|_F$ | $\|\hat{\hat{a}} - \check{a}\|_F / \|\hat{\hat{a}}\|_F$ |
| 24 | 5 | 1.0823269 | 1.2179537 | 6 | 1.0798311 | 1.2184973 | 196.3141 | 96.40112 |
| 76 | 6 | 0.8112295 | 0.8688194 | 6 | 0.8112295 | 0.8688194 | 0.0000 | 0.0000 |

$r = 3, q = 1, p = 2, T_{in} = 150, T_{out} = 100$, using the "OLS" estimates for the covariances:

| | min. norm $(\hat{\hat{a}})$ | | | min. norm modified $(\bar{a})$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | stable models: 100 | | | stable models: 100 | | | | |
| # | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\hat{s}$ | $\|\hat{\Sigma}_{\nu,in}\|_F$ | $\|\hat{\Sigma}_{\nu,out}\|_F$ | $\|\hat{\hat{a}} - \bar{a}\|_F$ | $\|\hat{\hat{a}} - \bar{a}\|_F / \|\hat{\hat{a}}\|_F$ |
| 24 | 5 | 1.0823269 | 1.2179537 | 5 | 1.0823269 | 1.2179537 | 3.863031e-03 | 5.048549e-04 |
| 76 | 6 | 0.8112295 | 0.8688194 | 6 | 0.8112295 | 0.8688194 | 1.711411e-09 | 7.152293e-11 |

The following table summarizes the relative distances of the coefficient matrices of the 6 estimation procedures, where the $[i,j]$ entry is the mean of the 100 values of $\|a_i - a_j\|_F / \|a_i\|_F$ and $a_i$ stands for the estimation procedure corresponding to the row (column) name.

| | "YW" $\hat{\hat{a}}$ | "YW" $\check{a}$ | "YW" $\bar{a}$ | "OLS" $\hat{\hat{a}}$ | "OLS" $\check{a}$ | "OLS" $\bar{a}$ |
|---|---|---|---|---|---|---|
| "YW" $\hat{\hat{a}}$ | 0.000000000 | 0.4153833 | 0.005088675 | 6.1772326157 | 39.14128 | 6.1772141251 |
| "YW" $\check{a}$ | 0.177130922 | 0.0000000 | 0.177211369 | 6.2117941034 | 26.25715 | 6.2117761834 |
| "YW" $\bar{a}$ | 0.005099621 | 0.4158970 | 0.000000000 | 6.1781671285 | 39.14076 | 6.1781492104 |
| "OLS" $\hat{\hat{a}}$ | 0.899327859 | 1.2114222 | 0.900065710 | 0.0000000000 | 23.13627 | 0.0001211652 |
| "OLS" $\check{a}$ | 1.014572210 | 1.0157145 | 1.014571321 | 0.2396362768 | 0.00000 | 0.2396362775 |
| "OLS" $\bar{a}$ | 0.899344008 | 1.2114457 | 0.900082485 | 0.0001212050 | 23.13672 | 0.0000000000 |

The statistics show that the estimated coefficient matrices differ quite a lot, with one exception (and that is not surprising): The minimum norm solution and the modified minimum norm solution are almost the same if the same estimator for the covariances is used. Especially the "brutal" method (using all eigenvalues which are numerically not zero) gives coefficient matrices which definitely do not coincide with the estimators of the other methods. Nevertheless the fit of all methods is (almost) the same and all polynomials are stable. For the practical point of view there is no winner.

## 5.3 Order Estimation

We have already discussed the problem of choosing a solution of the Yule-Walker equations, and have seen that the minimum norm solution is a reasonable candidate. Nevertheless we have always assumed that we know the lag (order) of the underlying autoregressive system. To be more precisely we have always assumed to know the minimum lag of the underlying system. In the theory of regular AR systems we do not face this problem as the lag of the AR system is uniquely defined by the highest index of the last coefficient matrix unequal to zero, i.e. if an AR system of higher order is used all coefficient matrices with higher index are zero. This property does not hold in the singular case as the AR polynomial is not unique and the order of the system can be chosen arbitrarily large (without adding only zero matrices as opposed to the regular case). This leads us to the following definition.

**Definition 5.3.1. (minimal order)** *The **minimal order** of a singular autoregressive system $[a(z), b]$ is the smallest order $p$ $(a(z) = I - a_1 z - ... - a_p z^p)$ of all observationally equivalent systems $[\bar{a}(z), b]$.*

We know already that for fixed $p$ all solutions of the Yule-Walker equations result in the same covariance matrix of the errors $\Sigma_\nu^p$. Of course this implies, if $p$ denotes the minimal order, that for $\tilde{p} > p$ the population covariance matrices fulfill $\Sigma_\nu^{\tilde{p}} = \Sigma_\nu^p$. As the sample covariance matrix $\hat{\Sigma}_\nu^{\tilde{p}} = \sum_{t=1}^{T-\tilde{p}} \hat{\nu}_t^{\tilde{p}} \hat{\nu}_t^{\tilde{p}\prime}$, where $\hat{\nu}_t^{\tilde{p}} = f_t - \hat{a}_1 f_{t-1} - ... - \hat{a}_{\tilde{p}} f_{t-\tilde{p}}$, is a consistent estimator of $\Sigma_\nu^{\tilde{p}}$, an Information Criterion,

analogously to the regular case, such as AIC or BIC, seems to be reasonable.

$$AIC(p) = \log(\det(\hat{\Sigma}_\nu^p)) + p * r^2 * 2/T$$

$$BIC(p) = \log(\det(\hat{\Sigma}_\nu^p)) + p * r^2 * \log(T)/T$$

where $r$ is the dimension of the process. Unfortunately things are more complicated as AIC and BIC use the logarithm of the determinant of the estimated covariance of the errors. If it happens, that the theoretical covariance is singular from the minimal order onwards only, everything is fine as $AIC(p) = -\infty$ would still select the correct minimal $p$ (if one would decide that the order is the first value where the Information Criteria is minus infinity). But the following example shows that this is not always fulfilled.

**Example 5.3.2.** *Let*

$$f_t = \begin{bmatrix} f_{t,1} \\ f_{t,2} \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \end{bmatrix} + \begin{bmatrix} -1/4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_{t-2,1} \\ f_{t-2,2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t,1} \\ 0 \end{bmatrix}$$

*be a stable AR(2) process, where $f_{t,2} = f_{t-1,1}$. Thus projecting $f_t$ onto $f_{t-1}$ gives already a perfect description of $f_{t,2}$ and thus the corresponding error is zero, which implies that the covariance matrix of the error is already singular. Consequently $AIC(1) = -\infty$ and the order will be most likely underestimated. Note that the order can still be estimated correctly as the sample estimator of the covariance matrix of the order does not have to be singular, but of course consistency cannot be shown.*

This example is on the one hand a dumb example, because setting up a multivariate AR model for a process whose first component is a univariate AR process and its second component is the same process lagged by one, is not what AR models are made for. On the other hand it shows how easily such a problem can occur if miss handled. The next example will show that the example above is not the only possibility where this phenomenon can occur.

**Example 5.3.3.** *Let*

$$f_t = \begin{bmatrix} f_{t,1} \\ f_{t,2} \\ f_{t,3} \end{bmatrix} = \begin{bmatrix} 1/2 & -1/4 & 1/4 \\ 1/4 & 1/4 & 1/8 \\ 0 & 1/2 & 1/4 \end{bmatrix} \begin{bmatrix} f_{t-1,1} \\ f_{t-1,2} \\ f_{t-1,3} \end{bmatrix} + \begin{bmatrix} 1/8 & -1/8 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_{t-2,1} \\ f_{t-2,2} \\ f_{t-2,3} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \varepsilon_t$$

*be a stable AR(2) process. Obviously the best predictor for the components two and three is already available if only one lag is used. Thus the covariance matrix of the error, fitting an AR(1), has already the correct entries except at the $[1,1]$ position. Thus it is singular and the vector $[0,1,-1]'$ is in its kernel. That means that the linear combination $f_{t,2} - f_{t,3}$ can be perfectly explained by $f_{t-1}$.*

Therefore we need an Information Criteria that can handle the singularity of the covariance matrix of the errors. A possibility is to measure the covariance matrix with a norm instead of the determinant.

We will see below that the Frobenius norm is a possible candidate.

Before we present the Information Criteria we will extend the results of [Whittle, 1963], where a recursive algorithm for the coefficient matrices of the AR polynomial $a(z)$ is presented, to the singular case.

### 5.3.1 Recursive Algorithm

The recursive algorithm presented here shows how the coefficient matrices of certain solutions of the Yule-Walker equations are related if the order increases. For this reason we have to introduce an additional index indicating how many lags are used

$$f_t = a_1^p f_{t-1} + ... + a_p^p f_{t-p} + \nu_t^p \tag{5.18}$$

Let $[a_1^{p*}, ..., a_p^{p*}]$ be a solution of the Yule-Walker equations

$$[a_1^p, ..., a_p^p]\Gamma_p = [\gamma_1, ..., \gamma_p] \tag{5.19}$$

such that

$$\sum_{k=0}^{p} d_k^{p*}\gamma_{j-k} = 0, \quad j = 1, ..., p \tag{5.20}$$

with $d_0^{p*} = I$ and $d_k^{p*} = -a_k^{p*}, k = 1, ..., p$. Furthermore we will use the solutions of the equations

$$[\tilde{\alpha}_1^p, ..., \tilde{\alpha}_p^p] \underbrace{\mathbb{E}\begin{bmatrix} x_{t-p} \\ \vdots \\ x_{t-1} \end{bmatrix} [x'_{t-p}, \dots, x'_{t-1}]}_{=:\bar{\Gamma}(p)} = [\gamma'_1, ..., \gamma'_p] \tag{5.21}$$

Such that for a solution $[\tilde{\alpha}_1^{p\triangle}, ..., \tilde{\alpha}_p^{p\triangle}]$ of (5.21) we have

$$\sum_{k=0}^{p} \alpha_k^{p\triangle}\gamma_{k-j} = 0, \quad j = 1, ..., p \tag{5.22}$$

where $\alpha_0^{p\triangle} = I$ and $\alpha_k^{p*} = -\tilde{\alpha}_k^{p*}, k = 1, ..., p$.

Next we define the matrices

$$\begin{aligned}
\Sigma_p^* &= \sum_{k=0}^{p} d_k^{p*}\gamma_{-k} \\
\bar{\Sigma}_p^{\triangle} &= \sum_{k=0}^{p} \alpha_k^{p\triangle}\gamma_k \\
\delta_p^* &= \sum_{k=0}^{p} d_k^{p*}\gamma_{p+1-k} \\
\bar{\delta}_p^{\triangle} &= \sum_{k=0}^{p} \alpha_k^{p\triangle}\gamma_{k-p-1}
\end{aligned} \tag{5.23}$$

**Lemma 5.3.4.** *Let $f_t$ have an AR($p_0$) representation $f_t = a_1 f_{t-1} + ... + a_{p_0} f_{t-p_0} + \nu_t$, then the following equalities hold, where the corresponding matrices are defined as above.*

*(i)* $\Sigma_p = \Sigma_p^* = \Sigma_p^{\triangledown}$

*(ii)* $\bar{\Sigma}_p = \bar{\Sigma}_p^{\triangle} = \bar{\Sigma}_p^{\diamondsuit}$

*(iii)* $\delta_p = \delta_p^* = \delta_p^{\triangledown}$

*(iv)* $\bar{\delta}_p = \bar{\delta}_p^{\triangle} = \bar{\delta}_p^{\diamondsuit}$

*(v)* $\delta_p = \bar{\delta}_p'$

*(vi)* $d_k^{p+1*} = d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}, k = 1, ..., p$

*(vii)* $\alpha_k^{p+1\triangle} = \alpha_k^{p\diamondsuit} + \alpha_{p+1}^{p+1\triangle} d_{p-k+1}^{p*}, k = 1, ..., p$

*(viii)* $d_{p+1}^{p+1*}\bar{\Sigma}_p = -\delta_p, \quad \alpha_{p+1}^{p+1\triangle}\Sigma_p = -\bar{\delta}_p$

*(ix)* $\Sigma_{p+1} = [I - d_{p+1}^{p+1*}\alpha_{p+1}^{p+1\triangle}]\Sigma_p$

*(x)* $\delta_p = \sum_{j=0}^{p}\sum_{l=0}^{p} d_j^{p*}\gamma_{p+1-j-l}\alpha_l^{p\triangle'}$

*Proof.* ***(Lemma 5.3.4.)***
(i)

$$\begin{aligned}
\Sigma_p^* &= \sum_{k=0}^{p} d_k^{p*}\gamma_{-k} = \gamma_0 + [d_1^{p*}, ..., d_p^{p*}][\gamma_1, ..., \gamma_p]' \\
&= \gamma_0 - [a_1^{p*}, ..., a_p^{p*}]\mathbb{E}[[x_{t-1}', ..., x_{t-p}']' x_t'] = \gamma_0 - \mathbb{E}[\sum_{k=1}^{p} a_k^{p*} x_{t-k} x_t'] \\
&= \gamma_0 - \mathbb{E}[\hat{x}_{t|x_{t-1},...,x_{t-p}} x_t'] = \gamma_0 - \mathbb{E}[\sum_{k=1}^{p} a_k^{p\triangledown} x_{t-k} x_t'] = \Sigma_p^{\triangledown}
\end{aligned}$$

where $\hat{x}_{t|x_{t-1},...,x_{t-p}}$ denotes the projection of $x_t$ onto the space spanned by the components of $\{x_{t-1}, ..., x_{t-p}\}$
(ii) the same as (i)
(iii)

$$\begin{aligned}\delta_p^* &= \sum_{k=0}^{p} d_k^{p*}\gamma_{p+1-k} = \gamma_{p+1} + [d_1^{p*}, ..., d_p^{p*}][\gamma_p', ..., \gamma_1']' \\ &= \gamma_{p+1} - [a_1^{p*}, ..., a_p^{p*}]\mathbb{E}[[x_{t-1}', ..., x_{t-p}']' x_{t-p-1}'] = \gamma_{p+1} - \mathbb{E}[\sum_{k=1}^{p} a_k^{p*} x_{t-k} x_{t-p-1}'] \\ &= \gamma_{p+1} - \mathbb{E}[\hat{x}_{t|x_{t-1},...,x_{t-p}} x_{t-p-1}'] = \gamma_{p+1} - \mathbb{E}[\sum_{k=1}^{p} a_k^{p\triangledown} x_{t-k} x_{t-p-1}'] = \delta_p^{\triangledown}\end{aligned}$$

