TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

D I S S E R T A T I O N

Modeling of High-Dimensional Time Series by Generalized Dynamic Factor Models

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

O.Univ.Prof. Dipl.-Ing. Dr.techn. Manfred Deistler

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

Dipl.-Ing. Christiane Zinner
Matr.Nr.: 0026679
Seisgasse 8/11
1040 Wien

Wien, im Mai 2008

_____

Deutsche Kurzfassung

Diese Dissertation befasst sich mit der Modellierung von multivariaten Zeitreihen, bei denen die Querschnittsdimension im Verhältnis zur Anzahl der Beobachtungen relativ hoch ist. Dieses Problem ist vorallem in den letzten Jahren verstärkt ins Zentrum der Aufmerksamkeit gerückt, da in vielen Anwendungsgebieten, wie zum Beispiel in der Modellierung von Finanzzeitreihen, in der Makroökonomie oder in der Bioinformatik, vermehrt Datensätze verfügbar sind, die aus vielen unterschiedlichen Variablen bestehen. In dieser Situation sind traditionelle multivariate Zeitreihenanalysemethoden, z.B. autoregressive Modelle, ungeeignet, da die Parameterräume verglichen mit der Beobachtungszahl zu große Dimensionen aufweisen. Es werden daher Modelle benötigt, die die, sowohl im Querschnitt als auch über die Zeit, vorhandene Information komprimieren und dadurch die Komplexität reduzieren. Hierfür können Faktormodelle, die auf der Idee weniger gemeinsamer latenter Faktoren basieren, verwendet werden. In dieser Dissertation liegt der Schwerpunkt auf dem verallgemeinerten dynamischen Faktormodell (generalized dynamic factor model, GDFM), das von Forni *et al.* (2000), Forni and Lippi (2001) und Stock and Watson (1998) vorgeschlagen und analysiert worden ist.

Die Dissertation gliedert sich in folgende Teile.

In Kap. 1 werden die allgemeinen Annahmen für Faktormodelle, die in dieser Arbeit verwendet werden, eingeführt. Außerdem werden zwei klassische Modelle zur Dimensionsreduktion, das Hauptkomponenentenmodell und das Faktormodell mit idiosynkratischen Fehlern diskutiert.

In Kap. 2 wird das GDFM vorgestellt. Das GDFM verallgemeinert das (dynamische) Faktormodell mit idiosynkratischen Fehlern, bei dem angenommen wird, dass die Fehlerkomponenten unkorreliert sind, insofern als die Fehlerkompenenten hier „schwach korreliert“ sein dürfen.

Kap. 3 beschäftigt sich mit einer Strukturtheorie für GDFMs. Als erstes werden die Resultate von Forni and Lippi (2001), die eine Beziehung zwischen Eigenschaften der spektralen Dichte der Beobachtungen und der GDFM-Darstellung herstellen, zitiert. Zweitens wird die Beziehung zwischen dem Faktormodell mit idiosynkratischen Fehlern, dem Hauptkomponenentenmodell und dem GDFM analysiert. Drittens, werden die latenten Variablen (das ist jener Teil der Beobachtungen, der von den Faktoren bestimmt wird) unter der Annahme, dass diese stationär sind und ein singuläres, rationales Spektrum besitzen, näher analyisiert. Die besondere Betrachtungsweise dieser Dissertation beruht hier auf der Verwendung systemtheoretischer Methoden,

in Analogie zur Entwicklung von Strukturtheorien für ARMA- und Zustandsraumsystemen. Es wird eine minimale Zustandsraumdarstellung der latenten Variablen entwickelt und analysiert. Darauf aufbauend, wird eine minimale (quasi-)statische Darstellung des GDFM hergeleitet, wobei der statische Faktor i.A. ein ARMA-Prozess mit singulärem Spektrum ist. Für diesen ARMA-Prozess wird dann eine eindeutig identifizierbare Darstellung angegeben.

Kap. 4 handelt von der Prognose im GDFM. Es wird gezeigt, dass die Prognose der latenten Variablen auf die Prognose der statischen Faktoren zurückgeführt werden kann. Letztere wird dann mithilfe der ARMA-Darstellung aus dem vorherigen Kapitel berechnet.

Die Schätzung eines GDFM wird in den Kap. 5 bis 7 behandelt. In Kap. 5 werden mehrere, in der Literatur beschriebene Methoden zur Schätzung der latenten Variablen bzw. der Faktoren und Faktorladungen präsentiert und ihre asymptotischen Eigenschaften diskutiert. Beispiele für Schätzmethoden sind die dynamische bzw. die statische Hauptkomponentenanalyse, siehe Forni *et al.* (2000) bzw. Stock and Watson (1998). In Kap. 6 schlagen wir vor, das ARMA-Modell des statischen Faktors mittels eines „Autoregression-Regression"-Ansatzes zu schätzen und leiten für diesen Schätzer Konsistenzresultate her. In Kap. 7 geht es um die Selektion eines bestimmten GDFM. Insbesondere geht es also um die Schätzung der Faktorenanzahl, für die es eine Reihe von Methoden gibt (siehe z.B. Bai and Ng (2002) und Bai and Ng (2007)) und die Schätzung der strukturellen Parameter, durch die die ARMA-Darstellung des statischen Faktors eindeutig festgelegt wird.

In Kap. 8 präsentieren wir dann eine Monte-Carlo-Simulationsstudie. In dieser Studie wird die Prognosequalität von Prädiktoren, basierend auf den hier vorgeschlagenen Schätzern, mit den „üblichen" GDFM-Prädiktoren, basierend auf einer AR(1)-Darstellung der statischen Faktoren, verglichen.

Diese Arbeit wurde teilweise finanziert durch den österreichischen Fonds zur Förderung der wissenschaftlichen Forschung, FWF, Projektnummern P17065 und P17378.

Abstract

This thesis is concerned with the modeling of multivariate time series, when the cross-sectional dimension is relatively large compared to sample size. This problem has recently attracted substantial interest, since in many areas of application, such as finance, macroeconomics or bioinformatics, data sets consisting of a large number of variables have become increasingly available. In this context, traditional multivariate time series modeling, e.g. by autoregressive models, would result in parameter spaces of excessively high dimension compared to sample size. Therefore models compressing the information contained in both, the time and the cross-sectional dimension and thus reducing the complexity, are needed. For this purpose, factor models, exploiting the idea of a few underlying latent factors, may be used. In this thesis, the focus is on generalized dynamic factor models (GDFMs) as proposed and analyzed by Forni *et al.* (2000), Forni and Lippi (2001) and Stock and Watson (1998).

The thesis is organized as follows.

In Chapter 1 the general factor model framework used throughout this thesis is introduced. Besides, two classical model types used for dimension reduction, the principal component model and the factor model with idiosyncratic noise, respectively, are discussed.

Subsequently in Chapter 2, the GDFM is introduced and important concepts are presented . The GDFM generalizes the factor model with idiosyncratic noise, where the noise components are assumed to be uncorrelated, in that the noise components are allowed to be "weakly correlated".

In Chapter 3 we are concerned with a structure theory for GDFMs. First, main results of Forni and Lippi (2001) relating properties of the spectral density of the observations to the GDFM representation are reported. Second, the relation between the factor model with idiosyncratic noise, the principal component model and the GDFM is analyzed. Third we develop a structure theory for the latent variables (i.e. the part of the observations driven by the factors) under the assumption that the latent variables are stationary with singular, rational spectral density. The particular point of view in this thesis is the use of system theoretic methods in analogy to to the structure theory for state space and ARMA systems. Here we develop and analyze a minimal state space representation of the latent variables, and based on that a minimal (quasi-)static representation of the GDFM, where the static factor is in general given as an ARMA process with possibly singular spectral density. For this ARMA process, a uniquely identifiable ARMA

representation is then derived.

Chapter 4 deals with prediction in the GDFM. It is shown that the prediction of the latent variable reduces to the prediction of the minimal static factor. The latter can then be accomplished using its ARMA representation educed in the previous chapter.

Estimation of the GDFM is dealt with in Chapters 5 to 7. In Chapter 5 a number of methods proposed in the literature for the estimation of the latent variables, and the factors and factor loadings, respectively, are presented and their asymptotic properties are discussed. Examples are estimators based on dynamic PCA or static PCA, see Forni *et al.* (2000) and Stock and Watson (1998), respectively. In Chapter 6, we propose to estimate the ARMA representation of the static factor using an "autoregression-regression" approach, and derive consistency results for these estimates. In Chapter 7 we are concerned with the estimation of several integer-valued structural parameters specifying a GDFM (model selection problem). In particular these are the numbers of dynamic and static factors, for which a number of estimation methods exist (see e.g. Bai and Ng (2002) and Bai and Ng (2007)), and the structural indices specifying the unique ARMA representation of the static factors.

In Chapter 8 we then present a Monte Carlo simulation study, comparing the forecasting quality using predictors based on the estimators proposed in this thesis with the standard GDFM-predictors, that base on an AR(1)-representation of the static factors.

This work was partly funded by the Austrian science fund, FWF, project numbers P17065 and P17378.

Acknowledgements

I may not have gone where I intended to go, but I think I have ended up where I needed to be.
(Douglas Adams)

Since this thesis is largely devoted to the analysis of latent factors, I am gladly taking the opportunity to mention and thank some of the "human factors" who have − directly or indirectly − contributed to this thesis.

In the first place I would like to thank my supervisor Prof. Manfred Deistler. Not only because he prevented me from becoming a tax advisor (!) by convincing me of the sheer necessity of writing a thesis, but especially for his great support and encouragement during the past couple of years. I have enormously benefited from his vast knowledge of econometrics and system theory (not to mention a lot of other things!) and our many fruitful discussions.

I would also like to thank Prof. Marco Lippi, who during his visit in Vienna took plenty of time to answer my questions. Although we have not met personally yet, Prof. Brian D.O. Anderson contributed many influential ideas to this thesis. Thanks also to Prof. Engelbert Dockner for employing me and thus granting me the opportunity to finish my thesis without worrying to much about my bank account, and for the interest he took in my work.

I would like to express my deepest gratitude to my mother, whose pragmatical nature and excitement for engineering were very influential and to my father, who has, ever since I was a little child, taught me to be curious and interested in all matters of life and whose advice I highly appreciate. I want to thank my sister for being such an inspiring person and a really good friend. Last, but not least, I would like to thank my boyfriend Thomas for his love and support in any situation of life.

Contents

Chapter 1

Introduction

> *Three rules of work: Out of clutter find simplicity, from discord find harmony, in the middle of difficulty lies opportunity.* (Albert Einstein)

Reliable models explaining the behaviour of some observed time series are needed for many purposes, e.g. for identifying relevant impact factors and understanding underlying processes in an economy or for prediction. During the last decades, data sets consisting of a large number of time series have become increasingly available. Examples are macroeconomic time series, which are observed for many different countries or financial time series, which are available for merely innumerable assets. The variables in these data sets are typically highly correlated, such that univariate modeling would seem to be a waste of information (see Tiao (2001)). However, traditional multivariate time series modeling, for instance by autoregressive (AR) models, is plagued by the so called *curse of dimensionality*: the complexity of the model class (i.e. the dimension of the parameter space) shows quadratic dependence on the number of variables, n say, whereas the number of observed data points, for fixed sample size, T say, is linear in n. A standard alternative is to choose a few variables of the data set as target variables or outputs, select among the others the most relevant subset of inputs and set up a small model, e.g. an autoregressive model with exogenous inputs (ARX). But in the context of large data sets the so called *data snooping* problem (see White (2000)) arises. A large number (relative to sample size) of potential inputs and input combinations induces the risk of overfitting, apart from the computational burden of finding an optimal input set. Moreover, as stated in Heij *et al.* (1997), the ARX setting, where the inputs are not modeled, is genuinely inappropriate if the inputs are themselves noisy or if there is uncertainty about the classification of the data into inputs and outputs. Either case requires a more symmetric way of modeling. For all of these reasons multivariate time series models that can exploit the information contained in large data sets, but are still parsimonious, are needed. This is why (dynamic) factor models have been brought into play.

In general, factor models are used to explain the variability of high dimensional data consisting of many variables in terms of a few unobserved common factors. Based on the idea that the observations are driven by a few common factors, factor models are compressing the information

contained in high-dimensional data and thus reducing the complexity. Factor models allow of a sparse parametrization even when the number of variables is large. Since they are symmetric, in that there is no distinction between observed inputs and outputs, the problems related to input selection and the modeling of noisy inputs are (at least partly) avoided. Hence, factor models may be used to mitigate all of the described problems and have therefore attracted substantial interest during the past couple of years.

Certainly factor models are not new, indeed *factor analysis* has been first devised in the field of psychology in the beginning of the twentieth century. Particularly, Burt (1909) and Spearman (1904) observing that in mental ability tests the scores on seemingly unrelated subjects were positively correlated, postulated the hypothesis of a common latent factor, called *general intelligence*. Later this idea was extended, allowing for more than one factor, representing different mental abilities, see Thurstone (1932). Around the same time, another dimension reduction method, *principal component analysis*, has been introduced by Pearson (1901) and Hotelling (1933), with first applications in the field of biology, leading to principal component models.

Whereas the initial factor models were oriented to data originating from independent, identically distributed random variables (static factor models), the idea has been generalized to the modeling of multivariate time series (dynamic factor models). This development took place independently in different areas, such as signal processing (Brillinger (1981)), and econometrics (Geweke (1977), Sargent and Sims (1977) and Engle and Watson (1981)).

Besides in the 1980's, Chamberlain and Rothschild (1983) and Chamberlain (1983) introduced the *generalized static factor model* (then called *approximate factor model*) by weakening the original assumption of uncorrelatedness (idiosyncrasy) of the noise components. Today, the state of the art is the *generalized dynamic factor model* (GDFM) as proposed and analyzed by Forni *et al.* (2000), Forni and Lippi (2001) and Stock and Watson (1998). This model, that constitutes the main area of this thesis, generalizes and combines the dynamic factor model (with idiosyncratic noise) and the generalized static factor model.

In Section 1.1 below we are going to present the general factor model framework used throughout this thesis. In the remainder of this introductory chapter we will then present the two classical models mentioned above, i.e. the principal component model (Section 1.2) and the factor model with idiosyncratic noise (Section 1.3).

1.1 General framework

We consider an n-dimensional vector valued process of observations $(y_t),\ t \in \mathbb{Z}$, that is driven by two sources of variation that are in general unobserved: a q-dimensional factor process $(\xi_t),\ t \in \mathbb{Z}$, $(q << n)$, that is common to all n components of (y_t) and an n-dimensional noise

process $(u_t),\ t \in \mathbb{Z}$. Then y_t can be represented by a factor model equation of the form

$$y_t = \Lambda(z)\xi_t + u_t, \qquad t \in \mathbb{Z}, \tag{1.1.1}$$

where the transfer function $\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j,\ \Lambda_j \in \mathbb{R}^{n\times q}$ is called the *factor loading matrix* and z is used (throughout this thesis) both for a complex variable and for the backward shift operator on $\mathbb{Z}$, i.e. $z^j\xi_t = \xi_{t-j}$. Throughout $\Lambda(z)\xi_t =: \chi_t$ will be called the *latent variable*. The factor model (1.1.1) considered here is thus *linear*, *dynamic* and *time invariant*, but in general need not be *causal*.

Throughout we assume that

Assumption 1.1 (General Assumptions)**.**

a) The processes (ξ_t) and (u_t) are wide sense stationary with real valued entries, $\mathbb{E}\,\xi_t = 0$, $\mathbb{E}\,u_t = 0$ for all $t \in \mathbb{Z}$ and with covariances[1] $\Gamma_\xi(s) = \mathbb{E}\,\xi_t\xi'_{t-s}$ and $\Gamma_u(s) = \mathbb{E}\,u_t u'_{t-s}$ satisfying the summability conditions

$$\sum_{s=-\infty}^{\infty} \|\Gamma_\xi(s)\| < \infty, \sum_{s=-\infty}^{\infty} \|\Gamma_u(s)\| < \infty, \tag{1.1.2}$$

where $\|\ \|$ denotes a matrix norm.

b) The factor process (ξ_t) and the noise process (u_t) are mutually orthogonal at any leads and lags, i.e. $\mathbb{E}\,\xi_t u'_{t-k} = 0$ for all $t,\ k \in \mathbb{Z}$.

c) The coefficients Λ_j are absolutely summable, $\sum_{j=-\infty}^{\infty} \|\Lambda_j\| < \infty$.

d) The spectral density f_χ of (χ_t) is rational in $e^{-i\lambda}$ and of rank q for λ a.e. in $[-\pi, \pi]$.

e) The covariance $\Gamma_\chi(0)$ of (χ_t) is of rank $s \geq q$.

f) Each of the spectral densities f_y, f_χ and f_u of (y_t), (χ_t) and (u_t) respectively, has distinct eigenvalues.

Assumptions 1.1 a) and 1.1 c) imply that the spectral densities of (χ_t) and (u_t) exist and that (y_t) is a zero-mean stationary process with spectral density f_y. Hence the focus here is on stationary processes and in case of trend or difference stationary time series occurring in applications the trend will have to be eliminated in a prior step. Assumption 1.1 b) is standard in factor analysis and needed for identification of the latent variables and the noise. Then Assumption 1.1 b) implies that f_y can be written as[2]

$$\begin{aligned} f_y(\lambda) &= \Lambda(e^{-i\lambda}) f_\xi(\lambda) \Lambda(e^{-i\lambda})^* + f_u(\lambda) \\ &= f_\chi(\lambda) + f_u(\lambda), \end{aligned} \tag{1.1.3}$$

[1] Throughout this thesis A' will denote the transpose of a matrix A.

[2] Throughout this thesis the asterisk will denote the conjugate transpose, i.e. $\Lambda(e^{-i\lambda})^* = \Lambda(e^{i\lambda})'$.

where by Assumption 1.1 d) the first term on the right hand side, i.e. f_χ, is rational in $e^{-i\lambda}$ and of reduced rank q. Note that since every minor of f_χ is rational, each minor is either equal to zero or it only vanishes at finitely many points, and hence the rank of f_χ is constantly equal to q a.e. in $[-\pi, \pi]$. The assumption that f_χ is rational is a proper restriction of generality. The main reasons for its imposition are, as will be discussed in detail in Section 3.3, that it allows of a *causal* representation of (χ_t) and parametric modeling. Notice still that f_u is not restricted to be rational and that so far, the noise (u_t) may be any stationary process including correlated (over the cross-sectional dimension) and auto-correlated processes. Assumption 1.1 e) may be seen as the static analogue of Assumption 1.1 d), since both refer to the possible dimension reduction in the cross-sectional dimension. Assumption 1.1 f) is again a standard assumption imposed to avoid unnecessary complications related to eigenspaces of higher dimension than 1.

A special case often considered occurs when $\Lambda(z) = \Lambda$ is constant and (ξ_t) and (u_t) and thus (y_t) are white noise. In this case (1.1.1) is called *static* and the variance matrix of y_t, $\Gamma_y = \Gamma_y(0)$, is of the form:

$$\Gamma_y = \Lambda\Gamma_\xi\Lambda' + \Gamma_u \tag{1.1.4}$$

$$= \Gamma_\chi + \Gamma_u. \tag{1.1.5}$$

If $\Lambda(z) = \Lambda$ is constant, but (ξ_t) and (u_t) are not necessarily white noise, the model is sometimes called *quasi-static*.

Assumption 1.1 is not sufficient to determine a reasonable model class, in the sense that for given f_y, or Γ_y respectively, too many decompositions into latent variables and noise would be possible, see for instance Heij *et al.* (1997). Thus, in order to obtain reasonable model classes, further assumptions have to be imposed. This leads to principal component models, linear factor models with idiosyncratic noise and generalized linear factor models considered in this thesis.

1.2 The principal component model

The aim of principal component analysis (PCA) is to approximate the n-dimensional observed process (y_t) by a filtered version of itself, whose spectral density is of reduced rank q, such that the filtered version retains most of the original variance or, in other words, such that the variance of the residuals is minimized (see Brillinger (1981)). Geometrically, the n-dimensional observations y_t are projected onto a lower dimensional space spanned by the so called *principal components* of (y_t), defined such that the variance of the perpendiculars is minimized. It is emphasized that the focus in PCA is hence data compression or dimension reduction and there does not need to be an underlying factor structure.