(iv) the same as (iii)
(v) Define $\varepsilon_t^p$ as the remainder after subtracting the projection $\hat{x}_{t|x_{t-1},...,x_{t-p}}$ from $x_t$ and $\eta_t^p = x_t - \hat{x}_{t|x_{t+1},...,x_{t+p}}$. So we have

$$\delta_p := \sum_{k=0}^{p} d_k^p \gamma_{p+1-k} = \mathbb{E}[[\sum_{k=0}^{p} d_k^p x_{t-k}] x_{t-p-1}'] = \mathbb{E}[\varepsilon_t^p x_{t-p-1}']$$

and

$$\bar{\delta}_p := \sum_{k=0}^{p} \alpha_k^p \gamma_{k-p-1} = \mathbb{E}[[\sum_{k=0}^{p} \alpha_k^p x_{t+k}] x_{t+p+1}'] = \mathbb{E}[\eta_t^p x_{t+p+1}']$$

and

$$\begin{aligned}\mathbb{E}[\varepsilon_t^p x_{t-p-1}'] &= \mathbb{E}[\varepsilon_t^p [\eta_{t-p-1}^p - \alpha_1^p x_{t-p} - ... - \alpha_p^p x_{t-1}']'] = \mathbb{E}[\varepsilon_t^p \eta_{t-p-1}^{p'}] \\ &= \mathbb{E}[\eta_{t-p-1}^p \varepsilon_t^{p'}]' = \mathbb{E}[\eta_t^p \varepsilon_{t+p+1}^{p'}]' = \mathbb{E}[\eta_t^p x_{t+p+1}']'\end{aligned}$$

gives the disered result.
(vi)

$$\sum_{k=0}^{p+1} d_k^{p+1*}\gamma_{j-k} = \sum_{k=1}^{p} d_k^{p+1*}\gamma_{j-k} + \gamma_j + d_{p+1}^{p+1*}\gamma_{j-p-1} = 0.$$

So if

$$\sum_{k=1}^{p} [d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangledown}]\gamma_{j-k} + \gamma_j + d_{p+1}^{p+1*}\gamma_{j-p-1} = 0$$

holds for all $j = 1, ..., p$ the representation $d_k^{p+1*} = d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}, k = 1, ..., p$ is obviously valid.

$$\begin{aligned}
&\sum_{k=1}^{p}[d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}]\gamma_{j-k} + \gamma_j + d_{p+1}^{p+1*}\gamma_{j-p-1} \\
&= \underbrace{\sum_{k=0}^{p} d_k^{p\triangledown}\gamma_{j-k}}_{=0} + d_{p+1}^{p+1*}[\sum_{k=1}^{p}\alpha_{p-k+1}^{p\triangle}\gamma_{j-k} + \gamma_{j-p-1}] \\
&\overset{\tilde{k}=p-k+1}{=} d_{p+1}^{p+1*}[\sum_{\tilde{k}=1}^{p}\alpha_{\tilde{k}}^{p\triangle}\gamma_{j-(p-\tilde{k}+1)} + \gamma_{j-p-1}] \\
&= d_{p+1}^{p+1*}[\sum_{\tilde{k}=0}^{p}\alpha_{\tilde{k}}^{p\triangle}\gamma_{j-p-1+\tilde{k}}] = d_{p+1}^{p+1*}[\underbrace{\sum_{\tilde{k}=0}^{p}\alpha_{\tilde{k}}^{p\triangle}\gamma_{\tilde{k}-j}}_{=0}] = 0
\end{aligned}$$

(vii) the same as (vi)

(viii)

$$\begin{aligned}
0 &= \sum_{k=0}^{p+1} d_k^{p+1*}\gamma_{p+1-k} = \sum_{k=1}^{p} d_k^{p+1*}\gamma_{p+1-k} + \gamma_{p+1} + d_{p+1}^{p+1*}\gamma_0 \\
&\overset{(vi)}{=} \sum_{k=1}^{p}[d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}]\gamma_{p+1-k} + \gamma_{p+1} + d_{p+1}^{p+1*}\gamma_0 \\
&= \sum_{k=0}^{p} d_k^{p\triangledown}\gamma_{p+1-k} + d_{p+1}^{p+1*}[\sum_{k=1}^{p}\alpha_{p-k+1}^{p\triangle}\gamma_{p+1-k} + \gamma_0] \\
&= \delta_p + d_{p+1}^{p+1*}\sum_{k=1}^{p+1}\alpha_{p-k+1}^{p\triangle}\gamma_{p+1-k} \overset{\tilde{k}=p+1-k}{=} \delta_p + d_{p+1}^{p+1*}\underbrace{\sum_{k=0}^{p}\alpha_{\tilde{k}}^{p\triangle}\gamma_{\tilde{k}}}_{=\bar{\Sigma}_p}
\end{aligned}$$

The other equality can be shown in a similar way.

(ix)

$$\begin{aligned}
\Sigma_{p+1} &= \sum_{k=0}^{p+1} d_k^{p+1*}\gamma_{-k} = \gamma_0 + \sum_{k=1}^{p} d_k^{p+1*}\gamma_{-k} + d_{p+1}^{p+1*}\gamma_{-p-1} \\
&\overset{(vi)}{=} \gamma_0 + \sum_{k=1}^{p}[d_k^{p\triangledown} + d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}]\gamma_{-k} + d_{p+1}^{p+1*}\gamma_{-p-1} \\
&= \Sigma_p + \sum_{k=1}^{p} d_{p+1}^{p+1*}\alpha_{p-k+1}^{p\triangle}\gamma_{-k} + d_{p+1}^{p+1*}\gamma_{-p-1} \\
&= \Sigma_p + d_{p+1}^{p+1*}\sum_{k=1}^{p+1}\alpha_{p-k+1}^{p\triangle}\gamma_{-k} \overset{\tilde{k}=p+1-k}{=} \Sigma_p + d_{p+1}^{p+1*}\underbrace{\sum_{k=0}^{p}\alpha_{\tilde{k}}^{p\triangle}\gamma_{-p-1+\tilde{k}}}_{\bar{\delta}_p} \\
&\overset{(viii)}{=} \Sigma_p - d_{p+1}^{p+1*}\alpha_{p+1}^{p+1\triangle}\Sigma_p
\end{aligned}$$

(x)

$$\begin{aligned}
\delta_p &= \sum_{k=0}^{p} d_k^{p*}\gamma_{p+1-k} = \gamma_{p+1} + \sum_{k=1}^{p} d_k^{p*}\gamma_{p+1-k} \\
&\overset{(5.22)}{=} \gamma_{p+1} + \sum_{k=1}^{p} d_k^{p*}\left[-\sum_{l=1}^{p}\gamma_{p+1-k-l}\alpha_l^{p\triangle\prime}\right] \\
&= \gamma_{p+1} - \sum_{k=1}^{p}\sum_{l=1}^{p} d_k^{p*}\gamma_{p+1-k-l}\alpha_l^{p\triangle\prime} \quad \overset{\bar{\delta}_p'=\gamma_{p+1}+\sum_{l=1}^{p}\gamma_{p+1-l}\alpha_l^{p\triangle\prime}}{=} \\
&= \bar{\delta}_p' - \sum_{l=1}^{p}\gamma_{p+1-l}\alpha_l^{p\triangle\prime} - \sum_{k=1}^{p}\sum_{l=1}^{p} d_k^{p*}\gamma_{p+1-k-l}\alpha_l^{p\triangle\prime} \\
&= \delta_p - \underbrace{\sum_{k=0}^{p}\sum_{l=1}^{p} d_k^{p*}\gamma_{p+1-k-l}\alpha_l^{p\triangle\prime}}_{\Rightarrow=0}
\end{aligned}$$

Therefore

$$\begin{aligned}
&\sum_{k=0}^{p}\sum_{l=0}^{p} d_k^{p*}\gamma_{p+1-k-l}\alpha_l^{p\triangle'} \\
&= \sum_{k=0}^{p} d_k^{p*}\left[\gamma_{p+1-k}\alpha_0^{p\triangle'} + \sum_{l=1}^{p}\gamma_{p+1-k-l}\alpha_l^{p\triangle'}\right] \\
&= \underbrace{\sum_{k=0}^{p} d_k^{p*}\gamma_{p+1-k}}_{\delta_p} + \underbrace{\sum_{k=0}^{p}\sum_{l=1}^{p} d_k^{p*}\gamma_{p+1-k-l}\alpha_l^{p\triangle'}}_{=0} = \delta_p
\end{aligned}$$

□

### 5.3.2 A Consistent Estimator

In this section we want to show (strong) consistency of the following Information Criteria using the Frobenius norm of the estimated covariance matrix of the innovations.

$$IC(k) = \|\hat{\Sigma}_k\|_F + k * c * (2 * \log(\log(T))/T)^2) \tag{5.24}$$

where $T$ denotes the sample size and $c > 0$ ($r^4$ suggested, where $r$ is the dimension of the AR process).

We commence from an AR process $f_t$ which has a stable singular system

$$f_t = a_1 f_{t-1} + ... + a_p f_{t-p} + \nu_t \tag{5.25}$$

and denote by $\hat{\gamma}_j^T = 1/T\sum_{t=1}^{T-j} f_t f_{t-j}'$ the sample covariance estimator of $\gamma_j = \mathbb{E}f_t f_{t-j}'$ and by $\hat{\Gamma}_p^T = \begin{bmatrix} \hat{\gamma}_0^T & \hat{\gamma}_1^T & \cdots & \hat{\gamma}_{p-1}^T \\ \hat{\gamma}_1^{T\prime} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\gamma}_1^T \\ \hat{\gamma}_{p-1}^{T\prime} & \cdots & \hat{\gamma}_1^{T\prime} & \hat{\gamma}_0^T \end{bmatrix}$.

**Theorem 5.3.5.** *Let $f_t$ be generated by* (5.25) *and with minimal order $p_0$. Then*

$$\|\hat{\hat{a}}_p\hat{\alpha}_p\hat{\Sigma}_{\nu,p-1}\|_F = O(T^{-2}(2\log(\log(T)))^2), p > p_0$$

*where $\hat{\hat{a}}_p$ is the last coefficient of the minimum norm solution* (5.6) *and $\hat{\alpha}_p$ is the minimum norm solution estimator for the coefficient $\alpha_p$ of* (5.21). *$X = O(c(T))$ means that $X/c(T)$ is bounded almost surely.*

*Proof.* ***(Theorem 5.3.5.)***

Note that the results of Lemma 5.3.4 are still valid if the sample covariances and the corresponding estimated coefficient matrices are used.
Define $\hat{B}_p = \hat{\hat{a}}_p \hat{\hat{\Sigma}}^{1/2}_{\nu,p-1}$. Note that $\hat{\hat{\Sigma}}^{1/2}_{\nu,p-1} = \hat{b}$ where $\hat{\hat{\Sigma}}_{\nu,p-1} = [\hat{b}, 0] \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{b}' \\ 0 \end{bmatrix}$ and thus $\hat{\hat{\Sigma}}^{1/2}_{\nu,p-1} = \hat{b}\hat{\Sigma}^{1/2}$ might not be square. Of course $\hat{\hat{\Sigma}}^{-1/2}_{\nu,p-1}$ is then defined by inverting the non zero eigenvalues only. Thus $\hat{B}_p\hat{B}'_p = \hat{\hat{a}}_p \hat{\hat{\Sigma}}^{1/2}_{\nu,p-1} \hat{\hat{\Sigma}}^{1/2}_{\nu,p-1} \hat{\hat{a}}_p = \hat{\hat{a}}_p \hat{\hat{\Sigma}}_{\nu,p-1} \hat{\hat{a}}_p = \hat{\hat{a}}_p \hat{\alpha}_p \hat{\Sigma}_{\nu,p-1}$ holds by Lemma 5.3.4 (v) and (viii) and this implies that $\|\hat{\hat{a}}_{p+1}\hat{\alpha}_{p+1}\hat{\Sigma}_{\nu,p}\|_F = \|\hat{B}_p\hat{B}'_p\|_F = \sum_{i=1}^n \lambda_i^2$, where $\lambda_i$ is the i-th biggest eigenvalue of $\hat{B}_p\hat{B}'_p$.
As Quinn has shown in [Quinn, 1980] that $\hat{B}_p = O(T^{-1/2}(2\log(\log(T)))^{1/2})$ and has its limit points in the interval $[0, 1]$ ($\hat{B}_p = -\hat{\delta}_p \hat{\hat{\Sigma}}^{-1/2}_{\nu,p-1}$), it follows that $\sum_{i=1}^n \lambda_i^2 = tr(\hat{B}_p\hat{B}'_p\hat{B}_p\hat{B}'_p) = O(T^{-2}(2\log(\log(T)))^2)$ and thus $\limsup T^2(2\log(\log(T))))^{-2} tr(\hat{B}_p\hat{B}'_p\hat{B}_p\hat{B}'_p) = 1$ (see [Hannan, 1980]). □

**Corollary 5.3.6.** *Let $\hat{p}$ be chosen to minimize* (5.24) *over $p = 1, ..., M$ with $M \geq p_0$. Then $\hat{p} \to p_0, a.s.$ if and only if $c > 0$ holds.*

*Proof.* ***(Corollary 5.3.6.)***
$IC(p) - IC(p-1) = \|\hat{\Sigma}_{\nu,p}\| - \|\hat{\Sigma}_{\nu,p-1}\| + c * (2 * \log(\log(T))/T)^2)$
Consider the case $p \leq p_0$: $\|\hat{\Sigma}_{\nu,p}\|$ converges a.s. to $\|\Sigma_{\nu,p}\|$, $\|\hat{\Sigma_{\nu p}}\|$ is a non increasing function in $p$ and the inequality $\|\Sigma_{\nu,p}\| < \|\Sigma_{\nu,p-1}\|$ certainly holds for $p = p_0$. Thus there exists a $T_0$ such that $\|\hat{\Sigma_{\nu p}}\| < \|\hat{\Sigma}_{\nu,p-1}\|$ $a.s.$ holds $\forall T \geq T_0$ and therefore $IC(p_0) < IC(p), p < p_0, \forall T \geq T_0,\ a.s.$.
Next consider the case $p > p_0$:

$$\|\hat{\Sigma_{\nu p}}\| - \|\hat{\Sigma}_{\nu,p-1}\| = \|(I - \hat{A}^p_p\hat{\alpha}^p_p)\hat{\Sigma}_{\nu,p-1}\| - \|\hat{\Sigma}_{\nu,p-1}\| \geq \left|\|\hat{\Sigma}_{\nu,p-1}\| - \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\|\right| - \|\hat{\Sigma}_{\nu,p-1}\|$$

1. $\|\hat{\Sigma}_{\nu,p-1}\| \geq \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| \Rightarrow \|\hat{\Sigma_{\nu p}}\| - \|\hat{\Sigma}_{\nu,p-1}\| \geq -\|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\|$

2. $$\begin{aligned}\|\hat{\Sigma}_{\nu,p-1}\| \leq \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| \Rightarrow &\|\hat{\Sigma_{\nu p}}\| - \|\hat{\Sigma}_{\nu,p-1}\| \geq \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| - 2\|\hat{\Sigma}_{\nu,p-1}\| \\ &\geq \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| - 2\|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| = -\|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\|\end{aligned}$$

$\Rightarrow \left|\|\hat{\Sigma_{\nu p}}\| - \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\|\right| - \|\hat{\Sigma}_{\nu,p-1}\| \geq -\|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\|$. From Theorem 5.3.5 we know that there exists a $T_0$ such that $c * (2 * \log(\log(T))/T)^2) - \|\hat{A}^p_p\hat{\alpha}^p_p\hat{\Sigma}_{\nu,p-1}\| \geq 0\ a.s.\ \forall T \geq T_0$. □

Note that the proofs of Theorem 5.3.5 and Corollary 5.3.6 can be adopted also for the situation where only weak consistent estimators of $\gamma_j$ are available. In this case the $\|\hat{\hat{a}}_p\hat{\alpha}_p\hat{\Sigma}_{\nu,p-1}\|_F$ will only converge in probability, i.e. $\|\hat{\hat{a}}_p\hat{\alpha}_p\hat{\Sigma}_{\nu,p-1}\|_F = O_p(T^{-2}(2\log(\log(T)))^2), p > p_0$, where $X = O_p(c(T))$ means

that $plim_{T\to\infty} X/c(T) = M < \infty$.

Unfortunately the estimation is "only" consistent, that means it is nice to have this theoretical result but the practical use of this Information Criteria is not even worth mentioning. Maybe it can be upgraded in the future.

### 5.3.3 An Estimator for the Practice

The Information Criterion presented here consists of two criteria. The first one is similar to the AIC, with the difference, that the determinant of a $(q \times q)$ matrix (where $q$ is the dimension of the driving white noise) is calculated. The second one estimates the dimension of the one-step-ahead prediction errors for each lag. As this sequence is non increasing, the first order which selects the minimal dimension is selected. Finally the order is estimated by taking the maximum of the orders selected by these two criteria.

#### First Criterion

Let $M$ be a fixed integer with $M \geq p$, where $p$ is the true minimal order of the system. Then $\Sigma_{\nu,M} = \Sigma_{\nu,p}$, where $\Sigma_{\nu,p} = bb'$ is the covariance matrix of the one-step-ahead prediction errors. As $\Sigma_{\nu,k} = bb' + \Omega_k$, where $k$ is arbitrary, the rank of $\Sigma_{\nu,k}$ is larger than or equal to $q$. Furthermore $b'\Sigma_{\nu,k}b$ is regular. Thus the first Information Criterion is

$$IC_1(k) = \det(\hat{b}'\hat{\Sigma}_{\nu,k}\hat{b}) + kr^2 2/T \tag{5.26}$$

with $\hat{\Sigma}_{\nu,M} = \hat{b}\hat{b}'$, and the order is estimated by minimizing (5.26) for $k = 0, ..., M$. Unfortunately we cannot exclude the possibility that $b'\Omega_k b = 0$ holds for some $k < p$, and thus the order might be underestimated systematically. Therefore the second criterion is needed.

#### Second Criterion

We have already pointed out in section 5.1 that the "OLS" estimate (5.5) of the covariance matrix $\Gamma_p$ has the same kernel as the true one and thus the rank of $\Gamma_p$ can be estimated by counting the eigenvalues which are larger than a (numerically) reasonable boundary, e.g. $10^{-6}$. Thus, the rank $q_k$ of the covariance matrix, of the one-step-ahead prediction errors which is obtained by fitting an AR(k) model, can be estimated by $\hat{s}_{k+1} - \hat{s}_k$ where $\hat{s}_k = \mathrm{rk}(\hat{\Gamma}_k)$. Hence, $q_k = q_p$ for $k \geq p$, but the equality can also hold for $k < p$ (see e.g. example 5.3.2). Nevertheless, the criteria selects the smallest order which minimizes $\hat{q}_k$.

**Final Criterion**

We have shown, that none of the two criteria introduced above can estimate the true order. But the maximum of the orders selected by the two criteria is an estimate of the minimal order which can be seen as follows: Note that if $b'\Omega_k b = 0$ holds for some $k < p$ it follows that $q_k > q_p$. On the other hand if $q_k = q_p$ for some $k < p$, $b'\Omega_k b \neq 0$ holds.

Consistency of this estimation procedure might be shown in the future. We will see below that this criterion works considerably well.

### 5.3.4 Comparison of Order Estimates

Here we want to compare the AIC, BIC and the estimator presented in the previous section. We fill fix the sample size to $T = 150$. For the estimation of the coefficient matrices we will use the minimum norm solution presented in section 5.1. As the rank of the "OLS" sample covariance estimate is the same as the rank of the population covariance matrix we will determine all ranks which have to be used by counting the eigenvalues which are of modulus larger than or equal to $10^{-6}$.

We will estimate the covariance matrices with what we call the "Yule-Walker" (compare to (5.2)) and the "OLS" methods (compare to (5.3)).

"$AICq$" and "$q$" denote the "first" and "second" criterion described above. Consequently "combine" denotes the "final" criterion.

$r = 3, q = 3, 100$ simulations:

| "Yule-Walker" estimates | | | | | | "OLS" estimates | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | AIC | BIC | q | AICq | combine | | AIC | BIC | q | AICq | combine |
| $\hat{p} = 0$ | 0 | 0 | 93 | 0 | 0 | $\hat{p} = 0$ | 0 | 0 | 93 | 23 | 20 |
| $\hat{p} = 1$ | 27 | 62 | 2 | 27 | 24 | $\hat{p} = 1$ | 30 | 50 | 2 | 13 | 13 |
| $\hat{p} = 2$ | 66 | 38 | 3 | 66 | 67 | $\hat{p} = 2$ | 50 | 48 | 3 | 44 | 45 |
| $\hat{p} = 3$ | 7 | 0 | 2 | 7 | 9 | $\hat{p} = 3$ | 5 | 0 | 2 | 5 | 7 |
| $\hat{p} > 3$ | 0 | 0 | 0 | 0 | 0 | $\hat{p} > 3$ | 15 | 2 | 0 | 15 | 15 |

$r = 3, q = 1, 100$ simulations:

| | "Yule-Walker" estimates | | | | | | "OLS" estimates | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | AIC | BIC | q | AICq | combine | | AIC | BIC | q | AICq | combine |
| $p = 0$ | 0 | 0 | 0 | 4 | 0 | $p = 0$ | 0 | 0 | 0 | 4 | 0 |
| $p = 1$ | 40 | 54 | 5 | 92 | 5 | $p = 1$ | 68 | 68 | 5 | 90 | 5 |
| $p = 2$ | 45 | 36 | 74 | 4 | 74 | $p = 2$ | 26 | 26 | 74 | 6 | 74 |
| $p = 3$ | 8 | 4 | 21 | 0 | 21 | $p = 3$ | 2 | 2 | 21 | 0 | 21 |
| $p > 3$ | 7 | 6 | 0 | 0 | 0 | $p > 3$ | 4 | 4 | 0 | 0 | 0 |

**Summary**

In the case of a regular AR process the "final" criterion has a similar performance as the AIC in the case of "Yule-Walker" estimates of the covariances. In the case of "OLS" estimates the results of all criteria are bad but AIC is still best.
In the singular case the "final" criterion is the clear winner.
Summarizing we can say that the "final" criterion performs almost as well as the AIC in the regular case. In the singular case it outperforms the other estimation procedures and is indeed a good estimation procedure.

# Chapter 6

# Structure Theory of GDFMs with Oscillations

In chapter 2 the latent variables were assumed to be linearly regular stochastic processes. In our conference paper [Deistler et al., 2010b] we discuss what happens if the latent variables have also linearly singular components, to be more precise we allow the singular components to be harmonic processes. We will see that in this case the static factors have linearly singular components too and therefore they will have a singular AR representation (this is only a generic result of course) which has unit roots.

We commence from the observations

$$y_t = ((y_t^N)_{t\in\mathbb{Z}})_{N\in\mathbb{N}}$$

which can be represented by a sum of the latent variables $\chi_t$ and the idiosyncratic noise $u_t$, which are orthogonal to each other at any lead and lag

$$y_{i,t} = \chi_{i,t} + u_{i,t}$$

Now the latent variables $\chi_t$, can be represented by a sum of its linearly regular $\chi_t^r$ and its linearly singular part $\chi_t^s$, which are also orthogonal to each other at any lead and lag (Wold decomposition)

$$\chi_{i,t} = \chi_{i,t}^r + \chi_{i,t}^s$$

The linearly singular part is assumed to be a harmonic process. This leads us to a modified Assumption 1:

**Assumption 1'**

1. $y_t^N$ is a (wide sense) stationary zero mean process for all $N \in \mathbb{N}$. Let $y_t^N = y_t^{N,r} + y_t^{N,s}$ be its

decomposition in its linearly regular and linearly singular part. Then for all $N \in \mathbb{N}$

$$y_t^N = y_t^{N,r} + y_t^{N,s} = \chi_t^{N,r} + \chi_t^{N,s} + u_t^N$$

holds, i.e. $y_t^{N,s} = \chi_t^{N,s}$, which means that only the latent variables have a linearly singular component and thus the idiosyncratic noise is linearly regular.

2. $y_t^{N,r}, \chi_t^{N,r}, u_t^N$ fulfill Assumption 1 and Assumption 2 respectively for all $N \in \mathbb{N}$.

3. $$\chi_t^{N,s} = \sum_{j=1}^{h} C_j^N e^{-i\lambda_j t} r_j \tag{6.1}$$

with $C_j^N \in \mathbb{C}^{N\times 1}$ and the complex valued one dimensional random variables $r_j$ satisfy

- $\mathbb{E}r_j = 0, j = 1, ..., h$
- $\mathbb{E}r_j \bar{r}_l = 0, \forall j \neq l$
- $\lambda_{j+1} = -\lambda_{h-j}$, $r_{j+1} = \bar{r}_{h-j}$ and $C_{j+1}^N = \bar{C}_{h-j}^N$ for $j = 0, 1, \cdots, \lfloor h/2 - 1 \rfloor$
- $|r_j| = 1, j = 1, ..., h$

Note that harmonic processes are in general non ergodic. Since only a single trajectory is observed the randomness may be confined to the phases of $r_j$. Thus the normalization $|r_j| = 1$ can be justified.

**Assumption 5.**
*The non zero eigenvalues of $C_{j+1}^N C_{j+1}^{N*} + C_{h-j}^N C_{h-j}^{N*}$ diverge for $N \to \infty$.*

Note that by Assumption 1' only the latent variables are allowed to have a linearly singular component. As the latent variables are characterized by the projection of the observations on the aggregation space, the random variables $r_j$ have to be in the aggregation space, and thus Assumption 5 is necessary.