Commencing from the n-dimensional observed process (y_t) with spectral density f_y (or covariance Γ_y in the static case) we want to find a $(q \times n)$ filter $B(z) = \sum_j B_j z^j$, defining a

q-dimensional linear transformation of y_t,

$$\psi_t = B(z)y_t,$$

and an $(n \times q)$ filter $C(z) = \sum_j C_j z^j$, such that y_t can be expressed as

$$\begin{aligned} y_t &= C(z)\psi_t + \tilde{u}_t \\ &= \quad \tilde{\chi}_t \quad + \tilde{u}_t \end{aligned} \tag{1.2.1}$$

and such that

$$\operatorname{tr}\left(\mathbb{E}(\tilde{u}_t\tilde{u}_t')\right) = \operatorname{tr}\left(\mathbb{E}(y_t - C(z)\psi_t)(y_t - C(z)\psi_t)^*\right) \tag{1.2.2}$$

is minimal with respect to $B(z)$, $C(z)$ for fixed q, where throughout tr will denote the trace of a matrix. The q-dimensional process (ψ_t) (formed by the principal components of (y_t)) may however be interpreted as a factor process in the principal component (PC) model (1.2.1).

1.2.1 Static or quasi-static PCA

Let us first consider the static (or quasi-static) case, where $B(z) = B$ and $C(z) = C$ are constant matrices.

Since the trace of a matrix equals the sum of its eigenvalues, minimization of (1.2.2) is equivalent to the simultaneous minimization of all eigenvalues of $\mathbb{E}\,\tilde{u}_t\tilde{u}_t'$. The solution of this minimization problem is then obtained via the canonical representation (see (A.1.1)) of Γ_y, decomposed as

$$\Gamma_y = O_1\Omega_1O_1' + O_2\Omega_2O_2', \tag{1.2.3}$$

where Ω_1 and Ω_2 denote the diagonal matrices containing the q largest and $(n-q)$ smallest eigenvalues of Γ_y, respectively, arranged in descending order of magnitude and $O_1 = (o_1 \dots o_q)$ and $O_2 = (o_{q+1} \dots o_n)$ are the $(n \times q)$- and $n \times (n-q)$-dimensional orthogonal matrices, respectively, of corresponding normalized eigenvectors. Minimizers B and C of (1.2.2) are, as will be shown below, given by

$$B = O_1', \quad C = O_1. \tag{1.2.4}$$

With (1.2.4) and since by orthogonality of the eigenvectors we have $O_1O_1' + O_2O_2' = I_n$, the static PC model is defined as

$$\begin{aligned} \psi_t &= O_1'y_t, \\ \tilde{\chi}_t &= O_1\psi_t = O_1O_1'y_t, \\ \tilde{u}_t &= y_t - O_1O_1'y_t = O_2O_2'y_t, \\ \tilde{\Gamma}_\chi &= O_1\Omega_1O_1', \quad \tilde{\Gamma}_u = O_2\Omega_2O_2'. \end{aligned} \tag{1.2.5}$$

The q orthogonal scalar processes forming the q-dimensional process (ψ_t), are called the *principal components* of (y_t).

In the sequel we are going to show, that (1.2.4) are indeed minimizers. First observe that in the static case the objective function (1.2.2) can be written as

$$\operatorname{tr}(I_n - CB)\Gamma_y(I_n - CB)' = \operatorname{tr}(\Gamma_y^{1/2} - CB\Gamma_y^{1/2})(\Gamma_y^{1/2} - CB\Gamma_y^{1/2})'. \tag{1.2.6}$$

Second, Lemma A.1.2 implies, that the minimum of (1.2.6) is achieved, if

$$CB{\Gamma_y}^{1/2} = O_1{\Omega_1}^{1/2}O_1' \tag{1.2.7}$$

holds. Hence post multiplying both sides of the last expression by ${\Gamma_y}^{-1/2}$ we obtain, that in the minimum

$$CB = O_1{\Omega_1}^{1/2}O_1'{\Gamma_y}^{-1/2} = O_1{\Omega_1}^{1/2}O_1'(O_1{\Omega_1}^{-1/2}O_1' + O_2{\Omega_2}^{-1/2}O_2) = O_1O_1' \tag{1.2.8}$$

holds. Concluding we see that by choosing B and C as in (1.2.4) the eigenvalues of $(\Gamma_y^{1/2} - CB\Gamma_y^{1/2})(\Gamma_y^{1/2} - CB\Gamma_y^{1/2})'$ are minimized and hence a minimum of (1.2.6) is achieved.

Calling $\tilde{\chi}_t = O_1\psi_t$ and $\tilde{u}_t$ the latent variables and the noise respectively, in the PC model, we see that since $O_1'O_2 = 0$, $\tilde{u}_t \perp \psi_t$ and $\tilde{u}_t \perp \tilde{\chi}_t$, hence the decomposition into latent variables and noise corresponds to an orthogonal projection. Since the orthogonal projection is unique, the latent variables and the noise respectively are uniquely identifiable. Furthermore decomposition (1.2.3) is unique and corresponds to the decomposition of the variance into the part explained by the principal components and the noise respectively. In contrast, the matrices B and C are only identifiable up to pre multiplying B by any non-singular $(q \times q)$ matrix P and post multiplying C by P^{-1}, which has no impact on $\tilde{\Gamma}_\chi$ and $\tilde{\chi}_t$. By choosing B and C as in (1.2.4) we have adopted the normalization condition $C'C = I_q$.

1.2.2 Dynamic PCA

In the dynamic case, where $B(z)$ and $C(z)$ are in general two-sided filters, we may write (1.2.2) as

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}(I - C(e^{-i\lambda})B(e^{-i\lambda})f_y(\lambda)(I - C(e^{-i\lambda})B(e^{-i\lambda})^* d\lambda, \tag{1.2.9}$$

which is minimized if $\operatorname{tr}(I - C(e^{-i\lambda})B(e^{-i\lambda})f_y(\lambda)(I - C(e^{-i\lambda})B(e^{-i\lambda})^*$ is minimized for each λ. Proceeding analogously to the static case, the minimization of (1.2.2) is now based on the eigenvalue decomposition of the spectral density matrix $f_y(\lambda)$, for $\lambda \in [-\pi, \pi]$:

$$f_y(\lambda) = O_1(\lambda)\Omega_1(\lambda)O_1(\lambda)^* + O_2(\lambda)\Omega_2(\lambda)O_2(\lambda)^* \tag{1.2.10}$$

where the matrices $\Omega_1(\lambda)$, $\Omega_2(\lambda)$, $O_1(\lambda)$ and $O_2(\lambda)$ are defined analogously to the static case. Minimizers of (1.2.9) are then given by

$$B(e^{-i\lambda}) = O_1(e^{-i\lambda})^*, \quad C(e^{-i\lambda}) = O_1(e^{-i\lambda}), \tag{1.2.11}$$

where the $(n \times q)$-dimensional matrix function in $e^{-i\lambda}$, $O_1(e^{-i\lambda})$, is defined via the Fourier series expansion of $O_1(\lambda)$,

$$O_1(\lambda) = \sum_j \underline{O}_{1j} e^{-i\lambda j} = O_1(e^{-i\lambda}), \tag{1.2.12}$$

where the Fourier coefficients $\underline{O}_{1j}$ are determined as

$$\underline{O}_{1j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} O_1(\lambda) e^{i\lambda j} d\lambda. \tag{1.2.13}$$

With the minimizers (1.2.11), and since by orthogonality of the eigenvectors $O_1(e^{-i\lambda})O_1(e^{-i\lambda})^* + O_2(e^{-i\lambda})O_2(e^{-i\lambda})^* = I_n$, where $O_2(e^{-i\lambda})$ is defined in an analogous way, the variables in the dynamic PC model are:

$$\begin{aligned} \psi_t &= O_1(z)^* y_t, \\ \tilde{\chi}_t &= O_1(z)\psi_t = O_1(z)O_1(z)^* y_t, \\ \tilde{u}_t &= y_t - O_1(z)O_1(z)^* y_t = O_2(z)O_2(z)^* y_t, \\ \tilde{f}_\chi(\lambda) &= O_1(\lambda)\Omega_1(\lambda)O_1(\lambda)^*, \ \tilde{f}_u(\lambda) = O_2(\lambda)\Omega_2(\lambda)O_2(\lambda)^*. \end{aligned} \tag{1.2.14}$$

The q orthogonal scalar processes forming the q-dimensional process (ψ_t), are now called the dynamic principal components of (y_t).

Concerning identifiability the results from the static PC model may be taken over to the dynamic case: the decomposition of y_t into latent PC variables $\tilde{\chi}_t$ and PC noise $\tilde{u}_t$ (and of f_y into $\tilde{f}_\chi$ and $\tilde{f}_u$ respectively) is uniquely identifiable, whereas $B(z)$ and $C(z)$ are only identifiable up to regular transformations. By choosing $B(z)$ and $C(z)$ as in (1.2.11) we adopted the normalization condition $C(z)^*C(z) = I_q$.

To ensure that the expressions in (1.2.14) are meaningful in the sense that all infinite sums converge and hence the PC variables exist as stationary limits, we need the following results due to Brillinger (1981) and Forni and Lippi (2001), that show that $O_1(e^{-i\lambda})$ and $O_2(e^{-i\lambda})$ in (1.2.14) are indeed transfer functions with absolutely summable corresponding filters.

Lemma 1.2.1. *Under Assumption (1.1) the eigenvalues of f_y, $\omega_j : [-\pi, \pi] \to \mathbb{R}$, $j = 1, \ldots, n$, are Lebesgue-measurable and integrable in $[-\pi, \pi]$.*

Proof. First, recall that under Assumption (1.1) f_y is uniformly continuous and bounded (see Brillinger (1981), page 23) and therefore Lebesgue-measurable. Since $\int_\pi^{-\pi} f_y(\lambda) d\lambda = \Gamma_y(0)$ and y_t is stationary, it is also integrable. Eigenvalues are continuous functions of the matrix elements (which is equivalent to the statement that the roots of a polynomial are continuous functions of the coefficients, see e.g. Tyrtyshnikov (1997)) and therefore continuous with respect to λ. Since continuous functions of a measurable function are themselves measurable, see e.g. Royden

(1988), eigenvalues are measurable functions. They are integrable since for $1 \leq j \leq n$

$$0 \leq \omega_j(\lambda) \leq \sum_{k=1}^{n} \omega_k(\lambda) = \mathrm{tr}(f_y(\lambda))$$

and

$$\int_{\pi}^{-\pi} \mathrm{tr}(f_y(\lambda))d\lambda = \mathrm{tr}(\Gamma_y(0))$$

hold. ■

Lemma 1.2.2. *Under Assumption (1.1) the normalized eigenvectors of f_y, $o_j : [-\pi, \pi] \to \mathbb{R}^{n\times 1}$, $j = 1, \ldots, n$, are Lebesgue-measurable in $[-\pi, \pi]$.*

Proof. Given the eigenvalues $\omega_j(\lambda),\ j = 1, \ldots, n$, of $f_y(\lambda),\ \lambda \in [-\pi, \pi]$, the corresponding normalized eigenvectors $o_j(\lambda),\ j = 1, \ldots, n$, solve the following equations:

$$\begin{aligned} (f_y(\lambda) - \omega_j(\lambda)I_n)\, o_j(\lambda) &= 0 \\ o_j(\lambda)^* o_k(\lambda) &= 0,\ k = 1, \ldots, j-1,\ j > 1 \\ s.t. \qquad o_j(\lambda)^* o_j(\lambda) &= 1 \end{aligned} \tag{1.2.15}$$

Since f_y and ω_1 are measurable, for $j = 1$ the coefficients of (1.2.15) are measurable. Hence, for $j = 1$ the solution of (1.2.15) is measurable since it is a continuous function of the coefficients. By recursion, for $j = 2, \ldots, n$ the coefficients and thus the solutions of (1.2.15) are measurable. ■

As a consequence of Lemma 1.2.2 and since $O_1(\lambda)^* O_1(\lambda) = I_q$, the integral on the right hand side of (1.2.13) converges and hence the filter coefficients corresponding to the eigenvectors of f_y are well defined. Finally, the next lemma ensures that the filter coefficients are absolutely summable.

Lemma 1.2.3. *Under Assumption (1.1),*

$$\begin{aligned} \sum_j \|\underline{O}_{1j}\| &< \infty, \\ \sum_j \|\underline{O}_{2j}\| &< \infty, \end{aligned} \tag{1.2.16}$$

where $\underline{O}_{1j},\ j \in \mathbb{Z}$ is defined as in (1.2.13) and $\underline{O}_{2j},\ j \in \mathbb{Z}$ is defined analogously with $O_2(\lambda)$ in place of $O_1(\lambda)$.

Proof. See Brillinger (1981), Theorem 9.3.3. (for $P = 0$). ■

1.2.3 Estimation

For estimation of the PC model analog estimators are employed: The population second moments, Γ_y and f_y in (1.2.3) and (1.2.10) respectively, are replaced by their estimators - typically the sample covariance and a non-parametric estimator of the spectral density. As can be shown

under general assumptions, consistency of these estimators yields consistent estimators of Λ, ψ_t, $\tilde{\chi}_t$ and $\Gamma_{\tilde{u}}$ from (1.2.5) and $\Lambda(z)$, ψ_t, $\tilde{\chi}_t$ and $f_{\tilde{u}}$ from (1.2.14), respectively.

Choice of the number of principal components

In the PC model an underlying factor structure is not necessarily assumed. Consequently, the number of principal components q is not intrinsic in the sense, that it is not a property of Γ_y or f_y. By the choice of q, the degree of dimension reduction and, as a trade-off, the quality of approximation are determined. In dynamic PCA dimension reduction in the time dimension is performed by introducing a finite dimensional parametrization. Note, however, that even for a rational spectral density f_y, the matrices on the right hand side of (1.2.10) are not necessarily rational.

1.3 The factor model with idiosyncratic noise

Here, in addition to Assumption 1.1, it is assumed that the noise components are uncorrelated, i.e. that f_u (or in the static and quasi-static case Γ_u) is diagonal. In other words, it is assumed that the noise process (u_t) does not influence the correlation structure of the observed process (y_t) and is therefore called *idiosyncratic*. The main idea in classical factor analysis is thus to separate the latent variable described by the factors from the idiosyncratic noise. The factors here have a splitting property: For given factors, the components of (y_t) are conditionally uncorrelated.

1.3.1 Static or quasi-static factor model with idiosyncratic noise

Commencing from given Γ_y and number of factors q, we see from equation (1.1.4), i.e. $\Gamma_y = \Lambda\Gamma_\xi\Lambda' + \Gamma_u$, that there are two identifiability problems. The first is to identify pairs of matrices (Γ_χ, Γ_u), where Γ_χ is of rank q, positive semidefinite and symmetric, Γ_u is diagonal and positive semidefinite, and where (1.1.4) holds. The second is to identify matrices Λ and Γ_ξ corresponding to Γ_χ. Here we will assume that $\Gamma_\xi = I_q$, such that Λ is identifiable up to right multiplication by orthogonal matrices, i.e. up to (factor) rotations.

As far as the first problem is concerned, an inspection of both sides of (1.1.4) reveals that $\frac{1}{2}n(n+1)$ single equations (due to the symmetry of Γ_y) are used to determine $n + nq - \frac{1}{2}q(q-1)$ functionally independent parameters, where n is the number of free parameters in Γ_u and $nq - \frac{1}{2}q(q-1)$ is the number of free parameters in Γ_χ, which results from the fact that the orthogonality of any post multiplication of Λ already entails $\frac{1}{2}q(q-1)$ restrictions. For given n, the solution of

$$\frac{1}{2}n(n+1) - n + nq - \frac{1}{2}q(q-1) = 0$$

is called the *Lederman bound* and is explicitly given as $q_{max} = \frac{2n+1}{2} - \sqrt{\frac{(2n+1)^2}{4} - n^2 + n}$. Hence in general, it follows that if $q > q_{max}$, then Γ_χ and Γ_u will not be uniquely identifiable. But it

has been shown that Γ_χ and Γ_u are generically identifiable if $q \leq q_{max}$ (see Scherrer and Deistler (1998)).

Prior to estimation, a normalization condition for Λ has to be defined. Therefore consider a rescaling of the observations such that the noise variance becomes I_n, hence

$$\bar{\Gamma}_y = \Gamma_u^{-1/2}\Gamma_y\Gamma_u^{-1/2}\Gamma_y = \Gamma_u^{-1/2}\Lambda\Lambda'\Gamma_u^{-1/2} + I_n.$$

Then the matrix $\bar{\Gamma}_y - I_n$ is symmetric and of rank q, and may therefore be expressed by means of its canonical decomposition, i.e.

$$\bar{\Gamma}_y - I_n = \bar{V}\Delta\bar{V}',$$

where Δ is the diagonal matrix whose diagonal entries are the q non-zero eigenvalues of $\bar{\Gamma}_y - I_n$ and $\bar{V}$ is the matrix of corresponding normalized eigenvectors. Hence we may uniquely (apart from possible sign changes of its columns) define Λ by

$$\Lambda = \Gamma_u^{1/2}\bar{V}\Delta^{1/2}.$$

Or in other words we may choose Λ such that it satisfies the normalization condition

$$\Lambda'\Gamma_u^{-1}\Lambda = \Delta, \tag{1.3.1}$$

where Δ is a diagonal matrix whose positive, distinct elements are arranged in descending order of magnitude.

Estimation

In the case that $q \leq q_{max}$, estimators of Λ and Γ_u may be obtained from (iteratively) maximizing the Gaussian log-likelihood function, i.e. omitting a constant

$$L_T(\Lambda, \Gamma_u|\hat{\Gamma}_y) = \frac{T}{2}\log\det(\Lambda\Lambda' + \Gamma_u) - \frac{T}{2}\operatorname{tr}((\Lambda\Lambda' + \Gamma_u)^{-1}\hat{\Gamma}_y)$$

subject to $\Lambda \in \mathbb{R}^{n\times q}$, $\operatorname{rk}\Lambda = q$, $\Gamma_u > 0$ and a suitable normalization condition on Λ (e.g. (1.3.1)) guaranteeing uniqueness , see Lawley and Maxwell (1971), Chapter 4 for a detailed description. In the static case the ML-estimators can be shown to be consistent under general assumptions, see Anderson (1984). In the quasi static case $L_T(\Lambda, \Gamma_u|\hat{\Gamma}_y)$ is no longer the log-likelihood function - hence we may rather speak of a *quasi-ML* estimation - but nevertheless yields consistent estimators.

In contrast to the PC model, here the factors ξ_t are in general not functions of the observations, but have to be approximated by some (linear) function of y_t. This can be affected by applying the *regression method* (see Thomson (1951)), where the factors are approximated by a static transformation of y_t that minimizes the sum of squared errors, or *Bartlett's method* (see

Bartlett (1937)), where also the sum of squared errors is minimized, but the errors are rescaled by multiplication by $\Gamma_u^{-1/2}$.

Another difference to the PC model is that here q (i.e. the minimal q) is intrinsic, in the sense that it is a property of Γ_y. Tests for determining q have been proposed for instance by Anderson and Rubin (1956).

1.3.2 Dynamic factor model with idiosyncratic noise

The answer to the dynamic analogue of the first identifiability question above is, that, for given spectral density f_y, the spectra $f_\chi = \Lambda(e^{-i\lambda}) f_\xi(\lambda) \Lambda(e^{-i\lambda})^*$ and f_u are generically identifiable for $q \leq n - \sqrt{n}$ (see Scherrer and Deistler (1998)).

If $q \leq n - \sqrt{n}$ then $\Lambda(z)$ and f_u can be estimated under a suitable normalization condition and some additional structural assumptions by maximum likelihood estimation using the Kalman filter and the factors ξ_t can be estimated using the Kalman smoother (see Engle and Watson (1981) and Watson and Engle (1983)).

Tests for determining q in the dynamic case have been proposed by Geweke (1977) and Sargent and Sims (1977).

Chapter 2

The generalized dynamic factor model

For many applications the classical assumption of uncorrelatedness of the noise components, or to be more precise that f_u (or in the static case Γ_u) is diagonal, turns out to be too restrictive. "Local" dependency between the noise components (e.g. between asset prices of the same industrial sector or macroeconomic data from related countries) may occur, which makes it unlikely that a classical factor structure with a reasonably small number of common factors exists.

Moreover, in a number of applications, e.g. in cross-country business cycle analysis or asset pricing, the cross-sectional dimension (i.e. the number of observed time series) may be high, possibly exceeding sample size, obstructing the use of traditional time series methods and in turn increasing the probability of local dependence between noise components.

These two issues have lead to the development of *generalized factor models*, where the uncorrelatedness of the noise component is replaced by some weaker assumption. In the static case, generalized factor models (then called *approximate factor models*) were first introduced and analyzed by Chamberlain and Rothschild (1983) and Chamberlain (1983) in the context of asset pricing. The further generalization to the dynamic case and hence to time series modeling has been made by Forni and Lippi (2001), Forni *et al.* (2000) and Stock and Watson (1998). Weakening the assumption of uncorrelatedness, however, implies that the generalized factor model, in contrast to the factor model with idiosyncratic noise, is only identifiable asymptotically, i.e. for infinite cross-sectional dimension. At the same time, because of this genuine asymptotic framework, the model is especially applicable to large data sets and most remarkable, even for small sample sizes additional information can be gained from adding more time series.