Of course the theory of the previous chapters can be used for the linearly regular parts of $y_t$ and $\chi_t$ and thus we recall that we generically have a system for $\chi_t^{N,r}$ of the form

$$f_t^r = a_1 f_{t-1}^r + ... + a_p f_{t-p}^r + b\varepsilon_t \tag{6.2}$$
$$\chi_t^{N,r} = H_1^N f_t^r \tag{6.3}$$

where $H_1^N \in \mathbb{R}^{N\times r}$ has full rank. Obviously, the linearly singular part $\chi_t^{N,s}$ does not have a spectral density as its spectral distribution has jumps at $\lambda_j, j = 1, ..., h$ (and is constant in between).
As the linearly regular and linearly singular parts are orthogonal we have

$$\mathbb{E}\chi_t^N \chi_t^{N'} = \mathbb{E}\chi_t^{N,r}\chi_t^{N,r'} + \mathbb{E}\chi_t^{N,s}\chi_t^{N,s'} = H_1^N \mathbb{E} f_t^r f_t^{r'} H_1^{N'} + \sum_{j=1}^{h} C_j^N \mathbb{E} r_j \bar{r}_j C_j^{N*} = H^N f_t f_t' H^{N'} \tag{6.4}$$

where $f_t$ is a factor process consisting of components and linear combinations of $f_t^r$ and $f_t^s$, where

$$f_t^s = \begin{bmatrix} 1 & 0 & \dots & & & \dots & 0 & 1 \\ i & 0 & \dots & & & \dots & 0 & -i \\ 0 & 1 & 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & i & 0 & & & 0 & -i & 0 \\ & & \ddots & & & \iddots & & \\ & & & 1 & 1 & & & \\ & & & i & -i & & & \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_h \end{bmatrix} \tag{6.5}$$

Now let us discuss the dynamics of the $r$ dimensional process $f_t$. As the dimensions of the processes $f_t^r$ and $f_t^s$ do not have to be equal we decompose

$$f_t = \tilde{f}_t^r + \tilde{f}_t^s \tag{6.6}$$

according to Wold decomposition. This means that the Hilbert space spanned by $f_t$ is the orthogonal sum of the Hilbert spaces spanned by $\tilde{f}_t^r$ and $\tilde{f}_t^s$, i.e. $H_f(t) = \overline{span}\{f_{i,s}|i = 1, ...r, s \leq t\} = H_{\tilde{f}^r}(t) = \overline{span}\{\tilde{f}_{i,s}^r|i = 1, ...r, s \leq t\} \oplus H_{\tilde{f}^s}(t) = \overline{span}\{\tilde{f}_{i,s}^r|i = 1, ...r, s \leq t\}$. Therefore, as $\tilde{f}_t^s$ has an autoregressive representation with zero innovations process, $f_t$ has an autoregressive representation, which can be seen as follows

$$\begin{aligned} f_{t+1}|_{H_f(t)} &= \tilde{f}_{t+1}^r|_{H_f(t)} + \tilde{f}_{t+1}^s|_{H_f(t)} = \tilde{f}_{t+1}^r|_{H_{\tilde{f}^r}(t)\oplus H_{\tilde{f}^s}(t)} + \tilde{f}_{t+1}^s|_{H_{\tilde{f}^r}(t)\oplus H_{\tilde{f}^s}(t)} \\ &= \tilde{f}_{t+1}^r|_{H_{\tilde{f}^r}(t)} + \tilde{f}_{t+1}^r|_{H_{\tilde{f}^s}(t)} + \tilde{f}_{t+1}^s|_{H_{\tilde{f}^r}(t)} + \tilde{f}_{t+1}^s|_{H_{\tilde{f}^s}(t)} \\ &= \tilde{f}_{t+1}^r|_{H_{\tilde{f}^r}(t)} + \tilde{f}_{t+1}^s|_{H_{\tilde{f}^s}(t)} \end{aligned} \tag{6.7}$$

and as the one-step-ahead predictor of $\tilde{f}_{t+1}^r$ and $\tilde{f}_{t+1}^s$, given data until time $t$, only depend on a finite past, the one-step-ahead predictor of $f_{t+1}$ does so too and thus $f_t$ has an autoregressive representation. Note that in the extreme case, when $\mathbb{E}\chi_t\chi_t'$ has rank $r_1 + h$, where $r_1$ denotes the dimension of $f_t^r$, i.e.

$$\mathbb{E}\chi_t^N \chi_t^{N'} = [H_1^N, H_2^N] \begin{bmatrix} \mathbb{E}f_t^r f_t^{r'} & 0 \\ 0 & \mathbb{E}f_t^s f_t^{s'} \end{bmatrix} \begin{bmatrix} H_1^{N'} \\ H_2^{N'} \end{bmatrix} \tag{6.8}$$

holds, where $[H_1^N, H_2^N] \in \mathbb{R}^{N\times(r_1+h)}$ is of full rank, the factor process $f_t = \begin{bmatrix} f_t^r \\ f_t^s \end{bmatrix}$ can be reconstructed by a linear transformation of the latent variables. In this case we get the whole history of $f_t^r$ and $f_t^s$ separately and not a history of only a linear combination of them. Therefore, as $f_t^s$ has an autoregressive representation with zero innovation process, the process $f_t$ has an autoregressive representation of the

form

$$f_t = \begin{bmatrix} f_t^r \\ f_t^s \end{bmatrix} = A_1 f_{t_1} + ... + A_{\tilde{p}} f_{t-\tilde{p}} + \begin{bmatrix} b\varepsilon_t \\ 0 \end{bmatrix} \tag{6.9}$$

where the $A_i \in \mathbb{R}^{(r+h)\times(r+h)}, i = 1, ..., \tilde{p}$ are block diagonal. Note, that in general $\mathrm{rk}(\mathbb{E}\chi_t\chi_t') = r_1 + h$ will not hold and thus the AR system will not have the block structure. Nevertheless we always result in a system of the form

$$f_t = a_1 f_{t-1} + ... + a_p f_{t-p} + \nu_t \tag{6.10}$$

$$\chi_t^N = H^N f_t \tag{6.11}$$

where the polynomial $a(z) = \sum_{j=1}^{p} a_j z^j$ has at least one unit root.

# Chapter 7

# Mixed Frequency

Traditional (multivariate) time series analysis is designed for time series which appear at each time point simultaneously. We call that case the single frequency case, which means that all time series appear at the same (sampling) frequency. Of course, the previous chapters follow exactly this assumption.

In this chapter we are concerned with the case where this assumption is no longer valid. To be more precise we consider the case of different sampling frequencies and call that case the mixed frequency case. An example for mixed frequency is a macro economic model, where the GDP and the employment rate are used. Whereas the GDP appears only every third month, the number of unemployed people is published every month.

One way to handle this situation is a very brutal one, namely to design a model for the data on the slowest frequency only. All data which appears more often is thrown away (or summed up if the variable is a flow variable). We will not discuss this approach in more detail as we want to use as much information as possible.

The traditional approach commences from a model for the data at the highest frequency. That means although not all data is available (observable), a theoretic model at the highest frequency is assumed and is approximated by the data that is available.

Of course a model at the highest frequency is desirable, as everything can be done with such a model (if available), such as forecasting and nowcasting. Of course there is a huge literature concerned with this topic e.g. [Giannone et al., 2008], [Chen and Zadrozny, 1998], [Zadrozny, 1990], [Ghysels et al., 2006], [Ghysels et al., 2007]. All of the listed papers have somehow an underlying model at the highest frequency and suggest algorithms how to estimate/ reconstruct the model, or explain the slow frequency variables via the high frequency variables only.

The model/ theory presented in this chapter is designed for available data only. This is different to the approaches mentioned above, as we do not present the theoretical model at the highest frequency, but present a way how to optimally use the "incomplete" data.

Many of the following results followed from the intensive discussions with Manfred Deistler, Brian Anderson and Weitian Chen.

## 7.1 Framework

We will restrict ourselves to the two frequency case as a multi frequency situation can be easily adopted. We denote with $y_t = \begin{bmatrix} y_t^f \\ y_t^s \end{bmatrix}$ the $N$ dimensional vector of all variables at time $t$ and it consists of two sub vectors $y_t^f$ and $y_t^s$, where $y_t^f$ (fast frequency) is the vector of components which are available for every time and $y_t^s$ (slow frequency) is observed only every second (even) time point. The following table shows the setting, where an $x$ means that the variables are observed and a 0 that they are not observed.

| t | ... | 1 | 2 | 3 | 4 | ... |
|---|---|---|---|---|---|---|
| $y_t^f$ | ... | x | x | x | x | ... |
| $y_t^s$ | ... | 0 | x | 0 | x | ... |

$(y_t^f)_{t\in\mathbb{Z}}$ and $(y_t^s)_{t=2s,s\in\mathbb{Z}}$ are (wide sense) stationary processes. That implies that covariances at all lags are available for $y_t^f$ but only every second covariance of $y_t^s$ is available (defined). Furthermore the cross covariances between $y_t^f$ and $y_k^s$ depend only on the difference $t-k$.
We denote by $H_y(t) = \overline{span}\{y_{i,s}|i=1,...N, s\le t\}$ the Hilbert space spanned by all past and present variables of $y_t$, where the not observed values of $y_t^s$ (for odd $t$) are treated as if they are zero and thus have no additional information.

Furthermore, we here consider linearly regular processes only.

## 7.2 Wold representation

Before we present our approach we want to discuss the traditional approach shortly:

The Wold representation of the traditional approach would be

$$y_t = \sum_{j=0}^{\infty} k_j \zeta_{t-j}$$

where $\zeta_t$ denotes the innovation from $t-1$ to $t$. This is too ambitious for the mixed frequency case considered here. Thus it is not possible to use the well known one-step-ahead predictor

$$y_{t|t-1} = \sum_{j=1}^{\infty} k_j \zeta_{t-j}$$

as it requires the Hilbert space $\overline{span}\{\zeta_s | s \leq t\}$ which in the mixed frequency case is not "available". To illustrate this point we give a small example.

**Example 7.2.1.** *Let $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \zeta_t$ be a theoretical model for the 2 dimensional process $y_t$ whose second component is observed only every second time point. Assume that the matrices $a_1$ and $a_2$ are known, then the "best" one-step-ahead predictor according to Wold representation would be*

$$y_{t|t-1} = a_1 y_{t-1} + a_2 y_{t-2} \tag{7.1}$$

*Now the second component of either $y_{t-1}$ or $y_{t-2}$ is missing for every $t$. It is clear that the one-step-ahead error $y_t - y_{t|t-1}$ according to (7.1) is $[a_2]_{[,2]} y_{t-2,2} + \zeta_t$, where $[a_2]_{[,2]}$ denotes the second column of $a_2$, for odd $t$, and $[a_1]_{[,2]} y_{t-1,2} + \zeta_t$ for even $t$. We want to ask the question as to what information is lost if the second component is not available every second time instance. As the coefficients $a_1, a_2$ belong to the projection of $y_t$ onto the space spanned by **all** components of $\{y_{t-1}, y_{t-2}\}$ it is clear that the projection of $y_t$ onto the space spanned by $\{y_{t-1,1}, y_{t-1,2}, y_{t-2,1}\}$ for odd $t$ and onto the space spanned by $\{y_{t-1,1}, y_{t-2,1}, y_{t-2,2}\}$ for even $t$ will differ. Hence using the coefficients according to the projection of the only theoretically available space spanned by all components of $\{y_{t-1}, y_{t-2}\}$ is not optimal.*

The above example shows that a model at the highest frequency is nice to have in mind, but it is not clear if it is really the "best" for e.g. forecasting. This is the reason why we want to introduce our approach.

We will construct the Wold representation the same way as in the well known single frequency case:

$$y_t = y_t|_{H_y(t-1)} + \underbrace{y_t - y_t|_{H_y(t-1)}}_{=:\varepsilon_t} \tag{7.2}$$

with the difference that $H_y(t-1)$ is the Hilbert space as defined above, namely consisting of observable components only, and not the Hilbert space using also the theoretical variables of $y_t^s$ for odd $t$.
We know that $\varepsilon_t$ is the new information of $y_t$ and it can be split into two parts namely into the part spanned by the information coming from $y_t^f$, and if $t$ is even, the part spanned by the additional information coming from $y_t^s$. Thus, we can decompose the Hilbert space $H_{\varepsilon_t}$ spanned by the components of $\varepsilon_t$ into

the orthogonal subspaces $H_{\varepsilon_t^f}$ and $H_{\varepsilon_t^{f\perp}}$ (if the latter is not empty):

$$H_{\varepsilon_t} = H_{\varepsilon_t^f} \oplus H_{\varepsilon_t^{f\perp}} \tag{7.3}$$

where

$$\varepsilon_t^f := y_t^f - y_t^f|_{H_y(t-1)} \tag{7.4}$$

Therefore $\varepsilon_t^{f\perp}$ is the orthogonal complement of $H_{\varepsilon_t^f}$ in the space spanned by $\varepsilon_t$ and is therefore a subspace of the space spanned by $\varepsilon_t^s$, where $\varepsilon_t^s = y_t^s - y_t^s|_{H_y(t-1)}$. Summarizing we have

$$H_y(t) = H_y(t-1) \oplus H_{\varepsilon_t} = H_y(t-1) \oplus H_{\varepsilon_t^f} \oplus H_{\varepsilon_t^{f\perp}} \tag{7.5}$$

This construction ensures that $H_\varepsilon^f(t) \oplus H_\varepsilon^{f\perp}(t) = H_y(t)$ holds, where $H_\varepsilon^f(t) = \overline{span}\{\varepsilon_s^f | s \leq t\}$ and $H_\varepsilon^{f\perp}(t) = \overline{span}\{\varepsilon_s^{f\perp} | s \leq t\}$ and note that $\varepsilon_t^{f\perp}$ is zero for odd $t$. Therefore $y_t$ can be written in terms of $\varepsilon_s, s \leq t$, where we have to distinguish between $\varepsilon_t^{f,o}$ and $\varepsilon_t^{f,e}$ indicating if $t$ is odd or even.

- $y_t^f$
  - $t$ even: $y_t^f = \varepsilon_t^{f,e} + k_{1,1}^{e,f}\varepsilon_{t-1}^{f,o} + ... + k_{1,0}^{e,f\perp}\varepsilon_t^{f\perp} + k_{1,1}^{e,f\perp}0 + k_{1,2}^{e,f\perp}\varepsilon_{t-2}^{f\perp} + k_{1,3}^{e,f\perp}0 + k_{1,4}^{e,f\perp}\varepsilon_{t-4}^{f\perp} + ...$
  - $t$ odd: $y_t^f = \varepsilon_t^{f,o} + k_{1,1}^{o,f}\varepsilon_{t-1}^{f,e} + ... + k_{1,0}^{o,f\perp}0 + k_{1,1}^{o,f\perp}\varepsilon_{t-1}^{f\perp} + k_{1,2}^{o,f\perp}0 + k_{1,3}^{o,f\perp}\varepsilon_{t-3}^{f\perp} + k_{1,4}^{o,f\perp}0 + ...$
- $y_t^s$
  - $t$ even: $y_t^s = k_{2,0}^{e,f}\varepsilon_t^{f,e} + k_{2,1}^{e,f}\varepsilon_{t-1}^{f,o} + ... + k_{2,0}^{e,f\perp}\varepsilon_t^{f\perp} + k_{2,1}^{e,f\perp}0 + k_{2,2}^{e,f\perp}\varepsilon_{t-2}^{f\perp} + k_{2,3}^{e,f\perp}0 + k_{2,4}^{e,f\perp}\varepsilon_{t-4}^{f\perp} + ...$

where the coefficient matrices are defined in an evident way. It is important to distinguish between odd and even $t$ as the processes $(\varepsilon_t^f)_{t\in\mathbb{Z}}$ and $(\varepsilon_t^{f\perp})_{t\in\mathbb{Z}}$ are not stationary (as they are defined for all $t \in Z$). Thus, for example, the coefficients $k_{1,1}^{e,f}$ and $k_{1,1}^{o,f}$ do not have to be the same. For $y_t^s$ it is even more obvious as the coefficients $k_{2,i}^{e,f}, i = 0, 1, 2, ...$ are certainly not zero whereas the coefficients $k_{2,i}^{o,f}, i = 0, 1, 2, ...$ would be zero (we have not even defined this coefficients as we do not consider $y_t^s$ for odd $t$, but it would be the logical extension if we set $y_t^s = 0$ for odd $t$).

We result in a representation of $y_t$, depending on $t$, in terms of its innovations, but note, that $(y_t)_{t\in\mathbb{Z}}$, $(\varepsilon_t^f)_{t\in\mathbb{Z}}$ and $(\varepsilon_t^{f\perp})_{t\in\mathbb{Z}}$ are not stationary. Nevertheless we can use this representation to define a Wold representation of a blocked process, where all variables are stationary (as regarded at the slow frequency), which will be explained in the next section.

## 7.3 Blocking

Blocking is nothing else than stacking, but only available components. It is a frequently used approach in control engineering and therefore a huge literature exists which connects the "blocked" and the "unblocked" systems, see for instance [Colaneri and Longhi, 1995], [Bittanti and De Nicolao, 1993], [Bittanti et al., 1988]. In our case that means that we stack, e.g. $y_t^f$, $y_t^s$ and $y_{t+1}^f$. Thus we define

$$Y_{t+1} := \begin{bmatrix} y_t^f \\ y_t^s \\ y_{t+1}^f \end{bmatrix}, E_{t+1} := \begin{bmatrix} \varepsilon_t^{f,e} \\ \varepsilon_t^{f\perp} \\ \varepsilon_{t+1}^{f,o} \end{bmatrix} \tag{7.6}$$

Note that $Y_{t+1}$ and $E_{t+1}$ are defined only for even $t$. This construction ensures that the processes $(Y_{t+1})_{t=2s,s\in\mathbb{Z}}$ and $(E_{t+1})_{t=2s,s\in\mathbb{Z}}$ are stationary. By using the Wold representation of the previous section we get

$$Y_{t+1} = \begin{bmatrix} y_t^f \\ y_t^s \\ y_{t+1}^f \end{bmatrix} = \begin{bmatrix} [k_1^{e,f}(z^2)]^e & [k_1^{e,f\perp}(z^2)]^e & [k_1^{e,f}(z^2)]^o z^2 \\ [k_2^{e,f}(z^2)]^e & [k_2^{e,f\perp}(z^2)]^e & [k_2^{e,f}(z^2)]^o z^2 \\ [k_1^{o,f}(z^2)]^o & [k_1^{o,f\perp}(z^2)]^o & [k_1^{o,f}(z^2)]^e \end{bmatrix} \begin{bmatrix} \varepsilon_t^{f,e} \\ \varepsilon_t^{f\perp} \\ \varepsilon_{t+1}^{f,o} \end{bmatrix} = K(z^2)E_{t+1} \tag{7.7}$$

where

$$[k_i^{e,f}(z^2)]^e = \sum_{j=0}^{\infty} k_{i,2j}^{e,f} z^{2j}, i = 1,2 \tag{7.8}$$

$$[k_i^{e,f}(z^2)]^o = \sum_{j=0}^{\infty} k_{i,2j+1}^{e,f} z^{2j}, i = 1,2 \tag{7.9}$$

$$[k_i^{e,f\perp}(z^2)]^e = \sum_{j=0}^{\infty} k_{i,2j}^{e,f\perp} z^{2j}, i = 1,2 \tag{7.10}$$

$$[k_1^{o,f}(z^2)]^e = \sum_{j=0}^{\infty} k_{1,2j}^{o,f} z^{2j} \tag{7.11}$$

$$[k_1^{o,f}(z^2)]^o = \sum_{j=0}^{\infty} k_{1,2j+1}^{o,f} z^{2j} \tag{7.12}$$

$$[k_1^{o,f\perp}(z^2)]^0 = \sum_{j=0}^{\infty} k_{1,2j+1}^{o,f\perp} z^{2j} \tag{7.13}$$

Note that (7.7) is a Wold representation of $Y_{t+1}$ by construction. Also note that $K(0)$ has a lower block triangular structure as $\varepsilon_t^{f\perp}$ is not relevant for $y_t^f$ and $\varepsilon_{t+1}^{f,o}$ only effects $y_{t+1}^f$ of course.

## 7.4 Structure Theory

Analogously to the single frequency case we want to present a structure theory for the latent variables in the mixed frequency case. Let us commence from the $N$ dimensional vector of observations $y_t$ and split this vector into the two sub vectors $y_t^f$ consisting of the components which are available every time instance of dimension $N^f$ and $y_t^s$ consisting of the components which are available only every second (even) time instance of dimension $N^s$. Clearly $N = N^f + N^s$ holds.

Let us denote with $Y_{t+1}$ the $(2N^f + N^s)$ dimensional blocked vector

$$Y_{t+1} := \begin{bmatrix} y_t^f \\ y_t^s \\ y_{t+1}^f \end{bmatrix}$$

This vector can be treated as our former $y_t^N$ in the previous chapters such that

$$Y_{t+1} = X_{t+1} + U_{t+1} = \begin{bmatrix} \chi_t^f \\ \chi_t^s \\ \chi_{t+1}^f \end{bmatrix} + \begin{bmatrix} u_t^f \\ u_t^s \\ u_{t+1}^f \end{bmatrix} \tag{7.14}$$

where $(Y_{t+1})_{t=2s,s\in\mathbb{Z}}, (X_{t+1})_{t=2s,s\in\mathbb{Z}}$ and $(U_{t+1})_{t=2s,s\in\mathbb{Z}}$ fulfill Assumption 1 and Assumption 2 respectively and the spectral density of $Y_{t+1}$ satisfies the technical Assumption 3. That means if the $q$ largest eigenvalues of $f_Y$ diverge, where $q$ is the rank of $f_X$, the latent variables $X_{t+1}$ can be identified by letting $2N^f + N^s$ go to infinity. Note that it is not necessary that $N^f$ *and* $N^s$ diverge.

Thus it is feasible to model the latent variables $X_{t+1}$ by using the Wold representation introduced in the previous section:

$$X_{t+1} = \begin{bmatrix} \chi_t^f \\ \chi_t^s \\ \chi_{t+1}^f \end{bmatrix} = \begin{bmatrix} [k_1^{e,f}(z^2)]^e & [k_1^{e,f^\perp}(z^2)]^e & [k_1^{e,f}(z^2)]^o z^2 \\ [k_2^{e,f}(z^2)]^e & [k_2^{e,f^\perp}(z^2)]^e & [k_2^{e,f}(z^2)]^o z^2 \\ [k_1^{o,f}(z^2)]^o & [k_1^{o,f^\perp}(z^2)]^o & [k_1^{o,f}(z^2)]^e \end{bmatrix} \begin{bmatrix} \varepsilon_t^{f,e} \\ \varepsilon_t^{f^\perp} \\ \varepsilon_{t+1}^{f,o} \end{bmatrix} = K(z^2)E_{t+1} \tag{7.15}$$

Note that $q$, the rank of $f_X$, equals $q_e^f + q^{f^\perp} + q_o^f$ where $q_e^f$, $q^{f^\perp}$ and $q_o^f$ denote the dimensions of the processes $\varepsilon_t^{f,e}$, $\varepsilon_t^{f^\perp}$ and $\varepsilon_{t+1}^{f,o}$ respectively. Also note that $q_e^f = q_o^f$ does not have to hold.

Let

$$\mathscr{H} := \begin{bmatrix} K_0 & K_1 & K_2 & \dots \\ K_1 & K_2 & K_3 & \dots \\ K_2 & K_3 & \dots & \\ \vdots & & & \end{bmatrix} \tag{7.16}$$

where $K(z^2) = \sum_{j=0}^{\infty} K_j z^{2j}$ from (7.15), then we have

$$\begin{bmatrix} X_{t+1} \\ \hat{X}_{t+3|t+1} \\ \hat{X}_{t+5|t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} K_0 & K_1 & K_2 & \dots \\ K_1 & K_2 & K_3 & \dots \\ K_2 & K_3 & \dots & \\ \vdots & & & \end{bmatrix} \begin{bmatrix} E_{t+1} \\ E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix} \tag{7.17}$$

Consequently we can apply procedure presented in chapter 3 to define a state. If the blocked process $X_{t+1}$ has a rational spectrum, $K(z^2)$ is a rational spectral factor, and therefore $\mathscr{H}$ has finite rank. At this point we can again introduce the superscript $N$ for $X_{t+1}^N$, as we want to point out, that under the standard assumptions of the previous chapters, we can find again a state space representation which is independent of $N$ from an $N_0$ onwards. Note that the following structure theory has been already presented in chapter 3 for the single frequency case, and is repeated here for the sake of completeness. Let $n < \infty$ be the rank of $\mathscr{H}^\infty$ for $N = \infty$, such that Assumption 4 is fulfilled. Then from some $N_0$ onwards $\mathscr{H}^N$ has rank $n$. Let $S^{N_0}$ be the selector matrix which selects the first $n$ independent rows of $\mathscr{H}^{N_0}$ then we can define a state by

$$x_{t+1}^{N_0} := S^{N_0}\mathscr{H}^{N_0} \underbrace{\begin{bmatrix} E_{t+1} \\ E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix}}_{=:\mathbb{E}_{t+1}^-} = S^{N_0} \begin{bmatrix} K_0^{N_0} & K_1^{N_0} & K_2^{N_0} & \dots \\ K_1^{N_0} & K_2^{N_0} & K_3^{N_0} & \dots \\ K_2^{N_0} & K_3^{N_0} & \dots & \\ \vdots & & & \end{bmatrix} \begin{bmatrix} E_{t+1} \\ E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix} = \mathscr{H}_\alpha \begin{bmatrix} E_{t+1} \\ E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix} \tag{7.18}$$

Clearly the basis $\mathscr{H}_\alpha := S^{N_0}\mathscr{H}^{N_0}$ for the rows of $\mathscr{H}^{N_0}$ is also a basis for the rows of $\mathscr{H}^N$ with $N \geq N_0$. Therefore the state can be defined independently from the superscript $N$.