In the subsequent section we will give a precise formulation of the generalized dynamic factor model following to a great extent Forni and Lippi (2001).

2.1 Definition of the generalized dynamic factor model and properties of weakly correlated noise

We consider a double sequence

$$(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z}),$$

where the general Assumptions 1.1 hold true for every n-dimensional vector process $(y_t^n) = ((y_{1t}, y_{2t}, \ldots, y_{nt})'|\ t \in \mathbb{Z})$, if n is sufficiently large. In order to render explicit the dependence on n, the previously defined symbols may be provided with a superscript n, e.g. f_y^n denotes the spectral density of (y_t^n). Hence, we obtain a sequence of factor model equations

$$\begin{aligned} y_t^n &= \Lambda^n(z)\xi_t + u_t^n, \\ &= \quad \chi_t^n \quad + u_t^n,\ t \in \mathbb{Z},\ n = n_0, n_{0+1}, \ldots, \end{aligned} \tag{2.1.1}$$

where the q-dimensional factor ξ_t is independent of n, where the latent variable χ_t^n, the noise vector u_t^n and the transfer functions $\Lambda^n(z)$ are *nested*, in the sense that e.g. for $m \leq n$ the coefficient matrices Λ_j^m are the $(m \times q)$ top submatrices of Λ_j^n for all $j \in \mathbb{Z}$ and where n_0 is the smallest integer such that the Assumptions 1.1 d)-e) are satisfied.

As long as not stated otherwise, we will assume that the factor process (ξ_t) is orthogonal white noise with spectral density matrix $f_\xi = I_q$. This is no further restriction on the spectral densities f_χ^n of the latent variables, since our general Assumptions (1.1) always allow of this transformation. Given f_χ^n the factors ξ_t are then identifiable up to unitary transformations, i.e. transformations of the form $U(z)\xi_t$, where $U(e^{-i\lambda})U(e^{-i\lambda})^* = I_q$ holds.

Remark 2.1.1. Notice, that there is no symmetry in the two indices i, used for the cross-sectional dimension, and t, used for the time dimension: while the time dimension is strictly ordered, such that terms such as past, present and future are meaningful, the cross-sectional dimension is not and may be permuted without any consequences.

Instead of requiring f_u^n to be diagonal, we assume that its largest eigenvalue, $\omega_{u,1}^n : [-\pi, \pi] \to \mathbb{R}$ say, remains bounded as n increases, whereas the first q eigenvalues of f_χ^n, $\omega_{\chi,j}^n : [-\pi, \pi] \to \mathbb{R}, j = 1, \ldots, q$ say, increase without bound as n increases. Let $\mathcal{L}$ denote the Lebesgue measure on $\mathbb{R}$ and recall that a real function $\omega : [-\pi, \pi] \to \mathbb{R}$ is *essentially bounded* if there exists an $\bar{\omega} \in \mathbb{R}$ and a subset $M \subset [-\pi, \pi]$ such that $\mathcal{L}(M) = 0$ and $\omega(\lambda) \leq \bar{\omega}$ for any λ in $[-\pi, \pi] \setminus M$.

Assumption 2.1 (Generalized dynamic factor model)**.**

a) $\sup_n \omega_{u,1}^n$ *is essentially bounded.*

b) $\sup_n \omega_{\chi,j}^n(\lambda) = \infty$*, a.e. in* $[-\pi, \pi]$ *for* $j = 1, \ldots, q$*.*

The sequence of nested equations (2.1.1) together with the set of Assumptions 1.1 and 2.1 define the generalized dynamic q-factor model (q-GDFM, henceforth). In other words if a double

sequence $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ can be written as (2.1.1), where Assumption 1.1 holds for every (y_t^n), $n \geq n_0$, and where in addition the spectral densities of the latent variables and of the noise respectively satisfy Assumption 2.1, it is said to allow for a q-GDFM representation.

Informally speaking, a sequence of time series $(u_t^n), n \in \mathbb{N}$, will be said to be weakly correlated if by averaging an increasing number of time series (over cross-section and time) its variance can be caused to vanish. Then, as will be shown below, Assumption 2.1a) is a necessary and sufficient condition for weak correlation.

Example 2.1.1. As a simple example, consider a double sequence (u_{it}) of mutually uncorrelated independent and identically distributed (i.i.d.) variables with $\mathrm{var}(u_{it}) = \sigma^2$, for which Assumption 2.1a) is clearly satisfied (since all eigenvalues of f_u^n are equal, constant and independent of n). Taking a sequence of averages, e.g. the arithmetic means $\bar{u}_t^n = \frac{1}{n}\sum_{i=1}^n u_{it}$, the variances $\mathrm{var}(\bar{u}_t^n) = \frac{1}{n}\sigma^2$ tend to 0 as n increases. ■

On the other hand, Assumption 2.1b) ensures that every factor has a minimum amount of influence on infinitely many observations, which is needed to provide identifiability. Notice that we neither require uniform boundedness nor "uniform divergence". As we will see, essential boundedness of $\omega_{u,1}^n$ is sufficient for weak correlation. And cases where $\omega_{\chi,j}^n$, $j = 1, \ldots, q$ does not diverge for every $\lambda \in [-\pi, \pi]$ may arise in very common situations.

Example 2.1.2. Consider the 1-factor model given in Forni and Lippi (2001), i.e.

$$y_{it} = (1 - z)\xi_t + u_{it}, \tag{2.1.2}$$

where

$$f_\chi^n(\lambda) = \begin{pmatrix} 1 - e^{-i\lambda} \\ 1 - e^{-i\lambda} \\ \vdots \end{pmatrix} (1 - e^{i\lambda}, 1 - e^{i\lambda}, \ldots) = (2 - 2\cos(\lambda))1_n,$$

with $1_n = (1, 1, \ldots 1,)'(1, 1, \ldots 1)$, vanishes for $\lambda = 0$ and $\omega_{\chi,1}^n = (2 - 2\cos(\lambda))n$ diverges for all $\lambda \in [-\pi, \pi] \setminus \{0\}$. Furthermore, since under Assumption 1.1 f_χ^n is continuous, divergence of $\omega_{\chi,1}^n$ is arbitrarily slow in a neighbourhood of 0. ■

Summarizing, Assumption 2.1 captures a basic idea of the generalized factor model: by adding time series one may increase the amount of information by averaging out the noise term and concentrating the latent variables.

In the sequel we are going to specify the terms averaging and weak correlation used above. Therefore let us first recall the following definitions. We denote by $L_2^n([-\pi, \pi], \mathbb{C}, f_y^n)$ the *frequency domain* of the stationary process (y_t^n), i.e. the complex linear space of n-dimensional row vectors $k^n = (k_1, \ldots, k_n)$, such that

(i) k_i is a measurable complex function on $[-\pi, \pi]$ and

(ii) $\int_{-\pi}^{\pi} k^n(e^{-i\lambda}) f_y^n(\lambda) k^n(e^{-i\lambda})^* d\lambda < \infty$.

Then two elements k^n and l^n in $L_2^n([-\pi,\pi],\mathbb{C},f_y^n)$ are considered equal if

$$\int_{-\pi}^{\pi} (k^n(e^{-i\lambda}) - l^n(e^{-i\lambda})) f_y^n(\lambda) (k^n(e^{-i\lambda}) - l^n(e^{-i\lambda}))^* d\lambda = 0.$$

Hence strictly speaking the elements of $L_2^n([-\pi,\pi],\mathbb{C},f_y^n)$ are equivalence classes of functions, not functions. Defining the inner product as $\langle k^n, l^n \rangle_{f_y^n} = \int_{-\pi}^{\pi} k^n(e^{-i\lambda}) f_y^n(\lambda) l^n(e^{-i\lambda})^* d\lambda$ and the norm as $\|k^n\|_{f_y^n} = \sqrt{\langle k^n, k^n \rangle_{f_y^n}}$, then $L_2^n([-\pi,\pi],\mathbb{C},f_y^n)$ is a Hilbert space. In the sequel the notation $L_2^n([-\pi,\pi],\mathbb{C})$ will be used for $L_2^n([-\pi,\pi],\mathbb{C},I_n)$ and $\|k^n\| = \|k^n\|_{I_n}$.

Definition 2.1.1. Let $k^n \in L_2^n([-\pi,\pi],\mathbb{C}) \cap L_2^n([-\pi,\pi],\mathbb{C},f_y^n)$, $n \in \mathbb{N}$. The sequence $(k^n | n \in \mathbb{N})$ is a *dynamic averaging sequence* (a DAS henceforth), if $\lim_n \|k^n\| = 0$.

Example 2.1.3. Clearly,

$$k^n = \frac{1}{n}(\underbrace{1,1,\cdots,1}_{n \text{ times}}),$$

corresponding to a sequence of arithmetic means, is a DAS (albeit static). ■

Definition 2.1.2. The double sequence $(u_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$ is *weakly correlated* if $k^n(z)u_t^n \to 0$ in mean square as $n \to \infty$ for any DAS $(k^n | n \in \mathbb{N})$.

The next theorem (see Forni and Lippi (2001), Theorem 1) shows, that in general averaging-out the noise term is possible, if and only if Assumption 2.1 a) is satisfied.

Theorem 2.1.1. *Let the double sequence $(u_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$ be such that for every $n \in \mathbb{N}$ the process (u_t^n), $t \in \mathbb{Z}$, is stationary with $\mathbb{E}\, u_t^n = 0$ and absolutely summable autocovariances (Assumption 1.1a)). Then the following statements are equivalent:*

1. *$(u_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$ is weakly correlated.*
2. *$\sup_n \omega_{u,1}^n$ is essentially bounded.*

Proof. Define $\omega_{u,1}(\lambda) = \sup_n \omega_{u,1}^n(\lambda)$. Recall the definition of the essential supremum: $\operatorname{ess\,sup} \omega_{u,1} = \inf\{\bar{\omega} : \mathcal{L}(\lambda : \omega_{u,1}(\lambda) > \bar{\omega}) = 0\}$, if $\omega_{u,1}$ is essentially bounded $\operatorname{ess\,sup} \omega_{u,1} < \infty$.

First, we will show that if $\omega_{u,1}$ is not essentially bounded, $(u_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$ will not be weakly correlated.

In general, let $\alpha < \operatorname{ess\,sup}\, \omega_{u,1}$. Then there exist an $s \in \mathbb{N}$ and a $k^s \in L_2^s([-\pi,\pi],\mathbb{C}) \cap L_2^s([-\pi,\pi],\mathbb{C},f_y^s)$ such that $\|k^s\| = 1$, $\|k^s(z)u_t^s\|^2 \geq \alpha$ and thus $\mathcal{L}(\{\lambda : \omega_{u,1}^s(\lambda) \geq \alpha\}) > 0$.

Now suppose that $\omega_{u,1}$ is not essentially bounded. Then there exists a sequence α_s with $\alpha_s \to \infty$ and a corresponding sequence $k^{m_s} \in L_2^{m_s}([-\pi,\pi],\mathbb{C}) \cap L_2^{m_s}([-\pi,\pi],\mathbb{C},f_y^{m_s})$ such

that $\|k^{m_s}\| = 1$, $\|k^{m_s}(z)u_t^{m_s}\|^2 \geq \alpha_s$ for every s. Therefore $l^{m_s} = k^{m_s}/\|k^{m_s}(z)u_t^{m_s}\|$ is a DAS and $\|l^{m_s}(z)u_t^{m_s}\|^2 = 1$. Thus, (1) implies (2).

For the opposite direction we have

$$\begin{aligned}\mathrm{var}(k^n(z)u_t^n) &= \int_\pi^{-\pi} k^n(e^{-i\lambda})f_u^n(\lambda)k^n(e^{-i\lambda})^* d\lambda \leq \int_\pi^{-\pi} \omega_{u,1}^n(\lambda)k^n(e^{-i\lambda})k^n(e^{-i\lambda})^* d\lambda \\ &\leq \operatorname{ess\,sup}\omega_{u,1} \int_\pi^{-\pi} k^n(e^{-i\lambda})(\lambda)k^n(e^{-i\lambda})^* d\lambda,\end{aligned}$$

which converges to 0 as $n \to \infty$ if k^n is a DAS. ■

Chapter 3

A structure theory for GDFMs

Structure theory in general deals with the analysis of relations between observed processes and internal model parameters (see Deistler (2001b)). In Section 3.1 we will be concerned with the relation between the spectral density of the observations and the GDFM representation (2.1.1). Second, in Section 3.3 we will be concerned with the relation between the spectral density of the latent variables and realizations of the latent variables in terms of autoregressive moving average (ARMA) or state space models. Structure theory thus may be seen as a somewhat idealized problem, since the starting point of the analysis consists of the population second moments rather than of data. However, structure theory provides important insight into the properties of the model class under consideration and is thus an indispensable step towards model selection and estimation from data.

Additionally, in Section 3.2, the relations between the factor model with idiosyncratic noise, the principal component model and the GDFM will be analyzed.

3.1 A characterization of the GDFM in terms of the spectral density of the observations

In the sequel we are interested in the question, under which circumstances a sequence of observations (y_t^n) allows for a q-GDFM representation. We are hence looking for conditions on the spectral densities f_y^n of the observations, that imply an underlying q-GDFM.

As we will see, the relevant condition is, that exactly q of the eigenvalues of f_y^n are unbounded as n increases. Here we will directly consider the more general dynamic case and we will largely quote the proofs of Forni and Lippi (2001). For the special case that f_y^n is constant with respect to λ for all $n \in \mathbb{N}$ (static case), analogous statements in terms of the covariance matrices $\Gamma_y^n(0)$ have been first made by Chamberlain (1983) and Chamberlain and Rothschild (1983). A by-product of this analysis is the remarkable result, that asymptotically (as n tends to infinity),

in a certain sense, the dynamic PC model and the GDFM coincide.

The following theorem (see Forni and Lippi (2001), Theorem 2) represents the main result of this section.

Theorem 3.1.1. *Let the double sequence $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ be such that for every $n \in \mathbb{N}$ the process (y_t^n), is stationary with $\mathbb{E}\, y_t^n = 0$, spectral density f_y^n and absolutely summable autocovariances $\Gamma_y^n(s)$, and let $\omega_j^n : [-\pi, \pi] \to \mathbb{R}$, $j = 1, \ldots, n$, denote the eigenvalues of f_y^n in descending order of magnitude. Then $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ allows for a q-GDFM representation if and only if*

1. $\sup_n \omega_{q+1}^n$ *is essentially bounded.*
2. $\sup_n \omega_q^n(\lambda) = \infty$, *a.e. in* $[-\pi, \pi]$.

First it will be shown that the conditions stated in Theorem 3.1.1 are necessary for the existence of a q-GDFM, hence it will be shown that if $(y_t^n), n \in \mathbb{N}$, has a q-GDFM representation then the conditions 1. and 2. are satisfied.

Proof. (if-part)
Since (y_t^n) has a q-GDFM representation, f_y^n can be decomposed as

$$f_y^n(\lambda) = f_\chi^n(\lambda) + f_u^n(\lambda),$$

where all matrices are Hermitian, non-negative definite. Applying corollary A.1.1 we have for $j = 1, \ldots, n$:

$$\begin{aligned} &\text{a)} \quad \omega_j^n(\lambda) \geq \omega_{\chi,j}^n(\lambda), \\ &\text{b)} \quad \omega_j^n(\lambda) \leq \omega_{\chi,j}^n(\lambda) + \omega_{u,1}^n(\lambda). \end{aligned} \tag{3.1.1}$$

Hence Assumption 2.1b) together with a) implies statement 2. Assumption 2.1a) together with b) and the fact that $\operatorname{rk} f_\chi^n \leq q$ for all $n \in \mathbb{N}$ sufficiently large and for all $\lambda \in [-\pi, \pi]$ and therefore $\omega_{\chi,q+1}^n(\lambda) = 0$ for all $\lambda \in [-\pi, \pi]$ imply statement 1. ■

The proof that the conditions of Theorem 3.1.1 are also sufficient, is much more complicated and will be performed in several steps. Let us first introduce some new definitions. Let $\mathbb{H}_y = \overline{\operatorname{span}}(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$, i.e. the Hilbert space spanned by the scalar random variables $y_{it}, i \in \mathbb{N}, t \in \mathbb{Z}$.

Definition 3.1.1. An element $z_t \in \mathbb{H}_y$ is called an *aggregate*, if there exists a DAS $(k^n), n \in \mathbb{N}$, such that $z_t = \lim_n k^n(z) y_t^n$. The set of all aggregates will be denoted by $\mathscr{A}(y)$.

Lemma 3.1.1. *$\mathscr{A}(y)$ is a closed subspace of $\mathbb{H}_y$.*

For a proof, see Forni and Lippi (2001).

Remark 3.1.1. Obviously, if $(y_t^n), n \in \mathbb{N}$, is weakly correlated, then $\mathscr{A}(y) = \{0\}$. Furthermore if y_t^n has a q-GDFM representation, for any aggregate z_t we have

$$z_t = \lim_n k^n(z)\chi_t^n + \lim_n k^n(z)u_t^n,$$

and since the second term on the right hand side vanishes as n tends to infinity, we obtain that $\mathscr{A}(y) \subseteq \mathbb{H}_\chi$ (where $\mathbb{H}_\chi = \overline{\text{span}}(\chi_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$).

Outline of the proof

The proof will be organized as follows: first it will be shown, that if the conditions of Theorem 3.1.1 are satisfied $\mathscr{A}(y)$ contains a q-dimensional orthogonal white noise, (z_t) say, with spectral density equal to I_q, and that $\mathscr{A}(y)$ is actually equal to the space spanned by the scalar components of this orthogonal white noise, i.e.

$$\mathscr{A}(y) = \overline{\text{span}}(z_{jt}|j = 1, \dots, q,\ t \in \mathbb{Z}).$$

Then, consider the projection equations

$$\begin{aligned} y_t^n &= \text{proj}(y_t^n|\mathscr{A}(y)) + e_t^n, \\ &= \underbrace{C^n(z)z_t}_{\gamma_t^n} + e_t^n,\ n \in \mathbb{N},\ t \in \mathbb{Z}, \end{aligned} \tag{3.1.2}$$

where obviously the matrices $C^n(z)$ and the perpendiculars e_t^n are nested, with corresponding (nested) spectral densities

$$f_y^n(\lambda) = C^n(e^{-i\lambda})C^n(e^{-i\lambda})^* + f_e^n = f_\gamma^n + f_e^n,\ n \in \mathbb{N}. \tag{3.1.3}$$

Let $\omega_{\gamma,j}^n(\lambda)$ and $\omega_{e,j}^n(\lambda)$, $j = 1, \dots, n$ denote the eigenvalues of f_γ^n and f_e^n respectively, in descending order of magnitude. Applying corollary A.1.1 we have for $\lambda \in [-\pi, \pi]$

$$\omega_{\gamma,q}^n(\lambda) \geq \omega_q^n(\lambda) - \omega_{e,1}^n(\lambda).$$

Hence, in order to prove that (3.1.2) is a q-GDFM representation and thus to complete the proof, it then remains to show that the perpendiculars $(e_t^n), n \in \mathbb{N}$, are weakly correlated.

To construct the q-dimensional orthogonal white noise (z_t) belonging to $\mathscr{A}(y)$, we will consider the (dynamic) principal components of (y_t^n) and rescale them to have unit spectrum, hence for $n \geq n_0$ we will consider the q-dimensional process

$$\psi_t^n = \Omega_1^n(z)^{-1/2}O_1^n(z)^* y_t^n,\ t \in \mathbb{Z}, \tag{3.1.4}$$

where the matrices $\Omega_1^n(z)$ and $O_1^n(z)$ have been defined in (1.2.10); and call it the *n-th order principal components* of $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$.

To ensure meaningfulness of (3.1.4) and to avoid unnecessary complications, we impose the additional assumption, that the first q eigenvalues of f_y^n are strictly positive,

$$\omega_j^n(\lambda) > 0, \text{ for } n \in \mathbb{N} \text{ sufficiently large, } j = 1, \dots, q \text{ and } \lambda \in [-\pi, \pi].$$

Notice, that this new assumption does not mean any loss of generality, since under the assumptions imposed so far, there always exists a double sequence $(\phi_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$ such that ϕ_{it} is uncorrelated with all leads and lags of ξ_t and u_t and for all $n \in \mathbb{N}$ (ϕ_t^n) is white noise with spectral density equal to I_n. Defining $\breve{u}_t^n = u_t^n + \phi_t^n$ and $\breve{y}_t^n = \chi_t^n + \breve{u}_t^n$, for $n \in \mathbb{N}$ and $t \in \mathbb{Z}$, $\breve{y}_t^n$ still satisfies the assumptions of the theorem and since $f_{\breve{y}}^n = f_y^n + I_n$, it follows that all eigenvalues of f_y^n are greater or equal to 1 and that conditions (1) and (2) hold for (y_{it}) if and only if they hold for $\breve{y}_{it}$.