We want to recall, that we do not require that $N^s$, the number of variables of $\chi_t^s$, *and* $N^f$, the number of variables of $\chi_t^f$, tend to infinity. For example if $\chi_{t,i}^s$, i.e the $i$-th component of $\chi_t^s$, has further information for $\chi_{t+1}^f$ (for $N^f \to \infty$ the number of components of $\chi_{t+1}^f$, which are affected, has to tend to infinity; for the sake of simplicity we assume that all components are affected), the divergence of the eigenvalue of $f_X$, corresponding to the associated component of $\varepsilon_t^{f\perp}$can be guaranteed for $N_f \to \infty$ (whereas $N^s$

does not have to tend to infinity).

$$x_{t+1} = S^{N_0}\mathscr{H}^{N_0}\mathbb{E}^-_{t+1} = S^N_0 \underbrace{\begin{bmatrix} K^{N_0}_1 & K^{N_0}_2 & \dots \\ K^{N_0}_2 & K^{N_0}_3 & \dots \\ K^{N_0}_3 & \dots & \\ \vdots & & \end{bmatrix}}_{=:\tilde{\mathscr{H}}^{N_0}} \begin{bmatrix} E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix} + S^{N_0} \begin{bmatrix} K^{N_0}_0 \\ K^{N_0}_1 \\ K^{N_0}_2 \\ \vdots \end{bmatrix} E_{t+1} \tag{7.19}$$

Define $\tilde{F}^{N_0}$ by expressing $\tilde{\mathscr{H}}^N_0$ in terms of the basis $\mathscr{H}_\alpha$ (as $\tilde{\mathscr{H}}^N_0$ is a submatrix of $\mathscr{H}^N_0$)

$$\mathscr{H}^{N_0} = \tilde{F}^{N_0}\mathscr{H}_\alpha \tag{7.20}$$

then we can define $F$ independently of $N$

$$F := S^{N_0}\tilde{F}^{N_0} \tag{7.21}$$

with the consequence that

$$x_{t+1} = S^{N_0}\mathscr{H}^{N_0}\mathbb{E}^-_{t+1} = S^{N_0} \begin{bmatrix} K^{N_0}_1 & K^{N_0}_2 & \dots \\ K^{N_0}_2 & K^{N_0}_3 & \dots \\ K^{N_0}_3 & \dots & \\ \vdots & & \end{bmatrix} \begin{bmatrix} E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix} + S^N_0 \begin{bmatrix} K^{N_0}_0 \\ K^{N_0}_1 \\ K^{N_0}_2 \\ \vdots \end{bmatrix} E_{t+1} \tag{7.22}$$

$$= S^{N_0}\tilde{F}^N_0 \underbrace{\mathscr{H}_\alpha \begin{bmatrix} E_{t-1} \\ E_{t-3} \\ \vdots \end{bmatrix}}_{x_{t-1}} + S^{N_0} \begin{bmatrix} K^{N_0}_0 \\ K^{N_0}_1 \\ K^{N_0}_2 \\ \vdots \end{bmatrix} E_{t+1}E_{t+1} = Fx_{t-1} + S^{N_0} \begin{bmatrix} K^{N_0}_0 \\ K^{N_0}_1 \\ K^{N_0}_2 \\ \vdots \end{bmatrix} E_{t+1} = Fx_{t-1} + GE_{t+1} \tag{7.23}$$

as $G := S^{N_0} \begin{bmatrix} K^{N_0}_0 \\ K^{N_0}_1 \\ K^{N_0}_2 \\ \vdots \end{bmatrix}$ can be chosen independently from $N$ too. Defining $H^{N_0}$ by expressing the first row block of $\mathscr{H}^{N_0}$ in terms of $\mathscr{H}_\alpha$ finishes the procedure.

$$\begin{bmatrix} K^{N_0}_0 & K^{N_0}_1 & K^{N_0}_2 & \dots \end{bmatrix} = H^{N_0}\mathscr{H}_\alpha \tag{7.24}$$

Therefore we result in

$$x_{t+1} = Fx_{t-1} + GE_{t+1}$$
$$X_{t+1}^N = \begin{bmatrix} \chi_t^{f,N} \\ \chi_t^{s,N} \\ \chi_{t+1}^{f,N} \end{bmatrix} = H^N x_{t+1}$$

## 7.5 Forecasting

Here we want to present the optimal forecasting procedure for our approach. A nowcasting procedure, see e.g. [Giannone et al., 2008], would be also desirable but might be very difficult, as our approach uses the space spanned by all observable components only.

In the previous section we have presented the state space construction for the blocked process

$$\bar{X}_{t+1}^N = \begin{bmatrix} \chi_t^{f,N} \\ \chi_t^{s,N} \\ \chi_{t+1}^{f,N} \end{bmatrix} \tag{7.25}$$

$$\bar{x}_{t+1} = \bar{F}\bar{x}_{t-1} + \bar{G}\bar{E}_{t+1} \tag{7.26}$$
$$\bar{X}_{t+1}^N = \begin{bmatrix} \chi_t^{f,N} \\ \chi_t^{s,N} \\ \chi_{t+1}^{f,N} \end{bmatrix} = \bar{H}^N \bar{x}_{t+1}, t = 2s, s \in \mathbb{Z} \tag{7.27}$$

where $\bar{E}_{t+1} = \begin{bmatrix} \varepsilon_t^{f,e} \\ \varepsilon_t^{f,\perp} \\ \varepsilon_{t+1}^{f,o} \end{bmatrix}$. Thus it is very easy to build the best forecast for $X_{t+3}^N$ having information up to time $t+1$:

$$\hat{\bar{X}}_{t+3|t+1}^N = \bar{H}^N \bar{F} \bar{x}_{t+1} \tag{7.28}$$

Now as $\bar{X}_{t+1}^N = \begin{bmatrix} \chi_t^{f,N} \\ \chi_t^{s,N} \\ \chi_{t+1}^{f,N} \end{bmatrix}$ it is clear that $\chi_{t+3}^{f,N}$ could be forecasted better if information were available up to time $t+2$. Unfortunately, the described procedure cannot adopt to the new information and build a

new/ better forecast $\chi^{f,N}_{t+3|t+2}$. Thus we repeat the whole procedure described above for the vector

$$\tilde{X}^N_t = \begin{bmatrix} \chi^{f,N}_{t-1} \\ \chi^{f,N}_t \\ \chi^{s,N}_t \end{bmatrix} \tag{7.29}$$

and obtain a second system

$$\tilde{x}_{t+2} = \tilde{F}\tilde{x}_t + \tilde{G}\tilde{E}_t \tag{7.30}$$

$$\tilde{X}^N_t = \begin{bmatrix} \chi^{f,N}_{t-1} \\ \chi^{f,N}_t \\ \chi^{s,N}_t \end{bmatrix} = \tilde{H}^N \tilde{x}_t, t = 2s, s \in \mathbb{Z} \tag{7.31}$$

with $\tilde{E}_t = \begin{bmatrix} \varepsilon^{f,o}_{t-1} \\ \varepsilon^{f,e}_t \\ \varepsilon^{f\perp}_t \end{bmatrix}$. Thus

$$\tilde{X}^N_{t+2|t} = \tilde{H}^N \tilde{F} \tilde{x}_t \tag{7.32}$$

Therefore, (7.28) and (7.32) can be used for the best forecasts

$$\hat{\chi}_{t+1|t} \tag{7.33}$$

and

$$\hat{\chi}_{t+2|t} \tag{7.34}$$

irrespective if the forecast is for the fast or the slow frequency variables and if $t$ is odd or even.

## 7.6 Simulation

Here we want to give a comparison between our method and the naive method described at the beginning of this chapter. We start with simulating a 2 dimensional zero mean stable AR(2) process

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$$

where $a_i \in \mathbb{R}^{2\times 2}$, and $\varepsilon_t$ are the white-noise innovations and $\mathbb{E}\varepsilon_t\varepsilon_t'$ has full rank. We denote by $Y$ the $(T \times 2)$ sample of the process $y_t$ where $T$ is the sample size and split the sample into an insample matrix $Y_{in}$ and an out-of-sample matrix $Y_{out}$ of dimensions $(T_{in} \times 2)$ and $(T_{out} \times 2)$. Next we set the second component of $y_t$ equal to zero for odd $t$ and result in the modified sample $\tilde{Y}$ (again we split it in the insample and out-of-sample matrices). We compare three (four) approaches:

1. We use the sample $Y_{in}$ (which is not available in practice) to estimate the coefficients $a_1, a_2$. Next we use the whole sample $Y$ to get the one-step-ahead predictors (in- and out-of-sample). After that the predictors for the second component for odd $t$ are set equal to zero.

2. We stack the insample matrix $\tilde{Y}_{in}$ to two three dimensional matrices according to (7.25), (7.29) and estimate for both samples an AR(1) system. Afterwards the one-step-ahead predictors of both models using $\tilde{Y}$ are calculated. Of course we take always the best one-step-ahead forecasts which are available, i.e. the predictor for odd $t$ is taken from the model for (7.29) whereas for even $t$ the predictor is taken from the model for (7.25).

3. We estimate the coefficients $a_1, a_2$

   (a) by using $\tilde{Y}_{in}$, according to [Chen and Zadrozny, 1998], i.e. the columns of $\hat{\Gamma}_p$ which have zeros (by construction) are replaced by other columns until we have a square non singular matrix.

   (b) by using $\tilde{Y}_{in}$ and forget about the problem that certain covariances are zero.

   After that the one-step-ahead predictors are calculated by using $\tilde{Y}$. In the end all predictors corresponding to odd $t$ are set equal to zero.

All covariances are estimated by

$$\hat{\gamma}_{j,[k,l]} = 1/\min(T_k, T_l) \sum_{t=1}^{T-j} y_{t+j,k} y_{t,l}, j \geq 0$$
$$\hat{\gamma}_{j,[k,l]} = \hat{\gamma}_{-j,[l,k]}, j < 0$$

where $\hat{\gamma}_{j,[k,l]}$ denotes the $[k,l]$ entry of $\hat{\gamma}_j$ which is an estimate of $\gamma_j = \mathbb{E} y_t y'_{t-j}$, $y_{t,k}$ denotes the $k$ th component of $y_t$ and $T_k$ is the number of non zero (available) observations of $y_{t,k}, t = 1, ..., T_{in}$.

The methods are compared by calculating the Frobenius norm of the sample covariances of the insample and out-of-sample one-step-ahead prediction errors corresponding to the four methods

$$\|\Sigma_{in}\|_F, \|\Sigma_{out}\|_F$$

and a multiple $R^2$ statistic, by regressing the estimated $\hat{\tilde{Y}}_{in}$ onto $\tilde{Y}_{in}$ and $\hat{\tilde{Y}}_{out}$ onto $\tilde{Y}_{out}$:

$$R^2_{in} = \frac{tr(\hat{\tilde{Y}}'_{in} \tilde{Y}_{in} [\tilde{Y}'_{in} \tilde{Y}_{in}]^{-1} \tilde{Y}'_{in} \hat{\tilde{Y}}_{in})}{tr(\hat{\tilde{Y}}'_{in} \hat{\tilde{Y}}_{in})}$$

$$R^2_{out} = \frac{tr(\hat{\tilde{Y}}'_{out} \tilde{Y}_{out} [\tilde{Y}'_{out} \tilde{Y}_{out}]^{-1} \tilde{Y}'_{out} \hat{\tilde{Y}}_{out})}{tr(\hat{\tilde{Y}}'_{out} \hat{\tilde{Y}}_{out})}$$

The insample size $T_{in}$ is set to 150, the out-of-sample size $T_{out}$ is set to 100 and we simulate 100 processes with arbitrary coefficient matrices $a_1, a_2$ (subject to $I - a_1 z - a_2 z^2$ is stable). The following table shows the mean and standard deviation of the statistics of the 100 simulations

$\|\Sigma_{in}\|_F, T_{in} = 150$

| | all | wold | high freq. a | high freq. b |
|---|---|---|---|---|
| mean | 0.322 | 0.347 | 0.798 | 0.549 |
| sd | 0.30 | 0.29 | 0.70 | 0.59 |

$\|R^2_{in}\|_F, T_{in} = 150$

| | all | wold | high freq. a | high freq. b |
|---|---|---|---|---|
| mean | 0.775 | 0.745 | 0.671 | 0.664 |
| sd | 0.22 | 0.22 | 0.25 | 0.26 |

$\|\Sigma_{out}\|_F, T_{out} = 100$

| | all | wold | high freq. a | high freq. b |
|---|---|---|---|---|
| mean | 0.350 | 0.399 | 0.857 | 0.625 |
| sd | 0.33 | 0.34 | 0.77 | 0.72 |

$\|R^2_{out}\|_F, T_{out} = 100$

| | all | wold | high freq. a | high freq. b |
|---|---|---|---|---|
| mean | 0.768 | 0.725 | 0.664 | 0.654 |
| sd | 0.23 | 0.25 | 0.26 | 0.27 |

Of course the method, using all data (although this can never happen in practice), performs best, but our approach outperforms the naive method, which is very promising for future research.

## 7.7 Summary

We have introduced an approach for modeling mixed frequency data. We have restricted ourselves to the two frequency case, but we claim that this approach can be extended to the multi frequency case quite easily. It turns out that we have to be very careful about distinguishing between odd and even time points, i.e. if all data is available or not, as the coefficients in our Wold representation depend on this distinction. Especially the driving white noise, which is not stationary if considered for the "unblocked" system (but is stationary if considered for the "blocked" system), depends on whether $t$ is odd or even. As this thesis is concerned with Generalized Dynamic Factor Models we have shown how this procedure can be

embedded in the theory of GDFMs. Furthermore we have presented the optimal forecasting procedure in our framework, by using two models which differ by blocking different components. Finally we have compared this method to the naive method and it seems as if our approach is worth having a closer look at it.

# Chapter 8

# Summary and Outlook

## Summary

The thesis deals with high dimensional time series which are modeled as Generalized Dynamic Factor Models. The basic assumption of factor models is, that the observations can be split into two orthogonal summands, the latent variables and the idiosyncratic noise. Generalized Dynamic Factor Models have been introduced in [Forni and Lippi, 2001], [Forni et al., 2000] and [Stock and Watson, 2002a], [Stock and Watson, 2002b] and are used by many central banks, e.g. ECB and FED, for forecasting macro economic time series.

Thus the first main chapter in this thesis presents the major results of the paper [Forni and Lippi, 2001] which establishes the theoretical background and the characterization of this model class.

The following chapter presents what we call "Structure Theory", i.e. commencing from the (theoretically available) second moments of the latent variables we present a (the most general) possible procedure how to model the latent variables. This structure theory uses a state space system where the minimal state has a (singular) AR(1) representation. As the minimal state does not have to be a minimal static factor, a generic result is presented, in which the minimal static factor can be modeled as a singular autoregression.

Consequently the next chapter deals with the theory of singular autoregressions, which have been poorly discussed in the existing literature. We have presented an intensive study of the Yule-Walker equations for these models, which do not have to have a unique solution opposed to the regular case. Therefore we have focused on a special solution of the Yule-Walker equations, namely the minimum norm solution. We have seen that if the underlying process is purely linearly regular, the minimum norm solution yields a stable polynomial. If the underlying stationary process is the sum of a linearly regular and a linearly singular process the minimum norm solution will yield in an autoregressive polynomial which has no

zeros inside but has zeros both outside and on the unit circle. Furthermore it has the minimal number of unit roots among all solutions of the Yule-Walker equations and the linearly regular part of the underlying stationary process corresponds always to the transfer function and is given by $f_t^r = a(z)^{-1}b\varepsilon_t$. Beside the minimum norm solution we also present two different canonical representatives, as the minimum norm solution does not have to result in a left coprime pair $[a(z), b]$.