Further recall, that by Lemmas 1.2.1 - 1.2.3, if (y_t^n) is stationary with absolutely summable autocovariances for all $n \in \mathbb{N}$, then (ψ_t^n) from (3.1.4) is well defined for all $n \geq n_0$.

Since divergence of the first q eigenvalues of f_y^n is not required for all λ in $[-\pi, \pi]$ (recall e.g. model (2.1.2)), the proof must be done piecewise on $[-\pi, \pi]$, for which we will use the following set of functions:

Definition 3.1.2. Let $M \subseteq [-\pi, \pi]$. Then $\mathcal{K}_M$ denotes the subset of $L_2^{q \times q}([-\pi, \pi], \mathbb{C})$ whose elements C are such that $C(\lambda) = 0$ for $\lambda \notin M$ and $C(\lambda)^* C(\lambda) = I_q$ for $\lambda \in M$.

We can now begin to prove that the conditions of Theorem 3.1.1 are sufficient for the existence of a q-GDFM. First, as mentioned above, we will show that $\mathscr{A}(y)$ contains a q-dimensional orthogonal white noise with spectral density equal to I_q. Therefore we start with the (rescaled) m-th order principal components (ψ_t^m) and project them onto the space spanned by the n-th order principal components (ψ_t^n), $n > m$. We will show that, when n and m increase, the residuals of this projection vanish and we will end up with a sequence converging to an orthogonal white noise that belongs to $\mathscr{A}(y)$. However, since these considerations have to be done piecewise over $[-\pi, \pi]$, instead of ψ_t^n, we will consider transformations of the form $C(z)\psi_t^m$.

In order to avoid heavy notation, in the sequel we will use matrix products like $O_1^{m*} O_1^n$, where O_1^{m*} has m columns and O_1^n has $n > m$ rows, assuming that the missing columns of O_1^{m*} have been filled with zeros.

Lemma 3.1.2. *Suppose that the conditions of Theorem 3.1.1 hold. Let $M \subseteq [-\pi, \pi]$, $C \in \mathcal{K}_M$ and $n > m$. Consider the projection equation*

$$C(z)\psi_t^m = \text{proj}(C(z)\psi_t^m | \psi_\tau^n, \tau \in \mathbb{Z}) + \delta_t^{mn}.$$

Then the largest eigenvalue of the spectral density matrix of (δ_t^{mn}), $\mu(\lambda)$ say, is bounded from above,

$$\mu(\lambda) \leq \frac{\omega_{q+1}^n(\lambda)}{\omega_q^m(\lambda)}.$$

Proof. Consider the decomposition of y_t^n corresponding to a dynamic PCA,

$$\begin{aligned} y_t^n &= O_1^n(z)O_1^n(z)^* y_t^n + O_2^n(z)O_2^n(z)^* y_t^n. \\ &= O_1^n(z)\Omega_1^n(z)^{1/2}\psi_t^n + O_2^n(z)O_2^n(z)^* y_t^n. \end{aligned} \tag{3.1.5}$$

The terms on the right hand side are orthogonal, since $O_1^{n*}O_2^n = 0$. Since $\Omega_1^m(z)^{-1/2}O_1^m(z)^* y_t^n = \psi_t^m$, pre multiplying both sides by $C(z)\Omega_1^m(z)^{-1/2}O_1^m(z)^*$ yields the desired projection equation:

$$C(z)\psi_t^m = \underbrace{C(z)\Omega_1^m(z)^{-1/2}O_1^m(z)^*O_1^n(z)\Omega_1^n(z)^{1/2}}_{D(z)}\psi_t^n + \underbrace{C(z)\Omega_1^m(z)^{-1/2}O_1^m(z)^*O_2^n(z)O_2^n(z)^*}_{R(z)}y_t^n. \tag{3.1.6}$$

For the sequel we need two preliminary considerations, first from $O_1^nO_1^{n*} + O_2^nO_2^{n*} = I_n$, we get that $I_n - O_2^nO_2^{n*} \geq 0$ (where ≥ 0 is short for non-negative definite), hence

$$\omega_{q+1}^n I_n - \omega_{q+1}^n O_2^n O_2^{n*} \geq 0, \tag{3.1.7}$$

and second, from the definition of Ω_2^n we have

$$\omega_{q+1}^n O_2^n O_2^{n*} - O_2^n\Omega_2^n O_2^{n*} \geq 0. \tag{3.1.8}$$

Summing up (3.1.7) and (3.1.8) yields

$$\omega_{q+1}^n I_n - O_2^n\Omega_2^n O_2^{n*} \geq 0. \tag{3.1.9}$$

Pre multiplying (3.1.9) by $C\Omega_1^{m^{-1/2}}O_1^{m*}$ and post multiplying (3.1.9) by $O_1^m\Omega_1^{m^{-1/2}}C^*$ and observing that $C\Omega_1^{m^{-1/2}}O_1^{m*}O_2^n\Omega_2^nO_2^{n*}O_1^m\Omega_1^{m^{-1/2}}C^*$ is equal to $Rf_y^nR^*$ yields

$$\omega_{q+1}^n C\Omega_1^{m^{-1}}C^* - Rf_y^nR^* \geq 0 \tag{3.1.10}$$

Applying corollary A.1.1 and observing that the largest eigenvalue of $C\Omega_1^{m^{-1}}C^*$ is bounded from above by $\frac{1}{\omega_q^m}$ and that $Rf_y^nR^*$ is the spectral density of (δ_t^{mn}), we obtain

$$\frac{\omega_{q+1}^n(\lambda)}{\omega_q^m(\lambda)} \geq \mu(\lambda). \tag{3.1.11}$$

■

The next step will be the construction of a convergent sequence. Therefore we will need the following results providing that the spectral densities of a convergent sequence of processes will converge too. Lemma 3.1.3 recalls the well-known fact that L_1- or L_2- convergence of a sequence of functions is equivalent to almost sure convergence of a subsequence. And Lemma 3.1.4 is an application of Lemma 3.1.3.

Lemma 3.1.3. *Let $(g^n), n \in \mathbb{N}$, be a convergent sequence of functions $g^n \in L_k([-\pi,\pi],\mathbb{C})$, $k \in \{1,2\}$, i.e. (g^n) converges in the norm of $L_k([-\pi,\pi],\mathbb{C}), k \in \{1,2\}$. Then there exists a subsequence (g^{n_i}) such that $\lim_i g^{n_i}(\lambda) = g(\lambda)$ a.e. in $[-\pi,\pi]$.*

Proof. For a proof see e.g. Apostol (1974). ■

In the following lemma $S(x,z)$ will denote the cross spectrum of two processes (x_t) and (z_t), $t \in \mathbb{Z}$, belonging to $\mathbb{H}_y$.

Lemma 3.1.4. *Let (x_t^n), (z_t^n), $t \in \mathbb{Z}, n \in \mathbb{N}$, be two sequences of scalar processes belonging to $\mathbb{H}_y$ and costationary with (y_t^n) for any $n \in \mathbb{N}$, with $x_t = \lim_n x_t^n$ and $z_t = \lim_n z_t^n$. Then there exist subsequences $(x_t^{n_i})$, $(z_t^{n_i})$, such that a.e. in $[-\pi, \pi]$*

$$S(x,z,\lambda) = \lim_i S(x_t^{n_i}, z_t^{n_i}, \lambda).$$

Proof. Recall that the cross spectrum is related to the inner product in $\mathbb{H}_y$ through $\langle x_t^n, z_t^n \rangle = \int_{-\pi}^{\pi} S(x_t^n, z_t^n, \lambda) d\lambda$. Since the inner product is continuous, convergence of (x_t^n) and (z_t^n) implies that $\langle x_t^n, z_t^n \rangle - \langle x_t, z_t \rangle \to 0$, and hence $\int_{-\pi}^{\pi} |S(x_t^n, z_t^n, \lambda) - S(x_t, z_t, \lambda)| d\lambda \to 0$, i.e. $S(x_t^n, z_t^n, \lambda)$ is L_1-convergent. The result follows from Lemma 3.1.3. ■

Before proceeding, let us consider two further points.

First, recall, that under the conditions of Theorem 3.1.1 there exist a subset Π of $[-\pi, \pi]$ and a positive real $\bar{\omega}$, such that $\mathcal{L}([-\pi, \pi] \setminus \Pi) = 0$ and

(i) $\omega_{q+1}^n(\lambda) \leq \bar{\omega}$ for any $n \in \mathbb{N}$ and any $\lambda \in \Pi$ and

(ii) $\sup_n \omega_j^n(\lambda) = \infty$ for any $\lambda \in \Pi$.

Hence in the sequel we will restrict ourselves to the set Π, since if any statement holds a.e. in Π, it holds a.e. in $[-\pi, \pi]$. However, we still cannot assume that $\omega_q^n(\lambda) \geq \alpha_n$ for some non-decreasing and diverging sequence (α_n) for all λ in Π, since, as mentioned above, divergence may be arbitrarily slow in the neighbourhood of points of non-divergence, so that we still have to work on subsets of Π.

Then suppose $M \subseteq \Pi$ with $\mathcal{L}(M) > 0$ and let $(\alpha_n), n \in \mathbb{N}$, $\alpha_n \in \mathbb{R}_+$, denote a non-decreasing, diverging sequence, such that for any $\lambda \in M$, $\omega_q^n(\lambda) \geq \alpha_n$ holds. Notice, that the proof that such subsets M exist will be given below in Lemma 3.1.6 by specifying a construction. However, applying Lemma 3.1.2 to such a subset M, we have for any $\lambda \in M$,

$$\mu(\lambda) \leq \frac{\bar{\omega}}{\alpha_m}.$$

Second, normalized eigenvectors are not unique − eigenvectors belonging to eigenvalues of the same size may still be rotated and even if all eigenvalues are distinct they may still be multiplied by $e^{i\lambda}$ − and hence the rescaled principal components are not unique either. Consequently in order to construct a converging sequence we have to choose a certain transformation. For any $\lambda \in M$, where M is defined as above, the spectral densities of the projection equation,

$$C(z)\psi_t^m = D(z)\psi_t^n + R(z)y_t^n, \tag{3.1.12}$$

equal

$$I_q = D(e^{-i\lambda})D(e^{-i\lambda})^* + R(e^{-i\lambda})f_y^n(\lambda)R(e^{-i\lambda})^*. \tag{3.1.13}$$

Consider the eigenvalue decomposition

$$D(e^{-i\lambda})D(e^{-i\lambda})^* = H(e^{-i\lambda})\Delta(e^{-i\lambda})H(e^{-i\lambda})^*,$$

calling δ_q the smallest eigenvalue of DD^*, then from (3.1.13) and corollary A.1.1, we see that

$$1 \geq \delta_q \geq 1 - \mu \geq 1 - \frac{\bar{\omega}}{\alpha_m}. \tag{3.1.14}$$

Hence, taking m^* large enough, such that $\frac{\bar{\omega}}{\alpha_m} < 1$ for $m \geq m^*$, then $\delta_q(\lambda) > 0$ for $\lambda \in M$ and $m \geq m^*$, and the following definition is meaningful:

$$F(e^{-i\lambda}) = \begin{cases} H(e^{-i\lambda})\Delta(e^{-i\lambda})^{-\frac{1}{2}}H(e^{-i\lambda})^*D(e^{-i\lambda}) & \lambda \in M \\ 0 & \lambda \notin M \end{cases} \tag{3.1.15}$$

Note that, since on M,

$$FF^* = H\Delta^{-\frac{1}{2}}\underbrace{H^*DD^*H}_{\Delta}\Delta^{-\frac{1}{2}}H^* = I_q,$$

$F \in \mathcal{K}_M$.

The next lemma shows how to construct the converging sequence on such subsets M of Π.

Lemma 3.1.5. *Suppose that the conditions of Theorem 3.1.1 hold. Let $\Pi \subseteq [-\pi, \pi]$ be defined as above and $M \subseteq \Pi$ with $\mathcal{L}(M) > 0$ and let $(\alpha_n), n \in \mathbb{N}$, $\alpha_n \in \mathbb{R}_+$, denote a non-decreasing, diverging sequence, such that for any $\lambda \in M$, $\omega_q^n(\lambda) \geq \alpha_n$ holds. Then there exists a q-dimensional process (v_t), $t \in \mathbb{Z}$, such that $v_{jt} \in \mathscr{A}(y)$ for any $j \in 1, \ldots, q$ and $t \in \mathbb{Z}$ and the spectral density of (v_t) equals I_q for λ a.e. in M and 0 elsewhere.*

Proof. We will need the following preliminary result: let $C \in \mathcal{K}_M$ and let $F \in \mathcal{K}_M$ be defined as in (3.1.15). For τ: $0 \leq \tau \leq 2$ given, there exists an $m_\tau \in \mathbb{N}$, such that

$$\frac{\bar{\omega}}{\alpha_{m_\tau}} < \frac{\tau}{2} \tag{3.1.16}$$

and the largest eigenvalue of the spectral density of

$$(C(z)\psi_t^m - F(z)\psi_t^n),\ t \in \mathbb{Z} \tag{3.1.17}$$

is less or equal to τ for any $\lambda \in \Pi$ and $n > m \geq m_\tau$.
Let $n > m \geq m_\tau$. From the projection equation (3.1.6) we see that

$$C(z)\psi_t^m - F(z)\psi_t^n = R(z)y_t^n + (D(z) - F(z))\,\psi_t^n. \tag{3.1.18}$$

Hence for $\lambda \in M$ the spectral density of (3.1.17), S say, is equal to

$$\begin{aligned} S &= Rf_y^nR^* + DD^* - FD^* - DF^* + FF^* \\ &= I_q - DD^* + DD^* - FD^* - DF^* + I_q \\ &= 2I_q - FD^* - DF^*, \end{aligned}$$

where we used equation (3.1.13) and the fact that on M, $FF^* = I_q$. Observing, that $FD^* = H\Delta^{-\frac{1}{2}}H^*DD^* = H\Delta^{\frac{1}{2}}H^* = DF^*$, yields (for $\lambda \in M$)

$$S = 2H(I_q - \Delta^{\frac{1}{2}})H^*,$$

Hence for any $\lambda \in \Pi$ the largest eigenvalue of S is less or equal to $2(1-\delta_q^{1/2}) \leq 2(1-\delta_q)$, which by (3.1.14) is less or equal to $2(1-(1-\frac{\bar{\omega}}{\alpha_{m_\tau}})) < \tau$. (And $\tau < 2$ since otherwise $\delta_q \leq 0$ and the definition of F would make no sense.)

We will use this result to construct a sequence $(v_t^n), n \in \mathbb{N}$ that converges to (v_t).

First, set $\tau = \frac{1}{2^2}$, $s_1 = m_\tau$ (such that (3.1.16) is satisfied), $F^1 \in \mathcal{K}_M$ and $v_t^1 = F^1\psi_t^{s_1}$. Obviously, the spectral density matrix of (v_t^1) is I_q for $\lambda \in M$ and 0 for $\lambda \notin M$.

Second, set $\tau = \frac{1}{2^4}$, $s_2 = m_\tau$ (such that (3.1.16) is satisfied), determine D as in (3.1.6) with F^1 instead of C, s_1 instead of m and s_2 instead of n, determine F^2 as defined in (3.1.15) and set $v_t^2 = F^2\psi_t^{s_2}$. Again, the spectral density matrix of (v_t^2) is I_q for $\lambda \in M$ and 0 for $\lambda \notin M$. Moreover by the result derived above, the largest eigenvalue of the spectral density of $(v_t^1 - v_t^2)$ is bounded from above by $\frac{1}{2^2}$ for any $\lambda \in \Pi$. Considering

$$\|v_{jt}^1 - v_{jt}^2\|^2 = \mathrm{var}(v_{jt}^1 - v_{jt}^2) \leq \int_{-\pi}^{\pi} \frac{1}{2^2} d\lambda = \frac{2\pi}{2^2},$$

we see that $\frac{1}{\sqrt{2\pi}}\|v_{jt}^1 - v_{jt}^2\| \leq \frac{1}{2}$, for $j = 1, \ldots, q$.

By recursion, set $\tau = \frac{1}{2^{2^k}}$, $s_k = m_\tau$ (such that (3.1.16) is satisfied) and proceed as described for (v_t^2). The spectral density of (v_t^k) is then I_q for $\lambda \in M$ and 0 for $\lambda \notin M$ and $\frac{1}{\sqrt{2\pi}}\|v_{jt}^{k-1} - v_{jt}^k\| \leq \frac{1}{2^{k-1}}$. And since

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\|v_{jt}^k - v_{jt}^{k+h}\| &\leq \underbrace{\frac{1}{\sqrt{2\pi}}\|v_{jt}^k - v_{jt}^{k+1}\|}_{\leq \frac{1}{2^k}} + \underbrace{\frac{1}{\sqrt{2\pi}}\|v_{jt}^{k+1} - v_{jt}^{k+2}\|}_{\leq \frac{1}{2^{k+1}}} + \ldots + \underbrace{\frac{1}{\sqrt{2\pi}}\|v_{jt}^{k+h-1} - v_{jt}^{k+h}\|}_{\leq \frac{1}{2^{k+h-1}}} \\
&\leq \sum_{j=0}^{h-1} \frac{1}{2^{k+j}} = \frac{1}{2^k}(2 - \frac{1}{2^{h-1}}) < \frac{1}{2^{k-1}} \qquad (3.1.19)
\end{aligned}$$

holds, every component of $(v_t^n), n \in \mathbb{N}$, is a Cauchy-sequence; (v_t) is then the vector of limits.

Since $(v_t^n) = F^n\psi_t^{s_n} = F\Omega_1^{s_n\,-1/2}O_1^{s_n*}y_t^n$, in order to show that $v_{jt} \in \mathscr{A}(y)$ for any $j \in 1, \ldots, q$, we need that every row of $G^n := F^n\Omega_1^{s_n\,-1/2}O_1^{s_n*}$ is a DAS. Since

$$\begin{aligned}
G^nG^{n*} &= F^n\Omega_1^{s_n\,-1/2}O_1^{s_n*}O_1^{s_n}\Omega_1^{s_n\,-1/2}F^{n*} \\
&= F^n\Omega_1^{s_n\,-1}F^{n*} \qquad (3.1.20)
\end{aligned}$$

we have

$$\begin{aligned}\lim_n \int_{-\pi}^{\pi} G^n(\lambda)G^n(\lambda)^* d\lambda &= \lim_n \int_{-\pi}^{\pi} F^n(\lambda)\Omega_1^{s_n}(\lambda)^{-1}F^n(\lambda)^* d\lambda \\ &\leq \lim_n \int_{-\pi}^{\pi} \frac{1}{\omega_q^{s_n}}(\lambda)d\lambda = 0, \end{aligned} \tag{3.1.21}$$

where the last equality follows from the Lebesgue Convergence Theorem, that may be applied since $\frac{1}{\omega_q^{s_n}}$ is by assumption bounded from above and converges to 0.

Finally, Lemma 3.1.4 and the construction of (v_t) imply that the spectral density of (v_t) equals I_q for λ a.e. in M and 0 elsewhere. ■

The next lemma shows how to explicitly construct the subsets M of Π and how to piece them together, such that eventually we will obtain a q-dimensional orthogonal white noise, that belongs to $\mathscr{A}(y)$.

Lemma 3.1.6. *Suppose that the conditions of Theorem 3.1.1 hold. Then there exists a q-dimensional white noise (z_t), $t \in \mathbb{Z}$, with spectral density matrix equal to I_q a.e. in $[-\pi, \pi]$ and such that every component z_{jt}, $j = 1, \ldots, q$, $t \in \mathbb{Z}$, is an aggregate.*

Proof. We start with $S_0 = \Pi$, set $k = 1$, choose m_1 as the smallest $m \in \mathbb{N}$, such that

$$\mathcal{L}(\lambda \in S_0 : \omega_q^m(\lambda) > 1) > \pi$$

and define

$$S_1 = \{\lambda \in S_0 : \omega_q^{m_1}(\lambda) > 1\}.$$

We continue by setting $k = 2$, choose m_2 as the smallest $m \in \mathbb{N}$, such that

$$\mathcal{L}(\lambda \in S_1 : \omega_q^m(\lambda) > 2) > \pi$$

and define

$$S_2 = \{\lambda \in S_1 : \omega_q^{m_2}(\lambda) > 2\}.$$

By recursion, for $k \in \mathbb{N}$, choose m_k as the smallest $m \in \mathbb{N}$, such that

$$\mathcal{L}(\lambda \in S_{k-1} : \omega_q^m(\lambda) > k) > \pi$$

and define

$$S_k = \{\lambda \in S_{k-1} : \omega_q^{m_k}(\lambda) > k\}.$$

Then we set

$$M_1 = S_0 \cap S_1 \cap \ldots \cap S_k \cap \ldots$$

By construction $\mathcal{L}(M_1) \geq \pi$.