In the next chapter we have presented an estimation procedure for the minimum norm solution. It turns out, that the estimated AR polynomial is, analogously to the regular case, always stable, even if a linearly singular component is present. A simulation shows that the trivial way of estimating the AR coefficients (namely inverting the corresponding covariance matrix as long as it is numerically possible) is possible but the estimated coefficients seem to have a very large variance and it is not clear what happens for $T \to \infty$ (where $T$ is the sample size). Furthermore, the simulation shows that the forecasting performances are almost the same such that no winner can be identified. We have also presented possible ways to estimate the order of a singular AR process, as the well known Information Criteria such as AIC and BIC need a regular covariance matrix of the driving white noise. A consistent estimator is presented but as the practical value of this procedure is not worth mentioning another estimator is introduced. It is shown that our suggested estimating procedure produces comparable results to AIC in the regular case and outperforms the AIC in the singular case.

Next it is shown, that in the case where the latent variables have linearly singular components (we restricted ourselves to harmonic processes) the corresponding minimal static factor has such linearly singular components too and thus the corresponding autoregressive polynomial has a unit root. This case can be handled of course with our structure theory as we also discussed singular AR processes with linearly singular components in the previous chapters.

The last chapter deals with the mixed frequency problem. This scenario appears if multivariate time series are modeled where the variables appear with different sampling frequencies. We present a Wold representation for this case and use this representation to present a mini-phase spectral factor of a blocked (stationary) process. Furthermore this theory is embedded in the context of GDFMs such that the latent variables can be modeled according to our mixed frequency procedure. Finally a small simulation compares our approach to the naive method where our method seems to be clearly advantageous to the other one.

## Outlook

There are still a number of open questions concerning estimation procedures. Our idea, to estimate the needed coefficient matrices, is the following:

1. Denoise the data, i.e. apply PCA or dynamic PCA to obtain the latent variables.
2. Using the result from the first step and estimate a singular autoregression polynomial for the minimal static factor (which is obtained by a static linear transformation of the latent variables).

The problem is, that the first step is done without taking into account the dynamics of the minimal static factor. Thus a recursive procedure might be a logical extension of our existing one.

Furthermore the estimation of $s$, the rank of the covariance matrix $\Gamma_p$, is not solved. If a true singular autoregressive process is observed, the suggested procedure is ok but not really exciting. Once the singular AR process is not observed directly, but has to be estimated in a first step (as in the theory of GDFMs), this problem is still open.

Of course the last chapter of this thesis concerning the mixed frequency problem is only a first step for handling this situation. It is not clear up to now how to model the variables at their finest rates (to unblock the system again), i.e. the fast variables at the fast frequency and the slow variables at the slow frequency or maybe even at the high frequency. Consequently the interpolation, i.e. nowcasting, of the slow variables is still open, which is a very interesting question as a nowcast of the GDP is very valuable. Last but not least a generic zeroless result analogous to the single frequency case would be desirable too.

# Chapter 9

# Useful Preliminary Work

## 9.1 Polynomial Matrices

Here we want to summarize some properties of polynomial matrices

$$a(z) = \sum_{j=0}^{p} a_j z^j \tag{9.1}$$

where $a_j \in \mathbb{R}^{n \times m}$.

**Definition 9.1.1. (unimodular).** *An $(n \times n)$ polynomial matrix $a(z)$ is called unimodular if its determinant is a constant unequal to zero.*

**Definition 9.1.2. (coprime).** *Two polynomial matrices $[a(z), b(z)]$ ($[a(z)', b(z)']'$) are called left (right) coprime if the matrix $[a(z), b(z)]$ ($[a(z)', b(z)']'$) has full row (column) rank for all $z \in \mathbb{C}$.*

**Lemma 9.1.3.** *For the polynomial matrices $[a(z), b(z)]$ (where $det(a(z)) \not\equiv 0$) the following statements are equivalent:*

*(i)* $[a(z), b(z)]$ *is left coprime.*

*(ii) The identity matrix $I$ is a greatest left divisor of $[a(z), b(z)]$.*

*(iii) There exist polynomial matrices $g, h$ such that*

$$ag + bh = I \tag{9.2}$$

*holds.*

*(iv) The degree of $det(a)$ is minimal among all matrices $[\bar{a}(z), \bar{b}(z)]$ with $\bar{a}^{-1}\bar{b} = a^{-1}b$.*

*(v)* $[a(z), b(z)]$ *has full rank for all* $z \in \mathbb{C}$.

*Proof.* ***(Lemma 9.1.3.)*** See Lemma 2.2.1 page 40 in [Hannan and Deistler, 1988]. □

**Definition 9.1.4. (rowreduced).** *An* $(n \times m)$ *polynomial matrix* $a(z)$ *is called rowreduced (columnreduced) if its row (column) end matrix is of full rank. The* $i$*-th row (column) of the row (column) end matrix is the* $i$*-th row (column) of the coefficient matrix of* $a(z)$*,* $a_j$ *say, belonging to the maximal degree* $j$ *of the entries of the* $i$*-th row (column) of* $a(z)$*.*

## 9.2 Rational Transfer Functions

Rational transfer functions play an important role in many areas, especially in electrical engineering, control theory and in time series analysis. As a rational transfer function can be described in different ways we present here the most important ones and some properties and definitions of rational functions.

### 9.2.1 State Space Systems

The theory of state space systems is an important theory for many areas and therefore well discussed. Here we will present only a few definitions and theorems which are useful for our structure theory. A general form of a state space model is

$$\tilde{x}_{t+1} = \tilde{A}\tilde{x}_t + e_t \tag{9.3}$$

$$y_t = \tilde{C}\tilde{x}_t + u_t \tag{9.4}$$

where $\tilde{A} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}, \tilde{C} \in \mathbb{R}^{N\times\tilde{n}}$, $y_t$ is an $N$ dimensional vector of observed outputs, $\tilde{x}_t$ denotes the $\tilde{n}$ dimensional state and $e_t$ and $u_t$ are white noise vector processes of dimensions $\tilde{n}$ and $N$ respectively. Note that $e_t$ and $u_t$ do not have to be uncorrelated at the same time points but have to be uncorrelated for different time points.

In [Hannan and Deistler, 1988] it is shown that a stable state space system (9.3), (9.4) (stable means that the largest eigenvalue of $A$ is smaller than one in modulus) can always be formulated as $[A, B, C, D]$ state space system

$$x_{t+1} = Ax_t + B\varepsilon_t \tag{9.5}$$

$$y_t = Cx_t + D\varepsilon_t \tag{9.6}$$

where $A \in \mathbb{R}^{\tilde{n}\times\tilde{n}}, B \in \mathbb{R}^{\tilde{n}\times N}, C \in \mathbb{R}^{N\times\tilde{n}}, D \in \mathbb{R}^{N\times N}$, $x_t$ denotes an $\tilde{n}$ dimensional state and $\varepsilon_t$ denotes an $N$ dimensional white noise process which is also the one-step-ahead prediction error, i.e. it is the white noise process that drives the observed process $y_t$.

In both representations (9.3), (9.4) and (9.5), (9.6) it is assumed that the state collects the whole information from $y_s, s < t$ which is useful for the present and future observations $y_s, s \geq t$. These representations and their properties are well discussed in [Kailath, 1980] and [Hannan and Deistler, 1988] for instance.

In our structure theory we are using $[F, G, H]$ state space systems of the form

$$x_{t+1} = Fx_t + G\varepsilon_{t+1} \tag{9.7}$$

$$y_t = Hx_t \tag{9.8}$$

where $F \in \mathbb{R}^{n\times n}, G \in \mathbb{R}^{n\times q}, H \in \mathbb{R}^{N\times n}$, $x_t$ denotes an $n$ dimensional state and $\varepsilon_t$ is a white noise process of dimension $q$. Here, contrary to (9.5), (9.6), $x_t$ collects the whole information from $y_s, s \leq t$ which is useful for the present and future observations $y_s, s \geq t$. In the following we will discuss the properties of (9.7), (9.8) which will turn out to be, not surprisingly, almost the same as the properties of (9.5), (9.6).

Although the terminologies of observability and reachability are usually defined for $[A, B, C, D]$ systems (9.7), (9.8) (see [Hannan and Deistler, 1988], [Kailath, 1980]) we will see in the following that all results can be adopted for our $[F, G, H]$ systems (9.7), (9.8).

**Definition 9.2.1. (observable).** *A system* (9.7)*,* (9.8) *is called observable if the matrix*

$$\mathscr{O} = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix} \tag{9.9}$$

*has full column rank* $n$*.*

Therefore observability means that in the absence of noise we can determine the initial value of $x_0$ from $y_t, t = 0, ..., n-1$.

**Definition 9.2.2. (reachable).** *A system* (9.7)*,* (9.8) *is called reachable if the matrix*

$$\mathscr{R} = [G, FG, ..., F^{n-1}G] \tag{9.10}$$

*has full column rank* $n$*.*

Using (9.7) we observe that

$$x_{t+k} = [G, FG, ..., F^{k-1}G] \begin{bmatrix} \varepsilon_{t+k} \\ \vdots \\ \varepsilon_{t+1} \end{bmatrix} + F^k x_t \tag{9.11}$$

holds. Therefore, for $k \geq n$, reachability means that we can reach any state $x_{t+k}$ for given initial value $x_t$ if the innovations $\varepsilon_{t+i}, i = 1, ..., k$ are under our control.

**Theorem 9.2.3.** *The reachability matrix $\mathscr{R} = [G, FG, ..., F^{n-1}G]$ is of full rank if and only if $[\lambda I - F, G]$ has full rank for all $\lambda \in \mathbb{C}$.*

*Proof.* ***(Theorem 9.2.3.)*** See Theorems 2.4-8. on page 135 and 2.4-9. on page 136 in [Kailath, 1980]. □

If reachability does not hold one is often interested in certain eigenvalues of $F$.

**Definition 9.2.4. (reachable mode).** *Let $\lambda$ be an eigenvalue of $F$, then $\lambda$ is a reachable mode of the pair $[F, G]$ if the matrix $[\lambda I - F, G]$ has full (row) rank. Otherwise $\lambda$ is called an unreachable mode.*

**Definition 9.2.5. (minimal).** *A system* (9.7)*,* (9.8) *is called minimal if it is observable and reachable.*

Note that the definitions of reachability and observability for (9.7), (9.8) are the same as the definitions for a system (9.5), (9.6) by taking into account the different indexing and meaning of the state.
It is worth noting that the dimensions $\tilde{n}$ and $n$ of the minimal states of the state space representations (9.5), (9.6) and (9.7), (9.8) are *not* the same.
Let us look at a simple example to illustrate the difference between the systems (9.7), (9.8) and (9.5), (9.6).

**Example 9.2.6.** *Let $y_t = \varepsilon_t + \varepsilon_{t-1}$ be an MA(1) where $y_t$ and $\varepsilon_t$ are univariate processes. A minimal $[F, G, H]$ state space system would be*

$$x_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_{t+1} \tag{9.12}$$

$$y_t = [1, 1] x_t \tag{9.13}$$

*whereas a minimal $[A, B, C, D]$ state space system would be*

$$x_t = \varepsilon_{t-1} = 0 x_{t-1} + 1 \varepsilon_{t-1} \tag{9.14}$$

$$y_t = 1 x_t + 1 \varepsilon_t \tag{9.15}$$

*Which shows that the dimensions of the states of the different systems differ.*

### 9.2.2 Zeros and Poles

Let $w(z)$ be an $(N \times q)$, rational function of rank $k \leq \min(N, q)$. It is well known (see for instance [Hannan and Deistler, 1988] page 53) that the (finite) zeros and poles of $w(z)$ can be defined by its ***Smith-McMillan form*** $\Lambda(z)$

$$w(z) = u(z)\Lambda(z)v(z) \tag{9.16}$$

where $u(z)$ and $v(z)$ are unimodular matrices of dimension $(N \times N)$ and $(q \times q)$ respectively, and $\Lambda(z)$ is of dimension $(N \times q)$ and has zeros everywhere except in the diagonal of its upper $(k \times k)$ sub matrix. The entries of this diagonal matrix are of the form

$$\Lambda_{i,i} = n_i/d_i$$

where $n_i$ and $d_i$ are relatively prime, $n_i$ divides $n_{i+1}$ and $d_{i+1}$ divides $d_i$ for $i = 1, ..., k-1$. The matrix $\Lambda(z)$ is unique, whereas $u(z)$ and $v(z)$ are not. The zeros and poles of $w(z)$ are then defined as the zeros of the $n_i$s and $d_i$s respectively.
The following lemma gives a nice characterization of the zeros of $w(z)$ using its representation as a minimal state space system, which can be found in [Kailath, 1980] page 448.