We continue with the definition of M_2, where we proceed just as for M_1, but now we start with $S_0 = \Pi \setminus M_1$ in place of Π and take $\mathcal{L}(\Pi \setminus M_1)/2$ in place of π. By recursion we define M_k just as we defined M_1, but starting with $S_0 = \Pi \setminus (M_1 \cup M_2 \cup \dots M_{k-1})$ and with $\mathcal{L}(\Pi \setminus (M_1 \cup M_2 \cup \dots M_{k-1}))/2$ in place of π. Defining $M = \bigcup_k M_k$, we have that $M_k \cap M_j = \emptyset$, if $k \neq j$ and $\mathcal{L}(M) = \sum_k \mathcal{L}(M_k) = 2\pi$.

By construction, for every subset M_k defined as above, there exists a non-decreasing diverging sequence $(\alpha_n^k), n \in \mathbb{N}$, $\alpha_n^k \in \mathbb{N}$, such that for any $\lambda \in M_k$, $\omega_q^n(\lambda) \geq \alpha_n^k$ holds. Therefore by Lemma 3.1.5 for any $k \in \mathbb{N}$ there exists a q-dimensional vector process (v_t^k), $t \in \mathbb{Z}$, such that $v_{jt}^k \in \mathscr{A}(y)$ for any $j \in 1, \dots, q$ and $t \in \mathbb{Z}$ and the spectral density of (v_t^k) equals I_q for λ a.e. in M_k and 0 elsewhere. Defining $z_t = \sum_k v_t^k$, the result follows. ■

Next we will show, that the space spanned by $(z_{jt} | j = 1, \dots, q,\ t \in \mathbb{Z})$ is actually equal to $\mathscr{A}(y)$.

Lemma 3.1.7. *Suppose that the conditions of Theorem 3.1.1 hold and let (z_t), $t \in \mathbb{Z}$, be defined as in lemma 3.1.6. Then the closure of the space spanned by $(z_{jt} | j = 1, \dots, q,\ t \in \mathbb{Z})$ is equal to $\mathscr{A}(y)$.*

Proof. Let $\mathbb{H}_z = \overline{\text{span}}(z_{jt} | j = 1, \dots, q,\ t \in \mathbb{Z})$, x_t in $\mathscr{A}(y)$ and consider the projection

$$x_t = \text{proj}(x_t | \mathbb{H}_z) + r_t.$$

We want to show that the residuals r_t are 0. Therefore we consider the $(q+1)$-dimensional process $(z_t', r_t)'$ with spectral density, $S(\lambda)$ say:

$$S(\lambda) = \begin{pmatrix} I_q & 0 \\ 0 & f_r(\lambda) \end{pmatrix},$$

such that $\det S(\lambda) = f_r(\lambda)$ and show that the latter equals 0.
Since z_{jt} is an aggregate and x_t is an aggregate, r_t is also an aggregate and there exist DAS's $(k_j^n | n \in \mathbb{N})$, $j = 1, \dots, q+1$, such that

$$\begin{aligned} \lim_n k_j^n y_t^n &= z_{jt}, \quad j = 1, \dots, q \\ \lim_n k_{q+1}^n y_t^n &= r_t. \end{aligned}$$

From the definition of a DAS, we know that $\lim_n \int_\pi^\pi |k_j^n(\lambda)|^2 d\lambda = 0$ for $j = 1, \dots, q+1$, hence (k_j^n) converges in L_2-sense for $j = 1, \dots, q+1$. Applying Lemma 3.1.3, there exist subsequences of (k_j^n) that converge to 0 pointwise a.e. in $[-\pi, \pi]$. Moreover, calling $Z^n(\lambda)$ the spectral density of $(k_1^n y_t^n, \dots, k_{q+1}^n y_t^n)$, applying Lemma 3.1.4, there exists a subsequence of $Z^n(\lambda)$ that converges to $S(\lambda)$ a.e. in $[-\pi, \pi]$. Thus, with no loss of generality we can assume that (k_j^n) converges to 0 for $j = 1, \dots, q+1$ and $Z^n(\lambda)$ converges to $S(\lambda)$ a.e. in $[-\pi, \pi]$.

Now, for $j = 1, \dots, q+1$, set $p_j^n = k_j^n O_1^n$ and $q_j^n = k_j^n - p_j^n O_1^{n*}$. Then $k_j^n = p_j^n O_1^{n*} + q_j^n$ and since $p_j^n O_1^{n*}$ and q_j^n are orthogonal

$$|k_j^n(\lambda)|^2 = |p_j^n(\lambda)|^2 + |q_j^n(\lambda)|^2,$$

implying that (p_j^n) and (q_j^n) converge to 0 a.e. in $[-\pi, \pi]$, and

$$k_j^n y_t^n = p_j^n O_1^{n*} y_t^n + q_j^n y_t^n$$

is the orthogonal projection of $k_j^n y_t^n$ onto the n-th order principal component of y_t^n. Therefore $Z^n(\lambda)$ may be written as the sum of the spectral density of the first parts, $Z_1^n(\lambda)$ say, and the spectral density of the second parts, $Z_2^n(\lambda)$ say. Since $\operatorname{rk} Z_1^n(\lambda) = q$, Z_1^n is singular for any λ in $[-\pi, \pi]$. Moreover, since $q_j^n(\lambda)$ is orthogonal to $O_1^n(\lambda)$ we have

$$q_j^n f_y^n q_j^n \leq \omega_{q+1}^n |q_j^n(\lambda)|^2,$$

which together with the essential boundedness of ω_{q+1}^n for any $n \in \mathbb{N}$ and the fact that (q_j^n) converges to 0 a.e. in $[-\pi, \pi]$ implies that $\det Z^n(\lambda)$ converges to 0 a.e. in $[-\pi, \pi]$ and hence $\det S(\lambda) = 0$ a.e. in $[-\pi, \pi]$. ■

Up to now, it has been shown that under the conditions of Theorem 3.1.1, $\mathscr{A}(y)$ is spanned by the scalar components of q-dimensional white noise. Recall the sequence of nested projection equations (3.1.2), i.e. $y_t^n = \operatorname{proj}(y_t^n | \mathscr{A}(y)) + e_t^n = \gamma_t^n + e_t^n$. What remains to be shown is, that the residuals $(e_t^n), n \in \mathbb{N}$, are weakly correlated. Since in the sequel we are only interested in the perpendiculars of (3.1.2), we will disregard the q-dimensional basis $(z_t), t \in \mathbb{Z}$, of $\mathscr{A}(y)$, defined as the limit of linear transformations of principal components, but we will base on the following weaker concept of convergence.

Definition 3.1.3. Let $(v_t^n | n \in \mathbb{N},\ t \in \mathbb{Z})$ be a sequence of q-dimensional processes belonging to $\mathbb{H}_y$, costationary with (y_t^n) for any $n \in \mathbb{N}$ and consider the orthogonal projection

$$\begin{aligned} v_t^m &= \operatorname{proj}(v_t^m | v_\tau^n,\ \tau \in \mathbb{Z}) + \rho_t^{mn} \\ &= A^{mn}(z) v_t^n + \rho_t^{mn}. \end{aligned} \tag{3.1.22}$$

Let f_ρ^{mn} denote the spectral density of the perpendiculars (ρ_t^{mn}). Then $(v_t^n | n \in \mathbb{N},\ t \in \mathbb{Z})$ is said to create a *Cauchy sequence of spaces* if, for any $\varepsilon > 0$ and for λ a.e. in $[-\pi, \pi]$, there exists an $n^*(\varepsilon, \lambda)$, such that for all $n, m \geq n^*(\varepsilon, \lambda)$,

$$\operatorname{tr}(f_\rho^{mn}(\lambda)) < \varepsilon.$$

Remark 3.1.2. Definition 3.1.3 is weaker than convergence in the Hilbert space. To see this, consider a q-dimensional sequence (v_t^n) that converges in $\mathbb{H}_y$ and let $\varepsilon_t^{mn} = v_t^n - v_t^m$: for every component ε_{jt}^n, $j = 1, \ldots, q$, $\|\varepsilon_{jt}^{mn}\|^2 = \operatorname{var}(\varepsilon_{jt}^{mn}) = \int_{-\pi}^{\pi} f_{\varepsilon_j}^{nm}(\lambda) d\lambda \to 0$, which implies, $\operatorname{tr}(f_\varepsilon^{nm}(\lambda)) \to 0$ for $n, m \to \infty$ and λ a.e. in $[-\pi, \pi]$. Since (3.1.22) is a projection, we have $\operatorname{tr}(f_\rho^{nm}(\lambda)) \leq \operatorname{tr}(f_\varepsilon^{nm}(\lambda))$, and hence (v_t^n) creates a Cauchy sequence of spaces. By contrast, a sequence that creates a Cauchy sequence of spaces does not necessarily converge in the Hilbert space. Consider for example the rescaled principal components (ψ_t^n), which, as will be shown in the Lemma below, create a Cauchy sequence of spaces. Suppose for a moment, that (ψ_t^n) converges to some (ψ_t). As we have said above $(-1)^n(\psi_t^n)$ also defines a sequence of rescaled principal components, but this time it does not converge.

The first part of the proof was especially complicated because of the construction of a convergent sequence. Now, since we can base on the weaker concept of Cauchy sequences of spaces and do not need a convergent basis, all we need are the principal components, as the next lemmas will show.

Lemma 3.1.8. *Suppose that the conditions of Theorem 3.1.1 hold. Then the sequence of rescaled principal components* $(\psi_t^n | n \in \mathbb{N}, t \in \mathbb{Z})$ *creates a Cauchy sequence of spaces.*

Proof. For $n > m$, consider the projection

$$\begin{aligned} \psi_t^m &= \text{proj}(\psi_t^m | \psi_\tau^n, \tau \in \mathbb{Z}) + \rho_t^{mn} \\ &= D^{mn}(z)\psi_t^n + \rho_t^{mn}. \end{aligned} \tag{3.1.23}$$

From Lemma 3.1.2 we know that the largest eigenvalue of f_ρ^{mn} is not larger than $\frac{\omega_{q+1}^n}{\omega_q^m}$ and thus converges to 0 a.e. in $[-\pi, \pi]$, which implies that $\text{tr}(f_\rho^{mn}(\lambda))$ converges to 0 a.e. in $[-\pi, \pi]$ (since the trace of a matrix equals the sum of its eigenvalues).

To cover the case $m > n$, we consider the reverse projection, i.e.

$$\psi_t^n = D^{mn}(z)^* \psi_t^m + \rho_t^{nm}. \tag{3.1.24}$$

Taking the spectral densities of (3.1.23) and (3.1.24), we get

$$I_q = DD^* + f_\rho^{mn} = D^*D + f_\rho^{nm} \tag{3.1.25}$$

and since $\text{tr}(DD^*) = \text{tr}(D^*D)$, $\text{tr}(f_\rho^{mn}) = \text{tr}(f_\rho^{nm})$, which completes the proof. ■

The next lemma shows that a sequence of projections onto a sequence that creates a Cauchy sequence of spaces is convergent and hence explains why definition 3.1.3 is sufficient for our purposes.

Lemma 3.1.9. *Let the sequence* $(v_t^n | n \in \mathbb{N},\ t \in \mathbb{Z})$ *fulfill definition 3.1.3 with spectral density equal to* I_q *for any* $n \in \mathbb{N}$ *and let* $(x_t | t \in \mathbb{Z})$ *be a scalar process belonging to* $\mathbb{H}_y$ *and costationary with* (y_t^n) *for any* $n \in \mathbb{N}$. *Further let* x_t^n *denote the orthogonal projection* $x_t^n = \text{proj}(x_t | \overline{\text{span}}(v_{j\tau}^n, j = 1, \ldots, q,\ \tau \in \mathbb{Z}))$. *Then* x_t^n *converges in* $\mathbb{H}_y$.

Proof. Consider the two projections

$$\begin{aligned} x_t &= x_t^n + r_t^n = b^n(z)v_t^n + r_t^n \\ x_t &= x_t^m + r_t^m = b^m(z)v_t^m + r_t^m. \end{aligned} \tag{3.1.26}$$

Subtracting the second line from the first yields

$$x_t^n - x_t^m = b^n(z)v_t^n - b^m(z)v_t^m = r_t^m - r_t^n. \tag{3.1.27}$$

In the sequel, let $S(x,y,\lambda)$ denote the cross spectrum of (x_t) and (y_t) at $\lambda \in [-\pi,\pi]$. Using the middle expression of (3.1.27), the spectral density of $(x_t^n - x_t^m)$, $\mathcal{S}^{nm}$ say, can then be written as

$$\mathcal{S}^{nm}(\lambda) = b^n(e^{-i\lambda})b^n(e^{-i\lambda})^* - 2S(b^n(z)v_t^n, b^m(z)v_t^m, \lambda) + b^m(e^{-i\lambda})b^m(e^{-i\lambda})^*. \tag{3.1.28}$$

Using equation (3.1.27) the middle term on the right hand side, can be written as

$$\begin{aligned} & S((r_t^m - r_t^n) + b^m(z)v_t^m, b^m(z)v_t^m, \lambda) + S(b^n(z)v_t^n, b^n(z)v_t^n - (r_t^m - r_t^n), \lambda) \\ = \;& b^m(e^{-i\lambda})b^m(e^{-i\lambda})^* + \\ & + S((r_t^m - r_t^n), b^m(z)v_t^m, \lambda) - S((r_t^m - r_t^n), b^n(z)v_t^n), \lambda) + \\ & + b^n(e^{-i\lambda})b^n(e^{-i\lambda})^*, \end{aligned} \tag{3.1.29}$$

which yields

$$\mathcal{S}^{nm}(\lambda) = -S((r_t^m - r_t^n), b^m(z)v_t^m, \lambda) + S((r_t^m - r_t^n), b^n(z)v_t^n), \lambda), \tag{3.1.30}$$

which due to orthogonality of r_t^k and $b^k(z)v_t^k$ can be simplified to

$$\mathcal{S}^{nm}(\lambda) = S(r_t^n, b^m(z)v_t^m, \lambda) + S(r_t^m, b^n(z)v_t^n), \lambda). \tag{3.1.31}$$

Consider the projection equation of $b^m(z)v_t^m$ onto (v_t^n):

$$b^m(z)v_t^m = b^m(z)A^{mn}(z)v_t^n + b^m(z)\rho_t^{mn}, \tag{3.1.32}$$

hence $S(r_t^n, b^m(z)v_t^m, \lambda)$ reduces to $S(r_t^n, b^m(z)\rho_t^{mn}, \lambda)$. Next, observing that $b^m b^{m*}$ and f_r are bounded from above by f_x together with the fact that (v_t^n) creates a Cauchy sequence of spaces and hence $\mathrm{tr}(f_\rho^{mn}) \to 0$ for λ a.e. in $[-\pi,\pi]$, implies that $S(r_t^n, b^m(z)\rho_t^{mn}, \lambda)$ converges to 0 for λ a.e. in $[-\pi,\pi]$ as $m,n \to \infty$. Obviously, the same arguments hold true for $S(r_t^m, b^n(z)v_t^n), \lambda)$, such that altogether $\mathcal{S}^{nm}$ converges to 0 for λ a.e. in $[-\pi,\pi]$ as $m,n \to \infty$. Applying the Lebesgue Convergence Theorem (since both spectral densities of (x_t^n) and (x_t^m) are dominated by f_x), we see that $var(x_t^n - x_t^m) = \int_{-\pi}^{\pi} \mathcal{S}^{nm}(\lambda)d\lambda \to 0$ as $m,n \to \infty$, implying that (x_t^n) is a Cauchy sequence and hence convergent in $\mathbb{H}_y$. ■

Let us recall the n-th order dynamic PC model. For $n \in \mathbb{N}$, the projection equation corresponding to the dynamic PC model for y_t^n is of the form

$$\begin{aligned} y_t^n &= O_1^n(z)\Omega_1^n(z)^{1/2}\psi_t^n + O_2^n(z)O_2^n(z)^* y_t^n \\ &= O_1^n(z)O_1^n(z)^* y_t^n + O_2^n(z)O_2^n(z)^* y_t^n. \end{aligned} \tag{3.1.33}$$

Hence let $i \leq n$ and denote by $\underline{o}_{ki}^n(z)$, $k \in \{1,2\}$, the i-th row of $O_k^n(z)$, then we have for y_{it}

$$\begin{aligned} y_{it} &= \underline{o}_{1i}^n(z)O_1^n(z)^* y_t^n + \underline{o}_{2i}^n(z)O_2^n(z)^* y_t^n \\ &= \tilde{\chi}_{it}^n + \tilde{u}_{it}^n, \end{aligned} \tag{3.1.34}$$

where $\tilde{\chi}_{it}^n$ and $\tilde{u}_{it}^n$ denote the i-th component of the n-th order PC latent variable and the n-th order PC noise respectively. In addition we have already introduced the projection of y_t^n onto $\mathscr{A}(y)$ (see also (3.1.2)), then for $i \in \mathbb{N}$,

$$\begin{aligned} y_{it} &= \mathrm{proj}(y_{it}|\mathscr{A}(y)) + e_{it} \\ &= \gamma_{it} + e_{it}, \end{aligned} \tag{3.1.35}$$

where as distinct from the PC model γ_{it} and e_{it} do not depend on n, since $\mathscr{A}(y)$ does not depend on n. The next lemma reveals that as $n \to \infty$, the two representations (3.1.34) and (3.1.35) are equivalent.

Lemma 3.1.10. *Suppose that the conditions of Theorem 3.1.1 hold. Then the sequence of n-th order PC latent variables $\tilde{\chi}_{it}^n$, $n \in \mathbb{N}$, $i \leq n$ converges to $\gamma_{it} = \text{proj}(y_{it}|\mathscr{A}(y))$ in mean square as $n \to \infty$.*

Proof. By Lemma 3.1.8 (ψ_t^n) creates a Cauchy sequence of spaces, hence by Lemma 3.1.9 the projections of y_{it}^n onto (ψ_t^n), i.e. $\tilde{\chi}_{it}^n$, $i \leq n$ converge in $\mathbb{H}_y$ as $n \to \infty$. What remains to be shown is, that the limit belongs to $\mathscr{A}(y)$ and is a projection.

Since $\tilde{\chi}_{it}^n = \underline{o}_{1i}^n(z)O_1^n(z)^* y_t^n$, for the first part we need to show that $(\underline{o}_{1i}^n(\lambda))O_1^n(\lambda)^*)$, $n \in \mathbb{N}$, is a DAS, hence we need to show that

$$\int_{-\pi}^{\pi} \underline{o}_{1i}^n(e^{-i\lambda})O_1^n(e^{-i\lambda})^* O_1^n(e^{-i\lambda})\underline{o}_{1i}^n(e^{-i\lambda})^* d\lambda = \int_{-\pi}^{\pi} \underline{o}_{1i}^n(e^{-i\lambda})\underline{o}_{1i}^n(e^{-i\lambda})^* d\lambda \tag{3.1.36}$$

converges to 0. Therefore, consider the spectral density of $\tilde{\chi}_{it}^n$, i.e. $\underline{o}_{1i}^n(e^{-i\lambda})\Omega_1^n(\lambda)\underline{o}_{1i}^n(e^{-i\lambda})^*$, for which the following inequalities hold:

$$\omega_q^n \underline{o}_{1i}^n \underline{o}_{1i}^{n\,*} \leq \underline{o}_{1i}^n \Omega_1^n \underline{o}_{1i}^{n\,*} \leq f_{y_i}, \tag{3.1.37}$$

implying

$$\underline{o}_{1i}^n \underline{o}_{1i}^{n\,*} \leq \frac{f_{y_i}}{\omega_q^n}, \tag{3.1.38}$$

where the latter converges to 0 a.e. in $[-\pi, \pi]$ and is bounded from above by assumption, so that by application of the Lebesgue Convergence Theorem the integral on the right hand side of (3.1.36) converges to 0 too.

For the second part, observe that by construction $\tilde{u}_{it}^n$ is orthogonal to ψ_{t-k}^n for any $n \in \mathbb{N}$, $k \in \mathbb{Z}$. Recall that $\mathscr{A}(y) = \overline{\text{span}}(z_{jt}|j = 1, \ldots, q, t \in \mathbb{Z})$ and that z_t has been defined as the limit of linear transformations of ψ_t^n. Continuity of the inner product implies that $\tilde{u}_{it} = \lim_n \tilde{u}_{it}^n$ is orthogonal to $\mathscr{A}(y)$ and since the orthogonal projection is unique, the result follows. ■

To conclude the proof of Theorem 3.1.1, we finally have to show that the double sequence (e_{it}), $i \in \mathbb{N}$, $t \in \mathbb{Z}$, from (3.1.2) is indeed weakly correlated.

Lemma 3.1.11. *Suppose that the conditions of Theorem 3.1.1 hold and let e_{it} be defined as in (3.1.35). Then $(e_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ is weakly correlated.*

Proof. From Lemma 3.1.10 we know, that $e_{it} = \lim_n \tilde{u}_{it}^n = \underline{o}_{2i}^n(z)O_2^n(z)^* y_t^n$. Let f_e^n denote the $(n \times n)$ spectral density matrix of $e_t^n = (e_{1t}, \ldots, e_{nt})'$, then with Lemma 3.1.4 and Lemma 3.1.10, f_e^n can be written as the $(n \times n)$ top-left submatrix of the limit of corresponding PCA spectral densities, i.e. for $m \geq n$

$$f_e^n = \lim_{m \to \infty} O_{2[1,\ldots,n]}^m \Omega_2^m O_{2[1,\ldots,n]}^{m\,*}, \tag{3.1.39}$$

where the numbers in square brackets denote the relevant rows of the matrix. Hence the first eigenvalue of f_e^n is certainly not larger than $\sup_m \omega_{q+1}^m$ for any λ in $[-\pi, \pi]$ and since this holds true for any $n \in \mathbb{N}$ the result follows. ■

Remark 3.1.3. The last results draw a direct connection between the (dynamic) PC model and the GDFM: under the conditions of Theorem 3.1.1 these two models coincide asymptotically. We will get back to that point in the course of estimation of the underlying variables (see Chapter 5).

In the sequel we will summarize some important remarks and conclusions concerning the last results. Indeed, the following corollaries are (nearly) immediate consequences of Theorem 3.1.1 or of parts of the proof.

Corollary 3.1.1. *If the double sequence $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ allows for a q-GDFM representation, then $\mathscr{A}(y) = \mathbb{H}_\chi = \mathbb{H}_\xi$.*

Proof. It has already been shown, that if $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ has a q-GDFM representation, $\mathscr{A}(y) \subseteq \mathbb{H}_\chi$ (see remark 3.1.1) and since $\chi_{it} = \lambda_i(z)\xi_t$, for all $i \in \mathbb{N}$, we have, that $\mathscr{A}(y) \subseteq \mathbb{H}_\chi \subseteq \mathbb{H}_\xi$.

On the other hand, it has been shown that if $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ has a q-GDFM representation, $\mathscr{A}(y) = \mathbb{H}_z$, hence $\mathbb{H}_z \subseteq \mathbb{H}_\xi$, where both (z_t) and (ξ_t) are q-dimensional orthogonal white noise processes, which in turn implies equality: $\mathbb{H}_z = \mathbb{H}_\xi$. ■

Corollary 3.1.2. *If the double sequence $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ allows for a q-GDFM representation, then the latent variables χ_{it} are the projections of y_{it} onto the aggregate space $\mathscr{A}(y)$,*

$$\chi_{it} = \text{proj}(y_{it}|\mathscr{A}(y)).$$

Proof. Since by corollary 3.1.1 $\chi_{it} \in \mathscr{A}(y)$ and since $\chi_{jt-k} \perp u_{it}$ and therefore $u_{it} \perp \mathscr{A}(y)$, $\chi_{it} = \text{proj}(y_{it}|\mathscr{A}(y))$. ■

Remark 3.1.4. In particular, it follows that if (y_t^n), $n \in \mathbb{N}$, has a q-GDFM representation γ_t^n and e_t^n defined through the projection equations (3.1.2) are equivalent to χ_t^n and u_t^n in (2.1.1).

As an immediate consequence of the uniqueness of the orthogonal projection, the decomposition of the observed process into latent variables and noise respectively, is unique. The following theorem precises this statement.

Theorem 3.1.2. *Suppose the double sequence $(y_{it}|i \in \mathbb{N}, t \in \mathbb{Z})$ allows for a q-GDFM representation (2.1.1). And suppose that there exists a p-dimensional white noise process (ζ_t), such that*

$$\begin{aligned} y_t^n &= C^n(z)\zeta_t + e_t^n \\ &= \gamma_t^n + e_t^n \end{aligned}$$

where $C^n(z) = \sum_j C_j^n z^j, \sum_j \|C_j^n\| < \infty$, *and such that the* q*-th largest eigenvalue of* f_γ^n *and the largest eigenvalue of* f_e^n *fulfill assumption 2.1. Then,* $p = q$, $\gamma_t^n = \chi_t^n$ *and* $e_t^n = u_t^n$.

Remark 3.1.5. Corollaries 3.1.1 and 3.1.2 imply that both the latent variables as well as the noise, respectively, belong to $\mathbb{H}_y$. Furthermore, since by Theorem 3.1.2, these variables are uniquely identified, there exists no representation where they do not belong to $\mathbb{H}_y$.

Remark 3.1.6. Notice however, that only the latent variables (χ_t^n), but neither the factor loadings $\Lambda^n(z)$ nor the factors (ξ_t) are uniquely identifiable. As mentioned above, since we assumed the factors to be orthogonal white noise the factors are identifiable up to unitary transformations, i.e. transformations of the form $U(z)\xi_t$, where $U \in L_2^{q\times q}([-\pi,\pi],\mathbb{C})$ and $U(e^{-i\lambda})U(e^{-i\lambda})^* = I_q$ holds and $\Lambda^n(z)$ may then be transformed accordingly as $\Lambda^n(z)U(z)^*$ without changing the latent variables.

Remark 3.1.7. We have defined the q-GDFM such that the factor loadings $\Lambda^n(z)$ may correspond to a two-sided filter. For many purposes, such as structural modeling or forecasting, a causal representation is required and as will be shown in Section 3.3, Assumption 1.1d) is sufficient for the existence a causal representation.

3.2 The relation to factor models with idiosyncratic noise

In the sequel, we want to examine the relation between the classical factor analytic model with idiosyncratic noise (see Section 1.3) and the GDFM. Since we have introduced the GDFM as a generalization of the classical factor model, we first want to show, that indeed under weak assumptions on the factor loadings a factor structure with idiosyncratic noise implies a generalized factor structure, hence, that the conditions of Theorem 3.1.1 are satisfied.

Recall, that the factor model with idiosyncratic noise is specified through

$$f_y^n(\lambda) = \Lambda^n(e^{-i\lambda})\Lambda^n(e^{-i\lambda})^* + f_u^n(\lambda), \tag{3.2.1}$$

where $\Lambda^n(e^{-i\lambda}) = \sum_j \Lambda_j^n e^{-i\lambda j}$, $\Lambda_j^n \in \mathbb{R}^{n\times q}$, f_u^n is diagonal for all λ in $[-\pi,\pi]$ and where we implicitly assumed that the spectral density matrix of the factors equals I_q.

Under our general assumptions all diagonal entries, $f_{u,i}$, $i = 1,\dots,n$, say, of f_u^n are bounded from above for all $n \in \mathbb{N}$ and let $\mu(\lambda) = \sup_{i\in\mathbb{N}} f_{u,i}(\lambda)$. Then, applying corollary A.1.1 yields $\omega_{q+1}^n \le \mu(\lambda)$ for all $n \in \mathbb{N}$ and $\lambda \in [-\pi,\pi]$. For the second condition of Theorem 3.1.1, consider

$$\Lambda^n(e^{-i\lambda})^* f_y^n(\lambda)\Lambda^n(e^{-i\lambda}) \ge (\Lambda^n(e^{-i\lambda})^*\Lambda^n(e^{-i\lambda}))^2,$$

implying that if all eigenvalues of $\Lambda^{n*}\Lambda^n$ diverge a.e. in $[-\pi,\pi]$, so do the first q eigenvalues of f_y^n.

Summarizing, if for $n \in \mathbb{N}$, (y_t^n) allows for a factor model with orthogonal noise and if all eigenvalues of $\Lambda^{n*}\Lambda^n$ diverge a.e. in $[-\pi, \pi]$, then the conditions of Theorem 3.1.1 are satisfied and hence there exists a q-GDFM representation (2.1.1). Since by Theorem 3.1.2 this representation is unique, it follows, that the two models have to coincide. Moreover, since by Lemma 3.1.10, asymptotically the q-GDFM representation is equivalent to the dynamic q-PC model, if there exists a factor structure with orthogonal noise, dynamic PCA and factor analysis are asymptotically equivalent.

Let us analyze the relation between PCA and factor analysis in more detail. In PCA the factor loadings are obtained from the eigenvalue decomposition of f_y^n. However, in the factor analytic approach the factor loadings are related to another eigenvalue problem (see Lawley and Maxwell (1971)). From equation (3.2.1) we get that $\Lambda\Lambda^* = f_y^n - f_u^n$. Now, consider the following transformation of f_y^n, that rescales the variates, such that the spectral density of the noise becomes I_n (recall that we used the same derivation to obtain a normalization condition for Λ in the static case, see (1.3.1)):

$$\begin{aligned} \bar{f}_y^n(\lambda) &= f_u^n(\lambda)^{-1/2} f_y^n(\lambda) f_u^n(\lambda)^{-1/2} \\ &= f_u^n(\lambda)^{-1/2} \Lambda^n(e^{-i\lambda}) \Lambda^n(e^{-i\lambda})^* f_u^n(\lambda)^{-1/2} + I_n. \end{aligned} \tag{3.2.2}$$

$\bar{f}_y^n(\lambda) - I_n$ is Hermitian and of rank q for almost all λ in $[-\pi, \pi]$, hence it can be written as

$$\bar{f}_y^n(\lambda) - I_n = \bar{V}_1^n(e^{-i\lambda}) \Delta^n(\lambda) \bar{V}_1^n(e^{-i\lambda})^*, \tag{3.2.3}$$

where $\Delta^n(\lambda)$ is the diagonal $(q \times q)$ matrix having the non-zero eigenvalues of $\bar{f}_y^n(\lambda) - I_n$ in its diagonal and $\bar{V}_1^n(e^{-i\lambda})$ is the $(n \times q)$ matrix of corresponding eigenvectors. Hence the factor loadings Λ^n can be expressed as

$$\Lambda^n(e^{-i\lambda}) = f_u^n(e^{-i\lambda})^{1/2} \bar{V}_1^n(e^{-i\lambda}) \Delta^n(e^{-i\lambda})^{1/2}. \tag{3.2.4}$$

Since $\bar{f}_y^n(\lambda)$ has the same eigenvectors as $\bar{f}_y^n(\lambda) - I_n$, and since by equations (A.2.1) to (A.2.3), we know, that $V_1^n(e^{-i\lambda}) = f_u^n(\lambda)^{-1/2} \bar{V}_1^n(e^{-i\lambda})$ are the first q generalized eigenvectors of the couple of matrices $(f_y^n(\lambda), f_u^n(\lambda))$, we can write (3.2.4) as

$$\Lambda^n(e^{-i\lambda}) = f_u^n(e^{-i\lambda}) V_1^n(e^{-i\lambda}) \Delta^n(e^{-i\lambda})^{1/2}. \tag{3.2.5}$$

Hence, apart from rescaling, the difference between the factor loadings occurring in PCA and factor analysis is, that in PCA the factor loadings are the first eigenvectors of f_y^n with respect to the identity matrix and in factor analysis they are the first generalized eigenvectors of f_y^n with respect to f_u^n. Above we have argued that if there is a classical factor structure with idiosyncratic noise, these two procedures will asymptotically yield the same decomposition into latent variables and noise. Actually it can be shown, that even if there only exists a q-GDFM representation, then under weak conditions on a sequence of matrices Σ^n, the generalized PCA of (f_y^n, Σ^n) will asymptotically yield the same decomposition into latent variables and noise as conventional dynamic PCA. For the static case this has been shown in Chamberlain and

Rothschild (1983), the extension to the dynamic case is immediate given the results from the last section. In classical factor analysis with idiosyncratic noise, one of the main difficulties in ML-estimation is the estimation of f_u^n. Hence the last statement provides a justification (at least from an asymptotic point of view) of PCA, which is computationally much simpler than factor analysis.

3.3 Latent variables with rational spectral density: a system theoretic analysis

In this section we will only consider the latent variables (χ_t^n) and disregard the noise part (u_t^n). The aim of this section is a deeper analysis of the latent part using system theoretic methods, i.e. methods related to autoregressive moving-average (ARMA) and state space realizations of transfer functions.

Recall that by Assumptions 1.1 a) and 1.1 c) (χ_t^n) is stationary with $\mathbb{E}\,\chi_t^n = 0$ and $(n \times n)$-dimensional spectral density $f_\chi^n(\lambda) = \Lambda^n(e^{-i\lambda}) f_\xi(\lambda) \Lambda^n(e^{-i\lambda})^*$. By Assumption 1.1 d) f_χ^n is rational in $e^{-i\lambda}$ and of normal rank[1] $q < n$ for n sufficiently large. In Section 3.3.1 below we will see that any rational $(n \times n)$ spectral density f_χ^n of normal rank q may be factorized as $f_\chi^n(\lambda) = (2\pi)^{-1} w^n(e^{-i\lambda}) w^n(e^{-i\lambda})^*$, where the $(n \times q)$ transfer function $w^n(z)$ is itself rational in z, analytic for all $|z| \leq 1$ and of full column rank q for all $|z| \leq 1$ and thus corresponds to a causal GDFM representation. As a consequence, this factorization gives rise to realizations of w^n in terms of (causal) ARMA or state space systems (see Section 3.3.2). To ensure that these realizations are parsimonious, in that the dimension of the parameter space only grows linearly in n, we will impose the following additional assumption.

Assumption 3.1 (McMillan degree)**.**
Let r be the state dimension of a minimal state space realization of w^n, i.e. the McMillan degree of w^n. Then for n sufficiently large, r does not depend on n.

Roughly speaking Assumption 1.1 d) is a measure for the complexity reduction that is possible in the cross-sectional dimension and Assumption 3.1 is a measure for the complexity reduction that is possible in the time dimension.

Moreover as will be shown in Section 3.3.3, under Assumptions 1.1 and 3.1 there always exists a quasi-static representation of χ_t^n (and thus of y_t^n), that in general can be achieved at the cost of a higher dimensional quasi-static factor. We will see that two distinct cases may occur: (1) the minimal static factor coincides with the minimal state of a state space realization and hence follows an AR(1) process or (2) the minimal state is of larger dimension than the minimal static factor. In Section 3.3.4 we will then show that in either case the minimal static factor is generated by an ARMA(P, Q) system (where $P = 1$ and $Q = 0$ in case (1)) and a uniquely identifiable and parsimonious ARMA representation will be given. In Section 3.3.5 we will summarize the properties of the quasi-static GDFM representation as derived above. And in Section 3.3.6 the pure autoregressive representation of the static factors will be further analyzed.

It is emphasized that the assumptions imposed here are more general than the assumptions usually imposed in the context of causal modeling (see for instance Stock and Watson (2002b),

[1]The normal rank of a rational matrix $f(\lambda)$ defined on $\mathbb{C}$ is defined as $\max_\lambda \operatorname{rk} f(\lambda)$, and equals the rank of $f(\lambda)$ with the possible exception of finitely many points.

Forni *et al.* (2005a) and Forni *et al.* (2005b)), since usually case (2) is ruled out by requiring for instance that $\Lambda^n(z)$ is polynomial of degree S and that ξ_t follows an AR(s) process with $s \leq S+1$ (see also Example 3.3.2 below).

3.3.1 Factorization of spectral densities and causal GDFM representations

The following theorem is a well-known result about the factorization of (singular) rational spectral densities and may be found in Rozanov (1967) or Hannan (1970).

Theorem 3.3.1. *An $(n \times n)$-dimensional rational spectral density matrix f_χ^n of normal rank $q \leq n$ can be factorized as*

$$f_\chi^n(\lambda) = \frac{1}{2\pi} w^n(e^{-i\lambda}) w^n(e^{-i\lambda})^*, \tag{3.3.1}$$

where the $(n \times q)$-dimensional matrix function $w^n(z)$ is rational in z, analytic for $|z| \leq 1$ and has rank q for all $|z| \leq 1$.

Proof. See Rozanov (1967) or Hannan (1970). ■

As a consequence, (χ_t^n) can be expressed as a causal linear transformation of a q-dimensional white noise process (ε_t):

$$\chi_t^n = w^n(z)\varepsilon_t = \sum_{j=0}^{\infty} W_j^n \varepsilon_{t-j}, \tag{3.3.2}$$

with $W_j^n \in \mathbb{R}^{n \times q}$, $j = 0, 1, \ldots$, $\sum_{j=0}^{\infty} \|W_j^n\|^2 < \infty$ and $\Gamma_\varepsilon = \mathbb{E}\,\varepsilon_t \varepsilon_t' = I_q$.

Let $\mathbb{H}_\chi^n(t^-)$ and $\mathbb{H}_\varepsilon(t^-)$ denote the Hilbert spaces spanned by $\{\chi_{i\tau} | i = 1, \ldots, n; \tau \leq t\}$ and $\{\varepsilon_{i\tau} | i = 1, \ldots, q; \tau \leq t\}$ respectively, then obviously (3.3.2) implies $\mathbb{H}_\chi^n(t^-) \subseteq \mathbb{H}_\varepsilon(t^-)$. Further, recall that w^n may be written as

$$w^n = ulv, \tag{3.3.3}$$

where the $(n \times n)$ and $(q \times q)$ polynomial matrices u and v are *unimodular*[2] and where the $(n \times q)$ rational matrix l, i.e. the *Smith-McMillan form* of w^n, is diagonal and displays the poles and zeros of w^n (see Hannan and Deistler (1988), Lemma 2.4.3). Defining a $(q \times n)$ matrix function,

$$w^{n-} = v^{-1}(l'l)^{-1}l'u^{-1}, \tag{3.3.4}$$

it is easily seen, that, since w^n has no poles and zeros for $|z| \leq 1$, the same holds for l and thus by its diagonality for $l'l$, w^{n-} is rational and has no poles and zeros for $|z| \leq 1$. Hence w^{n-} can be extended in a power series that is convergent in an open disc containing the unit disc and is thus a one-sided (causal) left inverse of w^n. As a consequence (ε_t) may also be represented as a causal linear transformation of (χ_t^n) implying that $\mathbb{H}_\chi^n(t^-) = \mathbb{H}_\varepsilon(t^-)$ holds. Hence (3.3.2) is the *Wold representation* of (χ_t^n) and the q-dimensional white noise (ε_t) forms a basis of the *linear innovations* of (χ_t^n), i.e. (ε_t) is *fundamental*.

[2] A polynomial matrix u is called unimodular if $\det u$ is constant, not equal to 0. A polynomial matrix u is unimodular iff its inverse u^{-1} exists and is a polynomial matrix.

Remark 3.3.1. If w^n is *zeroless*, i.e. if w^n has full column rank q for all $z \in \mathbb{C}$, it follows that the numerators of the Smith-McMillan form l in (3.3.3) are all equal to 1, implying that $(l'l)^{-1}$ and thus w^{n-} is polynomial. Hence in this case, ε_t depends only on a finite number of past and present χ_t^n.

Clearly, for given f_χ^n, the spectral factor w^n is not uniquely identifiable, which may be seen by the argument that post multiplying w^n by a constant orthogonal $(q \times q)$ matrix P with $P'P = PP' = I_q$ yields another feasible factorization of the form (3.3.1). The next theorem ensures, that post multiplication by a constant orthogonal $(q \times q)$ matrix is indeed the only indeterminacy.

Theorem 3.3.2. *Let $f^n(\lambda)$ satisfy the conditions of Theorem 3.3.1, then the factorization in (3.3.1) is unique up to post multiplication by a constant orthogonal matrix P with $P'P = PP' = I_q$.*

Proof. Let (χ_t^n) be an n-dimensional process with spectral density $f^n(\lambda)$. By what was said above due to its full rank for $|z| \leq 1$, the spectral factor $w^n(z)$ corresponds to the Wold representation $\chi_t^n = w^n(z)\varepsilon_t$ and (ε_t) is fundamental. It is well known that a fundamental process is defined up to multiplication by a unitary matrix P (see Rozanov (1967), Sec. II.3. and II.4.). Since we only consider real processes, the result follows. ■

Furthermore, for a sequence of nested spectral densities f_χ^n, $n \in \mathbb{N}$, it is immediately clear, that the spectral factors w^n, $n \in \mathbb{N}$, are again nested. Consequently (ε_t) in representation (3.3.2) is a q-dimensional factor process, sometimes also called a q-dimensional *fundamental shock*, and w^n are the corresponding rational and causal factor loadings, where (ε_t) and w^n are identifiable up to right multiplication of w^n by a constant orthogonal matrix and left multiplication of ε_t by its transpose, i.e. up to static factor rotations. Thus, assuming that f_χ^n is rational, is a sufficient condition for the existence of a causal GDFM representation.