**Lemma 9.2.7.** *Let $w(z)$ be an $(N \times q)$ rational transfer function such that $y_t = w(z)\varepsilon_t$ holds, and let* (9.7)*,* (9.8) *be a minimal state space system for $y_t$ with $w(z) = H[I - Fz]^{-1}G$, then the matrices $\begin{bmatrix} I - Fz & -G \\ H & 0 \end{bmatrix}$ and $\begin{bmatrix} I & 0 \\ 0 & c(z) \end{bmatrix}$, where $c(z)$ is the numerator matrix of a right coprime matrix fraction description $c(z), d(z)$ of $w(z) = c(z)d^{-1}(z)$, where $c(z)$ is of dimension $(N \times q)$ and $d(z)$ is of dimension $(q \times q)$, have the same Smith-Mc Millan form, i.e. they have the same (finite) zeros.*

*Proof.* ***(Lemma 9.2.7.)***
By definition we have

$$w(z) = c(z)d^{-1}(z) = H[I - Fz]^{-1}G$$

with $\begin{bmatrix} c(z) \\ d(z) \end{bmatrix}$ right coprime, and define $k(z) = [I - Fz]^{-1}G = \tilde{c}(z)\tilde{d}^{-1}(z)$ with $\tilde{c}(z), \tilde{d}(z)$ right coprime such that $c(z) = H\tilde{c}(z)$ holds. Note that a full rank condition on $H$ is not needed. By Lemma 9.1.3.(iii) we know that there exist polynomial matrices $X(z), Y(z), \bar{X}(z), \bar{Y}(z)$ of dimensions $(n \times n), (q \times n), (q \times n)$ and $(q \times q)$ such that

$$[I - Fz]X(z) + GY(z) = I_n \tag{9.17}$$

$$\bar{X}(z)\tilde{c}(z) + \bar{Y}(z)\tilde{d}(z) = I_q \tag{9.18}$$

hold. As $[I - Fz]\tilde{c}(z) - G\tilde{d}(z) = 0$ holds too we have

$$\begin{bmatrix} I - Fz & -G \\ \bar{X}(z) & \bar{Y}(z) \end{bmatrix} \begin{bmatrix} X(z) & \tilde{c}(z) \\ -Y(z) & \tilde{d}(z) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ -Q(z) & I_q \end{bmatrix} \tag{9.19}$$

where $Q(z) = -\bar{X}(z)X(z) + \bar{Y}(z)Y(z)$. As the two square block matrices on the left hand side of (9.19) are polynomial they are unimodular as the matrix on the right hand side of (9.19) is unimodular. Thus we have

$$\begin{bmatrix} I_n - Fz & -G \\ H & 0 \end{bmatrix} \begin{bmatrix} X(z) & \tilde{c}(z) \\ -Y(z) & \tilde{d}(z) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ HX(z) & H\tilde{c}(z) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ HX(z) & c(z) \end{bmatrix} \tag{9.20}$$

As matrices which are related by a multiplication of a unimodular matrix have the same Smith-Mc Millan form, the matrices $\begin{bmatrix} I_n & 0 \\ HX(z) & c(z) \end{bmatrix}$ and $\begin{bmatrix} I_n - Fz & -G \\ H & 0 \end{bmatrix}$ have the same Smith-Mc Millan form. As

$$\begin{bmatrix} I_n & 0 \\ -HX(z) & I_N \end{bmatrix} \begin{bmatrix} I_n & 0 \\ HX(z) & c(z) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & c(z) \end{bmatrix}$$

holds, and the first matrix on the left hand side of the equation above is unimodular, the result follows. □

### 9.2.3 McMillan degree

The McMillan degree is an important concept in the theory of linear systems. It is on the one hand an indicator of the dynamics of a system and on the other hand it links a lot of properties. We will use this concept for proper rational functions only.

**Definition 9.2.8. (proper rational function).** *A rational transfer function matrix $\tilde{w}(z)$ is said to be proper if*

$$\lim_{z\to\infty} \tilde{w}(z) < \infty$$

*and strictly proper if*

$$\lim_{z\to\infty} \tilde{w}(z) = 0$$

**Definition 9.2.9. (McMillan degree).** *Let $\tilde{w}(z)$ be a proper rational function and $\Lambda(z)$ its Smith-McMillan form from* (9.16)*, then the McMillan degree is the sum of the degrees of the denominator polynomials of $\Lambda(z)$.*

The concept can be extended to non proper rational functions too, see [Kailath, 1980] page 466.

In our framework we always use a rational transfer function $w(z)$ which is causal, i.e. has a power

series representation within and on the unit circle $w(z) = \sum_{j=0}^{\infty} w_j z^j$, and is therefore not proper. Thus we define a rational strictly proper function $\tilde{w}(z)$ which stands in a one-to-one relation with $w(z)$

$$\tilde{w}(z) = w(z^{-1})z^{-1} = \sum_{j=0}^{\infty} w_j z^{-j-1} \tag{9.21}$$

The interesting link for our structure theory is the following. Given a rational causal transfer function $w(z)$ then the McMillan degree of $\tilde{w}(z) = w(z^{-1})z^{-1}$, $n$ say, equals the dimension of a minimal state of an $[F, G, H]$ state space system of $w(z)$ (see also Theorem 3.2.2 (iii)).
Note that $w(z)$ and $\tilde{w}(z)$ have in general different McMillan degrees.

**Example 9.2.10.** *Let $w(z) = 1$ be a trivial transfer function which has McMillan degree 0, corresponding to $y_t = \varepsilon_t$. The corresponding $\tilde{w}(z) = w(z^{-1})z^{-1} = z^{-1}$ has McMillan degree 1. And a minimal $[F, G, H]$ state space system for $y_t$ is of order 1.*

$$\begin{aligned} x_{t+1} &= 0x_t + 1\varepsilon_{t+1} \\ y_t &= 1x_t \end{aligned} \tag{9.22}$$

# Bibliography

[Akaike, 1974] Akaike, H. (1974). Stochastic Theory of Minimal Realization. *IEEE Transactions on Automatic Control*, 19(6):667–674.

[Anderson and Deistler, 2008a] Anderson, B. D. O. and Deistler, M. (2008a). Generalized Linear Dynamic Factor Models - A Structure Theory. In *Conference on Decision and Control*.

[Anderson and Deistler, 2008b] Anderson, B. D. O. and Deistler, M. (2008b). Properties of Zero-free Transfer Function Matrices. *SICE Journal of Control, Measurement and System Integration*, 1(4):284–292.

[Anderson and Deistler, 2009] Anderson, B. D. O. and Deistler, M. (2009). Properties of Zero-free Spectral Matrices. *IEEE Transactions on Automatic Control*, 54(10):2365–2375.

[Anderson et al., 2009] Anderson, B. D. O., Deistler, M., Chen, W., and Filler, A. (2009). AR models of singular spectral matrices. In *Conference on Decision and Control*.

[Anderson et al., 2010] Anderson, B. D. O., Deistler, M., Chen, W., and Filler, A. (2010). Autoregressive models of singular spectral matrices. *submitted to Automatica*.

[Apostol, 1974] Apostol, T. M. (1974). *Mathematical Analysis*. Addison-Wesley.

[Bittanti et al., 1988] Bittanti, S., Colaneri, P., and De Nicolao, G. (1988). The Difference Periodic Riccati Equation for the Periodic Prediction Problem. *IEEE Transactions on Automatic Control*, 33(8):706–712.

[Bittanti and De Nicolao, 1993] Bittanti, S. and De Nicolao, G. (1993). Spectral Factorization of Linear Periodic Systems with Application to the Optimal Prediction of Periodic. *Automatica*, 29(2):517–522.

[Chamberlain, 1983] Chamberlain, G. (1983). Funds, Factors and Diversification in Arbitrage Pricing Models. *Econometrica*, 51(5):1305–1324.

[Chamberlain and Rothschild, 1983] Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure and Meanvariance Analysis on Large Asset Markets. *Econometrica*, 51(5):1281–1304.

[Chen and Zadrozny, 1998] Chen, B. and Zadrozny, P. A. (1998). An Extended Yule-Walker Method for Estimating a Vector Autoregressive Model with Mixed-Frequency. *Advances in Econometrics*, 13:47–73.

[Chen et al., 2010] Chen, W., Anderson, B. D. O., Deistler, M., and Filler, A. (2010). Solutions of Yule-Walker equations for Singular AR processes. *forthcoming in Journal of Time Series Analysis*.

[Colaneri and Longhi, 1995] Colaneri, P. and Longhi, S. (1995). The Realization Problem for Linear Periodic Systems. *Automatica*, 31(5):775–779.

[Darroch, 1965] Darroch, J. (1965). An Optimal Property of Principal components. *The Annals of Mathematical Statistics*, 36.

[Deistler, 2001] Deistler, M. (2001). *A Course in Time Series Analysis, Ed.: Pena, D.*, chapter 16, Linear Dynamic Multiinput/Multioutput Systems, pages 436–445. John Wiley, New York.

[Deistler et al., 2010a] Deistler, M., Anderson, B. D. O., Filler, A., Zinner, C., and Chen, W. (2010a). Generalized Linear Dynamic Factor Models - An approach via Singular Autoregressions. *European Journal of Control*, 16(3):211–224.

[Deistler et al., 2010b] Deistler, M., Filler, A., Anderson, B. D. O., Chen, W., and Felsenstein, E. (2010b). Singular Autoregressions for Generalized Dynamic Factor Models. *forthcoming in Conference on Decision and Control*.

[Deistler et al., 2010c] Deistler, M., Filler, A., and Funovits, B. (2010c). AR Systems and AR Processes: The Singular Case. *forthcoming in Brian Anderson special issue of Communications in Information and Systems*.

[Filler et al., 2009] Filler, A., Deistler, M., Anderson, B. D. O., Zinner, C., and Chen, W. (2009). Generalized Linear Dynamic Factor Models - An approach via Singular Autoregressions. In *European Conference of Control*.

[Forney jr., 1975] Forney jr., G. D. (1975). Minimal Bases of Rational Vector Spaces, with Applications to Multivariable Linear Systems. *SIAM Journal of Control*, 13(3):493–520.

[Forni et al., 2000] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

[Forni et al., 2005] Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, 100:830–840.

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

[Forni and Lippi, 2009] Forni, M. and Lippi, M. (2009). The unrestricted dynamic factor model. One-sided representation results. *Journal of Econometrics, forthcoming*.

[Geweke, 1977] Geweke, J. (1977). *Latent Variables in Socio-Economic Models, Ed.: Aigner, D. J.*, chapter 19, The Dynamic Factor Analysis of Economic Time Series Models, pages 365–383. North-Holland Amsterdam.

[Ghysels et al., 2006] Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies. *Journal of Econometrics*, 131(1-2):59–95.

[Ghysels et al., 2007] Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1):53–90.

[Giannone et al., 2008] Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(4):665–676.

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

[Hannan, 1980] Hannan, E. J. (1980). The Estimation of the Order of an ARMA Process. *Annals of Statistic*, 8(2):1071–1081.

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

[Ho and Kalman, 1966] Ho, B. L. and Kalman, R. E. (1966). Effective construction of linear state-variable models from input/output functions. *Regelungstechnik*, 14:545–548.

[Inouye, 1983] Inouye, Y. (1983). Modeling of Multichannel Time Series and Extrapolation of Matrix-Valued Autocorrelation Sequences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(1):45–55.

[Jolliffe, 2002] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics.

[Kailath, 1980] Kailath, T. (1980). *Linear Systems*. Prentice Hall, Englewood Cliffs, N.J.

[Lancaster and Tismenetsky, 1985] Lancaster, P. and Tismenetsky, M. (1985). *The Theory of Matrices*. 2nd Edition with Applications. Orlando: Academic Press.

[Quinn, 1980] Quinn, B. G. (1980). Order Determination for a Multivariate Autoregression. *Journal of the Royal Statistical Society*, 42(2):182 – 185.

[Robin and Smith, 2000] Robin, J. M. and Smith, R. J. (2000). Tests of Rank. *Econometric Theory*, 16:151–175.

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

[Sargent and Sims, 1977] Sargent, T. J. and Sims, C. A. (1977). Business cycle modelling without pretending to have too much a priori economic theory. *Working Paper 55, Federal Reserve Bank of Minneapolis*.

[Scherrer and Deistler, 1998] Scherrer, W. and Deistler, M. (1998). A Structure Theory for Linear Dynamic Errors-in-Variable Models. *SIAM Journal on Control and Optimization*, 36(6):2148–2175.

[Stock and Watson, 2002a] Stock, J. H. and Watson, M. W. (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the Americal Statistical Association*, 97:1167–79.

[Stock and Watson, 2002b] Stock, J. H. and Watson, M. W. (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20:147–62.

[Tummeltshammer, 2009] Tummeltshammer, C. (2009). Displacement Rank Approach for Texture Classification. Master's thesis, Vienna University of Technology.

[Whittle, 1963] Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50:129–134.

[Zadrozny, 1990] Zadrozny, P. (1990). Estimating a Multivariate ARMA Model with Mixed Frequency Data: An Application to Forecasting U.S. GNP at Monthly Intervals. *Working Paper No. 90-6, Research Department, Federal Reserve Bank of Atlanta*.

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

# Index

# Curriculum Vitae

## Alexander Filler

| | |
|---|---|
| Date of Birth: | 16. Juni 1983 |
| Place of Birth: | Mödling |
| Citizenship: | Austrian |

### Education

| | |
|---|---|
| 1993 - 2001 | Secondary School (BG/BRG Keimgasse in Mödling) |
| 2002 - 2007 | Dipl. Ing. (M.Sc.) in Technical Mathematics,<br>with focus on Economic Mathematics (Vienna University of Technology)<br>Master Thesis: Generalized Dynamic Factor Models -<br>Representation and Comparison of two Estimation Procedures<br>supervised by O.Univ.Prof. Dipl.-Ing. Dr.techn. Manfred Deistler |
| Since 11/ 2007 | Ph.D. student in Technical Mathematics (Vienna University of Technology)<br>Subject matter: Generalized Dynamic Factor Models -<br>Structure Theory and Estimation for Single Frequency and Mixed Frequency Data<br>supervised by Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Manfred Deistler |

### Professional Career

| | |
|---|---|
| 2006 | participant at the austrian Inforum project |
| 2007-2008 | Fin4Cast (Siemens AG Österreich), member of the financial forecast team |
| 2008-2009 | University of Vienna, project assistant |
| 2009-2010 | Vienna University of Technology, project assistant |
| 02/2010 | Australian National University, research |