The tall $(n \times q)$ transfer function w^n may then be realized as

- ARMA system
$$w^n(z) = a^n(z)^{-1}b^n(z),$$
where the $(n \times n)$ and $(n \times q)$ matrices $a^n(z)$ and $b^n(z)$ are relatively left prime polynomials in z and where $\det a^n(z) \neq 0$ for $|z| \leq 1$ (*stability assumption*) and $\operatorname{rk} b^n(z) = q$ for $|z| \leq 1$ (*miniphase assumption*) hold.

- Right matrix fraction description (right MFD)
$$w^n(z) = c^n(z)d(z)^{-1},$$
where $c^n(z)$ and $d(z)$ are $(n \times q)$ and $(q \times q)$ polynomial matrices respectively, that are relatively left prime and where $\det d^n(z) \neq 0$ for $|z| \leq 1$ and $\operatorname{rk} c^n(z) = q$ for $|z| \leq 1$ hold. This is the general form of the realization considered by Forni *et al.* (2005a).

- State space system (A, B, C^n),

$$x_t = Ax_{t-1} + B\varepsilon_t \tag{3.3.5}$$
$$\chi_t^n = C^n x_t, \tag{3.3.6}$$

with r-dimensional state vector x_t ($r \geq q$) and parameter matrices $A \in \mathbb{R}^{r\times r}$, $B \in \mathbb{R}^{r\times q}$, $C^n \in \mathbb{R}^{n\times r}$ and hence $w^n(z) = C^n(I - Az)^{-1}B$, where the stability assumption $\det(I - Az) \neq 0$ for $|z| \leq 1$ (or equivalently denoting the largest eigenvalue of A by $\omega_1(A)$, $|\omega_1(A)| < 1$) and the conjugated miniphase assumption $\operatorname{rk} C^n B = q$ and

$$\operatorname{rk}\begin{pmatrix} \frac{1}{z}I - A & -B \\ C^n & 0 \end{pmatrix} = r + q \text{ for} |z| \leq 1, z \neq 0$$

hold. Notice that, since only the matrix C^n, but neither of the matrices A and B depend on n, the dimension of the parameter space is proportional to n and that the parametrization of the system is thus quite sparse. This realization will be discussed below in more detail.

Before proceeding, let us examine two commonly used examples of latent variables with rational spectral density.

Example 3.3.1. Let the factor loading matrix $\Lambda^n(z)$ be polynomial of degree s, i.e.

$$\Lambda^n(z) = \sum_{j=0}^{s} \Lambda_j^n z^j, \Lambda_j^n \in \mathbb{R}^{n\times q}$$

and let the factors (ξ_t) be a q-dimensional orthonormal white noise process, i.e. $\xi_t = \varepsilon_t$ with $\mathbb{E}\,\varepsilon_t\varepsilon_t' = I_q$, $\mathbb{E}\,\varepsilon_t\varepsilon_{t-j}' = 0_q, j \neq 0$. Then $w^n(z) = \Lambda^n(z)$ and by defining the state as the sq-dimensional vector of stacked factors $x_t = (\xi_t', \xi_{t-1}', \ldots, \xi_{t-s}')'$ we get a state space realization

$$x_t = \underbrace{\begin{pmatrix} 0 & 0 & \ldots & 0 \\ I_q & 0 & \ldots & 0 \\ 0 & \ddots & & 0 \\ 0 & & I_q & 0 \end{pmatrix}}_{A\in\mathbb{R}^{(sq\times sq)}} x_{t-1} + \underbrace{\begin{pmatrix} I_q \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{B\in\mathbb{R}^{(sq\times q)}} \varepsilon_t \tag{3.3.7}$$

$$\chi_t^n = \underbrace{(\Lambda_0^n, \Lambda_1^n, \ldots, \Lambda_s^n)}_{C^n\in\mathbb{R}^{(n\times sq)}} x_t \tag{3.3.8}$$

If we further assume that $\operatorname{rk}(\mathbb{E}\,\chi_t^n\chi_t^{n'}) = \operatorname{rk}(\Lambda_0^n, \Lambda_1^n, \ldots, \Lambda_s^n) = sq$ holds, then x_t may be seen as a minimal quasi-static factor and (3.3.8) as the quasi-static representation of χ_t^n that has been made possible at the cost of a higher dimensional factor. ■

Example 3.3.2. Let $\Lambda^n(z)$ be of the same form as in the example above and let the factor (ξ_t) be a stationary AR(p)-process: $a(z)\xi_t = \varepsilon_t$, where $a(z) = A_0 - A_1 z - \ldots A_p z^p$, $A_j \in \mathbb{R}^{(q\times q)}$ with

$\operatorname{rk} A_0 = q$, $\det a(z) \neq 0$ for $|z| \leq 1$ and where (ε_t) is a q-dimensional orthonormal white noise process. Then $w^n(z) = \Lambda^n(z)a^{-1}(z)$. For $s \leq p+1$, this is the model considered for instance by Forni *et al.* (2005a). Setting $S = \max(p, s-1)$ and defining x_t as the Sq-dimensional vector of stacked factors $x_t = (\xi_t', \xi_{t-1}', \ldots, \xi_{t-S}')'$ we have a state space realization

$$x_t = \underbrace{\begin{pmatrix} A_0^{-1}A_1 & A_0^{-1}A_2 & \ldots & A_0^{-1}A_p & 0 \\ I_q & 0 & \ldots & & 0 \\ 0 & \ddots & & & 0 \\ 0 & & & I_q & 0 \end{pmatrix}}_{A \in \mathbb{R}^{(Sq \times Sq)}} x_{t-1} + \underbrace{\begin{pmatrix} A_0^{-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{B \in \mathbb{R}^{(Sq \times Sq)}} \varepsilon_t \tag{3.3.9}$$

$$\chi_t^n = \underbrace{(\Lambda_0, \Lambda_1, \ldots, \Lambda_s, 0_{n \times q(S-p)})}_{C^n \in \mathbb{R}^{(n \times Sq)}} x_t. \tag{3.3.10}$$

Consider the case $p \leq (s-1)$, then obviously the last $q(S-p)$ columns of C^n contain only zeros, and thus $\operatorname{rk} C < Sq$, implying that the minimal static factor does not coincide with the state vector x_t. Notice however, that the state space realizations given in these examples might be inconvenient in that the state vectors might be of higher dimension than necessary. ■

3.3.2 A state space representation of (χ_t^n)

Here we are going to show that any transfer function $w^n(z)$ corresponding to the Wold representation (3.3.2) and thus satisfying the conditions of Theorem (3.3.1) may indeed be realized as a state space realization of the form (3.3.5)- (3.3.6) by specifying a construction. This state space realization will be minimal, i.e. no other state space realization with lower dimensional state exists. For general aspects of the construction of state space realizations, we refer to Hannan and Deistler (1988) or Deistler (2001a).

Using equation (3.3.2) we may write

$$\begin{pmatrix} \chi_t^n \\ \chi_{t+1}^n \\ \chi_{t+2}^n \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} W_0^n & W_1^n & W_2^n & \ldots \\ W_1^n & W_2^n & W_3^n & \ldots \\ W_2^n & W_3^n & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}}_{\mathscr{H}_\infty^n} \underbrace{\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix}}_{\varepsilon_t^-} + \begin{pmatrix} 0 \\ W_0^n \varepsilon_{t+1} \\ W_0^n \varepsilon_{t+2} + W_1^n \varepsilon_{t+1} \\ \vdots \end{pmatrix}. \tag{3.3.11}$$

The matrix $\mathscr{H}_\infty^n$ is in block Hankel form (i.e. the (i,j) block only depends on the sum $i+j$) and is called the *Hankel matrix* of the transfer function $w^n(z)$. In the sequel $h^n(i,j)$ will denote the j-th row in the i-th block row of $\mathscr{H}_\infty^n$.

It can be shown that if $w^n(z)$ is rational then $\mathscr{H}_\infty^n$ is of finite rank, r say (for a proof see Hannan and Deistler (1988), Theorem 2.4.1.), and since $\operatorname{rk} w^n(z) = q$ for $|z| \leq 1$ implies that $\operatorname{rk} W_0^n = q$, then $r \geq q$. Hence there must be r linearly independent rows of $\mathscr{H}_\infty^n$ that form a basis for the

row space of $\mathscr{H}_\infty^n$. From the block Hankel form of $\mathscr{H}_\infty^n$ one immediately sees, that if $h^n(i,j) \in \text{span}\{h^n(i_1,j_1),\ldots,h^n(i_k,j_k)\}$, then $h^n(i+1,j) \in \text{span}\{h^n(i_1+1,j_1),\ldots,h^n(i_k+1,j_k)\}$. As a consequence, we may consider a basis with the property that, if $h^n(i+1,j)$ is in the basis, so is $h^n(i,j)$. Such a basis can be described by a *structure index* $\alpha = (r_1, r_2, \ldots, r_n)$ indicating that the $r = r_1 + \ldots + r_n$ basis rows are $h^n(1,1),\ldots,h^n(r_1,1),\ldots,h^n(1,n),\ldots,h^n(r_n,n)$, i.e. the i-th row of a block row is a basis row up to block row r_i. Let S_α^n be the $(r \times \infty)$-dimensional selector matrix (i.e. a matrix where each row contains only zeros except for a one in the column that corresponds to the selected row) that selects this row basis and write $\mathscr{H}_\alpha^n = S_\alpha^n \mathscr{H}_\infty^n$. Then we may define the r-dimensional state vector x_t^n, following the approach described in Kalman (1963), as

$$x_t^n = \mathscr{H}_\alpha^n \varepsilon_t^- . \tag{3.3.12}$$

Notice, that Assumption 3.1 is equivalent to assuming that $\text{rk}\,\mathscr{H}_\infty^n = r$ for n sufficiently large, where we are already anticipating the fact that the state vector x_t^n in 3.3.12 is of minimal possible dimension, which will be shown below. Moreover, as we will also see below, Assumption 3.1 implies that x_t^n, as well as the coefficient matrices A and B in (3.3.5) can be chosen independently of n. Now, let n_0 be the smallest cross-sectional dimension for which $\text{rk}\,\mathscr{H}_\infty^{n_0} = r$ holds. Obviously the rows of $\mathscr{H}_\infty^{n_0}$ are contained in any Hankel matrix $\mathscr{H}_\infty^n$ for $n \geq n_0$. Hence if $\mathscr{H}_\alpha^{n_0}$ are r linearly independent rows of $\mathscr{H}_\infty^{n_0}$ and thus form a basis they are also a basis of $\mathscr{H}_\infty^n$ for $n \geq n_0$. Consequently the structure index α can be chosen independently of n and so can the $(r \times \infty)$ matrix of basis rows $\mathscr{H}_\alpha = \mathscr{H}_\alpha^{n_0}$, hence we may rewrite (3.3.12) as

$$x_t = \mathscr{H}_\alpha \varepsilon_t^- . \tag{3.3.13}$$

Proceeding to x_{t+1} we have

$$\begin{aligned} x_{t+1} &= S_\alpha^n \mathscr{H}_\infty^n \varepsilon_{t+1}^- \\ &= S_\alpha^n \begin{pmatrix} W_1^n & W_2^n & W_3^n & \ldots \\ W_2^n & W_3^n & W_4^n & \ldots \\ W_3^n & W_4^n & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \end{pmatrix} + S_\alpha^n \begin{pmatrix} W_0^n \\ W_1^n \\ W_2^n \\ \vdots \end{pmatrix} \varepsilon_{t+1}, \end{aligned} \tag{3.3.14}$$

where the Hankel matrix in the first term on the right hand side is a submatrix of $\mathscr{H}_\infty^n$ and thus can be expressed as a linear transformation of the row basis $\mathscr{H}_\alpha$:

$$S_\alpha^n \begin{pmatrix} W_1^n & W_2^n & W_3^n & \ldots \\ W_2^n & W_3^n & W_4^n & \ldots \\ W_3^n & W_4^n & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix} = S_\alpha^n \tilde{A}^n \mathscr{H}_\alpha = A \mathscr{H}_\alpha . \tag{3.3.15}$$

Again, the matrix A can be chosen independently of n, since the structure index α is independent of n and S_α^n will always select the same rows of $\tilde{A}^n$.

Now, by setting

$$B = S_\alpha^n(W_0^n, W_1^n, W_2^n, \ldots)', \tag{3.3.16}$$

which is independent of n by the same argument we get

$$\begin{aligned} x_{t+1} &= A\mathscr{H}_\alpha \varepsilon_t^- + B\varepsilon_{t+1} \\ &= Ax_t + B\varepsilon_{t+1}. \end{aligned}$$

Furthermore $\chi_t^n = (W_0^n, W_1^n, W_2^n, \ldots)\varepsilon_t^-$ and since $(W_0^n, W_1^n, W_2^n, \ldots)$ is the first block row of $\mathscr{H}_\infty^n$ we may express it as

$$(W_0^n, W_1^n, W_2^n, \ldots) = C^n \mathscr{H}_\alpha, \tag{3.3.17}$$

and thus

$$\chi_t^n = C^n x_t,$$

where C^n is nested by construction (i.e. for $m > n$ the first n rows of C^n and C^m are identical), which concludes the construction of the state space system.

We are now going to introduce some important concepts for dealing with state space systems (see Hannan and Deistler (1988) for a more detailed presentation).

A state space realization (A, B, C^n) of a transfer function $w^n(z) = C^n(I - Az)^{-1}B$ is said to be *minimal* if A is of minimal possible dimension among all realizations $(\tilde{A}, \tilde{B}, \tilde{C^n})$ of $w^n(z)$.

A state space system (A, B, C^n) (or the pair (A, B)) is called *controllable* if the matrix

$$\mathscr{C}_r = (B, AB, \ldots, A^{r-1}B) \quad \in \mathbb{R}^{r \times rq} \tag{3.3.18}$$

is of full row rank r. $\mathscr{C}_r$ is called *controllability matrix*. If a system (A, B, C^n) is controllable and if $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1$ were under our control, then for $t \geq r$ the equation system

$$\begin{aligned} x_t &= Ax_{t-1} + B\varepsilon_t \\ &= \sum_{j=0}^{t-1} A^j B\varepsilon_{t-j} + A^t x_0 \\ &= (B, AB, \ldots, A^{t-1}B)(\varepsilon_t', \varepsilon_{t-1}', \ldots, \varepsilon_1')' + A^t x_0 \end{aligned} \tag{3.3.19}$$

is solvable for any x_t and any initial state x_0. Hence controllability means that if the innovations were manipulable, given an arbitrary initial state x_0, any state x_t could be reached.

A state space system (A, B, C^n) (or the pair (A, C^n)) is called *observable* if the matrix

$$\mathscr{O}_r = \begin{pmatrix} C^n \\ C^n A \\ \vdots \\ C^n A^{r-1} \end{pmatrix} \quad \in \mathbb{R}^{rn \times r} \tag{3.3.20}$$

is of full column rank r. $\mathscr{O}_r$ is called *observability matrix*. Assume that there were no innovations ε_t, then $\chi_t^n = C^n x_t = C^n A^t x_0$ for some initial state x_0. If (A, C^n) is observable and if there are no innovations ε_t, then the initial state x_0 will be uniquely determined (observed) from

$$\begin{pmatrix} \chi_0^n \\ \vdots \\ \chi_{r-1}^n \end{pmatrix} = \mathscr{O}_r x_0. \tag{3.3.21}$$

Notice that the rank of neither $\mathscr{C}_r$ nor $\mathscr{O}_r$ can be increased further by adding terms of the form $A^j B$ and CA^j, respectively, $j \geq r$ (Caley-Hamilton Theorem). Hence the infinite controllability and observability matrices

$$\mathscr{C} = (B, AB, A^2B, \ldots) \tag{3.3.22}$$

$$\mathscr{O} = (C^{n\prime}, A'C^{n\prime}, A'^2C^{n\prime}, \ldots)' \tag{3.3.23}$$

have the same rank as $\mathscr{C}_r$ and $\mathscr{O}_r$ respectively.

For a state space realization (A, B, C^n) of $w^n(z) = C^n(I - Az)^{-1}B = C^nB + C^nABz + C^nA^2Bz^2 + \ldots$ the Hankel matrix $\mathscr{H}_\infty^n$ is equal to the product of the infinite observability and controllability matrices:

$$\mathscr{H}_\infty^n = \mathscr{O}\mathscr{C} = \begin{pmatrix} C^nB & C^nAB & C^nA^2B & \ldots \\ C^nAB & C^nA^2B & C^nA^3B & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}. \tag{3.3.24}$$

It can be shown that a state space system (A, B, C^n) is minimal if and only if it is controllable and observable (see Hannan and Deistler (1988), proof of Theorem 2.3.3.). Consequently, the system (A, B, C^n) with r-dimensional state vector x_t defined as in (3.3.12) and parameter matrices A, B, C^n as in (3.3.15) - (3.3.17) is by construction minimal since $r = \operatorname{rk} \mathscr{H}_\infty^n$.

A minimal state space system (A, B, C^n) (or the matrix A) fulfills the *stability assumption* if the largest eigenvalue of A is strictly smaller than 1 in modulus,

$$|\omega_1(A)| < 1.$$

Clearly, $w^n(z) = C^n(I - Az)^{-1}B$ has a pole at $z_0 \in \mathbb{C}$ if and only if $\det(I - Az_0) = 0$ which holds if and only if $\frac{1}{z_0}$ is an eigenvalue of A. Thus, if w^n has no poles for $|z| \leq 1$, then A will satisfy the stability assumption.

A minimal state space system (A, B, C^n) fulfills the *miniphase assumption* if $\operatorname{rk}(C^nB) = q$ and if

$$\operatorname{rk} \begin{pmatrix} \frac{1}{z}I - A & -B \\ C^n & 0 \end{pmatrix} = r + q \text{ for} |z| \leq 1, z \neq 0.$$

The first condition is equivalent to $\operatorname{rk} w^n(0) = q$. The second condition is satisfied if $\operatorname{rk} w^n(z) = q$ for $|z| \leq 1, z \neq 0$. To see this, suppose for a moment that the second condition failed. Then there exists an $(r+q)$-dimensional vector $t = (t_1', t_2')'$, t_1 and t_2 not both equal to 0, such that for some $|z_0| < 1, z_0 \neq 0$,

$$\begin{aligned} 0 &= (I - Az_0)t_1 - z_0 B t_2 \\ 0 &= C^n t_1. \end{aligned}$$

Indeed if we assume $\operatorname{rk} B = q$, both t_1 and t_2 must be nonzero, and

$$0 = C^n (I - Az_0)^{-1} z_0 B t_2,$$

implying that $w^n(z_0) = C^n(I - Az_0)^{-1}B$ does not have full column rank. Hence, if either of the two conditions failed, $w^n(z)$ would not be of full rank for all $|z| < 1$.

Summarizing, we have proved the following theorem.

Theorem 3.3.3. *Let $(w^n(z))$, $n \in \mathbb{N}$, be a sequence of nested rational $(n \times q)-$dimensional transfer functions, such that for each n sufficiently large $w^n(z)$ is analytic for $|z| \leq 1$ and of constant rank q for $|z| \leq 1$ and suppose that Assumption 3.1 holds. Then for n sufficiently large, $w^n(z)$ may be expressed as $C^n(I - Az)^{-1}B$, where (C^n) is a sequence of nested matrices, (A, B) is controllable, (A, C^n) is observable and where (A, B, C^n) satisfies the stability and miniphase assumptions.*

3.3.3 Two distinct cases of state space realizations (A, B, C^n)

In the sequel we will assume that (A, B, C^n) is a minimal state space realization of $w^n(z)$ and we are going to examine the two distinct cases $\operatorname{rk} C^n = r$ (subsequently referred to as the generic case) and $\operatorname{rk} C^n < r$ (subsequently referred to as the non-generic case) in more detail. Notice already, that since $\Gamma_\chi^n = C^n \mathbb{E}\, x_t x_t' C^{n'}$ and since $\mathbb{E}\, x_t x_t'$ has full rank r, we have $\operatorname{rk} C^n = \operatorname{rk} \Gamma_\chi^n$, where the latter is equal to $s \geq q$ independently of n by Assumption 1.1 e). Hence in the generic case $r = s$ holds, whereas in the non-generic case $r > s$.

Let us first get back to equation (3.3.11): since (3.3.11) corresponds to the Wold representation, the infinite vector on the right hand side contains the prediction errors of the best linear predictions given information up to time t. Hence, denoting the best linear prediction of χ_{t+h}^n given $\chi_\tau^n, \tau \leq t$ by $\chi_{t+h|t}^n$, we obtain

$$\begin{pmatrix} \chi_t^n \\ \chi_{t+1|t}^n \\ \chi_{t+2|t}^n \\ \vdots \end{pmatrix} = \mathscr{H}_\infty^n \varepsilon_t^-. \tag{3.3.25}$$

Since the linear dependence structure has to be the same on both sides of (3.3.25), a structure index α indicating basis rows of $\mathscr{H}_\infty^n$ also indicates (scalar) components of $(\chi_t^n, \chi_{t+1|t}^n, \chi_{t+2|t}^n, \ldots)'$

that form a basis for the Hilbert space $\overline{\text{span}}(\chi_{i,t+h|t}|i = 1,\dots,n, h \geq 0)$. Hence the minimal state x_t defined as in (3.3.13) is itself a basis for $\overline{\text{span}}(\chi_{i,t+h|t}|i = 1,\dots,n, h \geq 0)$ and thus every information needed to predict future values of χ_t^n given the past of χ_t^n up to time t is contained in the state vector x_t. In other words, for given state x_t the future and the past of the process (χ_t^n) are conditionally uncorrelated (*splitting property* of the state).

Let us now consider a special selection of basis rows of $\mathscr{H}_\infty^n$, namely the first (in natural order) linear independent rows of $\mathscr{H}_\infty^n$. This selection clearly has the property, that if $h(i+1,j)$ is in the basis so is $h(i,j)$ and the corresponding structure index $\alpha = (r_1,\dots,r_n)$ is called the *Kronecker indices* of $\mathscr{H}_\infty^n$. For a given Hankel matrix (or transfer function respectively) Kronecker indices are unique by construction. Furthermore, under Assumption 3.1 the non-zero Kronecker indices are independent of n.

Then we see, that in the generic case, since C^n is of full rank r, there must be r linearly independent scalar components of χ_t^n, hence all basis rows of $\mathscr{H}_\infty^n$ corresponding to the Kronecker indices are to be found in the first block row of $\mathscr{H}_\infty^n$ and there are exactly r Kronecker indices equal to 1 (and the rest are equal to 0). Therefore in the generic case the minimal state x_t is contained in the space spanned only by present (scalar) variables $\chi_{it}, i = 1,\dots,n$ and may be seen as a minimal (quasi-) static factor with (3.3.6) being a minimal static factor representation, and where the static factor process x_t follows the AR(1)-process (3.3.5).

Conversely, in the non-generic case, since $\operatorname{rk} C^n = s$, where $q \leq s < r$, there are less then r linearly independent scalar components of χ_t^n, hence only s of the basis rows of $\mathscr{H}_\infty^n$ corresponding to the Kronecker indices are to be found in the first block row of $\mathscr{H}_\infty^n$, and (some of) the s non-zero Kronecker indices are greater than 1. Consequently in the non-generic case the minimal state x_t is not contained in the space only spanned by present (scalar) variables $\chi_{it}, i = 1,\dots,n$ and is no minimal (quasi-) static factor. Nevertheless, as we are going to show below, in the non-generic case there also exists a quasi-static factor model representation, where now the static factor follows an ARMA(P,Q)-process with orders P and Q depending on the maximum Kronecker index.

In the sequel we will assume that C^n is of the form $C^n = (C_1^n, 0_{n\times(r-s)})$. This does not mean any loss of generality, since given a (minimal) state space system $(\tilde{A}, \tilde{B}, \tilde{C}^n)$ where $\tilde{A} \in \mathbb{R}^{s\times s}$, $\tilde{B} \in \mathbb{R}^{s\times q}$ and $\tilde{C}^n \in \mathbb{R}^{n\times s}$ with $\operatorname{rk}\tilde{C}^n = s < r$ for all n sufficiently large, corresponding to $w^n(z)$, then there always exists a regular $(r \times r)$ matrix T independent of n, such that $\tilde{C}^n T^{-1} = C^n = (C_1^n, 0)$, where C_1^n is $n \times s$ and of full column rank s. Partitioning accordingly as

$$T\tilde{x}_t = x_t = \begin{pmatrix} f_t \\ x_t^{(2)} \end{pmatrix},\ T\tilde{A}T^{-1} = A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},\ T\tilde{B} = B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \tag{3.3.26}$$

we can rewrite the state space representation of χ_t^n as

$$\begin{aligned} f_t &= A_{11} f_{t-1} + A_{12} x_{t-1}^{(2)} + B_1 \varepsilon_t && (3.3.27) \\ x_t^{(2)} &= A_{21} f_{t-1} + A_{22} x_{t-1}^{(2)} + B_2 \varepsilon_t && (3.3.28) \\ \chi_t^n &= C_1^n f_t. && (3.3.29) \end{aligned}$$

Furthermore, as can be easily seen, we may assume throughout w.l.o.g. that $\Gamma_f = \mathbb{E}\, f_t f_t' = I_s$. Then f_t is identifiable up to static rotations.

Now, since C_1^n is of full column rank s, we may express f_t as

$$f_t = C_1^{n+} \chi_t^n \tag{3.3.30}$$

where $C_1^{n+} = (C_1^{n\prime} C_1^n)^{-1} C_1^{n\prime}$ is the pseudoinverse of C_1^n, hence the s-dimensional vector f_t is contained in the space spanned by present (scalar) variables $\chi_{it}, i = 1, \ldots, n$ and is thus a minimal static factor and (3.3.29) a static representation. Furthermore $\operatorname{rk} C_1^{n+} = s$ and (3.3.29) and (3.3.30) imply that for any $t \in \mathbb{Z}$, $\mathbb{H}_\chi(t^-) = \mathbb{H}_f(t^-)$, hence the prediction of χ_{t+h}^n given values χ_τ^n, $\tau \leq t$ only depends on the prediction of the static factor:

$$\chi_{t+h|t}^n = C_1^n f_{t+h|t}. \tag{3.3.31}$$

However, since due to observability of (A, C^n), A_{12} in (3.3.27) cannot be 0 (for otherwise the system would not be minimal), the dynamics of f_t and thus its prediction do still depend on $x_t^{(2)}$. Or, in other words, if $s < r$ the static factor does not allow for an AR(1)-representation. Eventually, observe that (3.3.30) implies that f_t can be written as

$$f_t = C_1^{n+} w^n(z) \varepsilon_t = k(z) \varepsilon_t = \sum_{j=0}^{\infty} K_j \varepsilon_{t-j}, \tag{3.3.32}$$

where

$$K_j = (I_s, 0_{s \times (r-s)}) A^j B \in \mathbb{R}^{s \times q}. \tag{3.3.33}$$

And it follows immediately that $k(z)$ preserves the properties of $w^n(z)$ given in Theorem 3.3.1, i.e. the $(s \times q)$ matrix $k(z)$ is rational in z, analytic for $|z| \leq 1$ and has rank q for $|z| \leq 1$.

3.3.4 A canonical singular ARMA realization

Here we are going to derive an ARMA representation for the static factor f_t. The difficulty here is indeed not to show that f_t has an ARMA representation, which is quite obvious, but to find a good ARMA representation. First recall, that in general ARMA realizations are highly non unique, that is for any given rational transfer function k many different *matrix fraction descriptions* (MFDs) $k = a^{-1} b$, where a and b are polynomial matrices respectively, corresponding to different *observationally equivalent* ARMA realizations, exist. Second, many of these realizations are not convenient, in that for example the number of real parameters involved is too high. Of course, both these aspects are of great importance for estimation (see for instance Lütkepohl

(2005), Section 7.1 or Hannan and Deistler (1988), Section 2.3 for detailed discussions of these problems). Thus, here we want to find an ARMA representation that is identifiable, in that for any given transfer function k a unique parametrization can be specified, and parsimonious, in that the dimension of the parameter space is small. Recall further, that the situation here differs from the standard ARMA case since (f_t) may be *dynamically singular*, i.e. (f_t) may have a singular spectral density, or in other words the static factors f_t may be of higher dimension than the fundamental shocks ε_t, resulting in a transfer function k that is non-square (or tall). In the standard ARMA case, a number of identifiable realizations have been analyzed, e.g. the *final equations form* or the *reversed echelon form* (see the references cited above for detailed descriptions) where the latter is usually preferable because it often yields parameter spaces of lower dimension. The realization and its derivation we are going to present here is thus similar to the reversed echelon ARMA form, but adapted to the singular case.

Denote the Hankel matrix of $k(z)$ by $\mathscr{H}_\infty(k)$ and the submatrix consisting of the first r block rows of $\mathscr{H}_\infty(k)$ by $\mathscr{H}_r(k)$. Then with (3.3.31) we may write

$$\begin{pmatrix} \chi_t^n \\ \chi_{t+1|t}^n \\ \vdots \\ \chi_{t+r-1|t}^n \end{pmatrix} = \underbrace{\operatorname{diag}(C_1^n, \dots, C_1^n)}_{(rn \times rs)} \mathscr{H}_r(k)\varepsilon_t^-, \tag{3.3.34}$$

where, denoting the first r block rows of $\mathscr{H}_\infty^n$ by $\mathscr{H}_r^n$, the right hand side is equal to $\mathscr{H}_r^n \varepsilon_t^-$. Since $\operatorname{rk} \mathscr{H}_r^n = \operatorname{rk} \mathscr{H}_\infty^n = r$ and $\operatorname{diag}(C_1^n, \dots, C_1^n)$ has full column rank sr, it follows that $\operatorname{rk} \mathscr{H}_r(k) = \operatorname{rk} \mathscr{H}_\infty(k) = r$. Moreover, applying the same argument to the first $j = 1, \dots, r$ block rows of (3.3.34),

$$\begin{pmatrix} \chi_t^n \\ \vdots \\ \chi_{t+j-1|t}^n \end{pmatrix} = \underbrace{\operatorname{diag}(C_1^n, \dots, C_1^n)}_{(jn \times jr)} \mathscr{H}_j(k)\varepsilon_t^-, \tag{3.3.35}$$

for $j = 1, \dots, r$, the rank of $\mathscr{H}_j(k)$ is the same as of $\mathscr{H}_j^n$, implying that up to permutations the Kronecker indices of $\mathscr{H}_\infty(k)$ are the same as the s non-zero Kronecker indices of $\mathscr{H}_\infty^n$; in particular all Kronecker indices of $\mathscr{H}_\infty(k)$ are strictly greater than zero.

Next we are going to show how to obtain a certain *matrix fraction description* (MFD) $k(z) = a^{-1}(z)b(z)$, where $a(z)$ and $b(z)$ are $(s \times s)$ and $(s \times q)$ polynomial matrices respectively. Therefore, let us first define a transfer function $\tilde{k}(z)$ in the *forward shift* z^{-1} as

$$\tilde{k}(z) = z^{-1}k(z^{-1}) = \sum_{j=0}^{\infty} K_j z^{-(j+1)}. \tag{3.3.36}$$

Then clearly $\tilde{k}(z)$ is also rational and $\tilde{k}(z)$ is *strictly proper*, that is $\lim_{|z|\to\infty} \tilde{k}(z) = 0$ (the denominator degrees are strictly higher than the numerator degrees). Note, that $\tilde{k}(z)$ is strictly

proper if and only if $k(z)$ is causal. Consider any MFD of $\tilde{k}(z)$,

$$\tilde{k}(z) = \tilde{a}^{-1}(z)\tilde{b}(z) \tag{3.3.37}$$

of $\tilde{k}(z)$, where $\tilde{a}(z)$ and $\tilde{b}(z)$ are polynomial matrices in z:

$$\tilde{a}(z) = \sum_{j=0}^{\tilde{p}} \tilde{A}_j z^j,\ \tilde{A}_j \in \mathbb{R}^{s\times s},\quad \tilde{b}(z) = \sum_{j=0}^{\tilde{q}} \tilde{B}_j z^j,\ \tilde{B}_j \in \mathbb{R}^{s\times q}.$$

Notice that MFDs always exist, since every rational matrix $\tilde{k}$ can be written as $\tilde{k} = c^{-1}N$, where c is least common denominator polynomial of the entries of $\tilde{k}$ and where N is a polynomial, and hence the existence of (3.3.37) is guaranteed. It then follows, since $z^{-1}k(z^{-1}) = \tilde{a}^{-1}(z)\tilde{b}(z)$, that

$$k(z) = \tilde{a}^{-1}(z^{-1})\left(\tilde{b}(z^{-1})z^{-1}\right) \tag{3.3.38}$$

is an MFD of $k(z)$ in the forward shift z^{-1}. To obtain an MFD of $k(z)$ in the backward shift z, we may set

$$k(z) = \underbrace{(\mathrm{diag}(z^{r_1},\ldots,z^{r_s})\tilde{a}(z^{-1}))}_{a(z)}{}^{-1}\ \underbrace{(\mathrm{diag}(z^{r_1},\ldots,z^{r_s})\tilde{b}(z^{-1})z^{-1})}_{b(z)}, \tag{3.3.39}$$

where r_i denotes the (maximum) degree of the i-th row of $\tilde{a}(z)$[3].The fact that $\tilde{k}(z)$ is strictly proper implies that the degree of the i-th row of $\tilde{b}(z)$ is strictly less than r_i. Hence $a(z)$ and $b(z)$ in 3.3.39 are indeed polynomial matrices.

To construct an MFD $\tilde{a}^{-1}\tilde{b}$ of $\tilde{k}$, we write (using (3.3.36) and (3.3.37))

$$\tilde{a}(z)\left(\sum_{j=0}^{\infty} K_j z^{-(j+1)}\right) = \tilde{b}(z). \tag{3.3.40}$$

From a comparison of coefficients in (3.3.40) corresponding to all negative powers of z we obtain

$$\left(\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_{\tilde{p}}, 0, \ldots\right)\mathscr{H}_\infty(k) = 0_{s\times\infty}, \tag{3.3.41}$$

and from a comparison of coefficients in (3.3.40) corresponding to all non-negative powers of z, we have

$$\left(\tilde{A}_0, \tilde{A}_1, \ldots, \tilde{A}_{\tilde{p}}, 0, \ldots\right)\begin{pmatrix} 0 & 0 & \ldots & \\ K_0 & 0 & \ldots & \\ K_1 & K_0 & 0 & \ldots \\ K_2 & K_1 & K_0 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \left(\tilde{B}_0, \tilde{B}_1, \ldots\right). \tag{3.3.42}$$

Consequently, given $\tilde{k}(z)$, every $\tilde{a}(z)$ satisfying (3.3.41) and such that $\det\tilde{a}(z) \not\equiv 0$ and determining $\tilde{b}(z)$ from (3.3.42) gives an MFD of $\tilde{k}(z)$. Next we will specify one particular MFD of

[3]As will be seen below, it is not coincidental that we use the same notation for the maximum degrees of $\tilde{a}(z)$ and the Kronecker indices.

$\tilde{k}(z)$ making use of the uniquely defined Kronecker indices corresponding to $k(z)$.

Hence, let $\alpha = (r_1, \ldots, r_s)$ be the Kronecker indices of $\mathscr{H}_\infty(k)$ and $h(1,1), \ldots, h(r_1,1), \ldots, h(s,1), \ldots, h(r_s, s)$ the corresponding basis rows. Then by expressing the rows $h(r_i+1, i)$, $i = 1, \ldots, s$ as linear combinations of all preceding basis rows

$$-h(r_i+1, i) = \sum_{j=1}^{s} \sum_{u=1}^{r_{ij}} \tilde{a}_{ij,u-1} h(u, j), \quad i = 1, \ldots, s \tag{3.3.43}$$

where

$$r_{ij} = \begin{cases} \min(r_i + 1, r_j) & \text{for } j < i \\ \min(r_i, r_j) & \text{for } j \geq i, \end{cases} \tag{3.3.44}$$

a special relation of the form (3.3.41) has been chosen, where the unique coefficients $\tilde{a}_{ij,u}$ are the (i,j) element of $\tilde{A}_u$, the (i,i) element of $\tilde{A}_{r_i}$ is $\tilde{a}_{ii,r_i} = 1$ and all other elements are equal to zero. Then $\tilde{p} = \max(r_1, \ldots, r_s)$.

Since for $k(z)$ given, the Kronecker indices are unique, we have uniquely defined $\tilde{a}(z)$. Next, $\tilde{b}(z)$ is obtained from (3.3.42) and is thus unique too. Notice that, (3.3.42) implies $\tilde{B}_j = 0$, $j > \tilde{p} - 1$. Moreover, the corresponding MFD of $\tilde{k}(z)$ has the following properties (see Hannan and Deistler (1988), section 2.5. for details).

(i) $\tilde{a}_{ii}$ are *monic* polynomials[4]. This is an immediate consequence of (3.3.43).

(ii) Denoting the order of a polynomial p by $v(p)$,

$$\begin{aligned} v(\tilde{a}_{ij}) \leq v(\tilde{a}_{ii}) = r_i \quad & j \leq i; \\ v(\tilde{a}_{ij}) < v(\tilde{a}_{ii}) \quad & j > i; \\ v(\tilde{a}_{ji}) < v(\tilde{a}_{ii}) \quad & j \neq i; \\ v(\tilde{b}_{ij}) < v(\tilde{a}_{ii}). & \end{aligned}$$

Again, these are straight forward consequences of (3.3.43). Moreover, it follows that $\tilde{a}(z)$ is both *row reduced* and *column reduced*[5] and hence $r = v(\det \tilde{a}(z))$, which can be shown to imply that

(iii) $(\tilde{a}, \tilde{b})$ is left prime, i.e. the only left divisors of $(\tilde{a}, \tilde{b})$ are unimodular matrices or in other words $\tilde{a}$ and $\tilde{b}$ have no common factors, except for unimodular ones and are thus chosen in the least redundant way.

Consequently, we obtain a unique ARMA realization of $k(z)$ in the forward shift z^{-1} as

$$k(z) = \tilde{a}^{-1}(z^{-1}) \left(\tilde{b}(z^{-1}) z^{-1} \right), \tag{3.3.45}$$

[4] A monic polynomial is a polynomial whose term of highest degree has a coefficient of 1.

[5] Let $a_i = \sum_{u=0}^{v(a_i)} a_i(u) z^u$ denote the i-th row of a $s \times s$ matrix a. Then the matrix $[a]_r = (a_1(v(a_1))', \ldots, a_s(v(a_s))')'$ is called *row end matrix* of a and a is called row reduced, if $\operatorname{rk} [a]_r = s$. Column reducedness is defined analogously.

where $(\tilde{a}(z), \tilde{b}(z))$ has the properties discussed above. This realization will be called the *(singular) echelon ARMA realization* of $k(z)$, for its obvious similarity to echelon ARMA realizations and where the term singular refers to the singular spectral density of the corresponding stationary process and the singular one-step prediction errors $K_0\varepsilon_t$. From (3.3.39) we then obtain a unique ARMA realization in the backward shift z that has similar properties and will be called the *reversed (singular) echelon ARMA realization* of $k(z)$.

The following simple example will illustrate the construction of a canonical ARMA realization.

Example 3.3.3. Consider the following non-generic minimal state space system:

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; C^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ & \vdots & \end{pmatrix}, \tag{3.3.46}$$

where $q = 1$, $s = \operatorname{rk} C^n = 2$ and $r = 3$, since

$$\mathscr{C}_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{3.3.47}$$

is of full (row) rank and

$$\mathscr{O}_3 = \left(\begin{array}{cccc|cccc|cccc} 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & 1 & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots \end{array} \right)' \tag{3.3.48}$$

is of full column rank and from $\mathscr{O}_3$ we obtain that the non-zero Kronecker indices are $r_1 = 1, r_2 = 2$.

The transfer function $k(z)$ corresponding to the static factor is

$$k(z) = (I_2, 0)(I_3 - Az)^{-1}B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} z + \begin{pmatrix} 1 \\ 0 \end{pmatrix} z^2 = \begin{pmatrix} z^2 \\ 1 \end{pmatrix}, \tag{3.3.49}$$

with Hankel matrix

$$\mathscr{H}_3(k) = \left(\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right), \tag{3.3.50}$$

of rank 3 and where the first basis rows are $h(1,1), h(2,1), h(1,2)$ implying that the corresponding Kronecker indices are $r_1 = 2, r_2 = 1$.

The corresponding strictly proper transfer function (3.3.36) is then

$$\tilde{k}(z) = z^{-1}k(z^{-1}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} z^{-1} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} z^{-2} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} z^{-3} = \begin{pmatrix} z^{-3} \\ z^{-1} \end{pmatrix}. \tag{3.3.51}$$

To obtain an MFD $\tilde{a}^{-1}\tilde{b}$ of (3.3.51) we may use (3.3.43) to determine $\tilde{a}$, hence

$$-h(3,1) = \tilde{a}_{11,0}h(1,1) + \tilde{a}_{12,0}h(1,2) + \tilde{a}_{11,1}h(2,1) \tag{3.3.52}$$

$$-h(2,2) = \tilde{a}_{21,0}h(1,1) + \tilde{a}_{22,0}h(1,2) + \tilde{a}_{21,1}h(2,1), \tag{3.3.53}$$

from which we get that $\tilde{a}_{11,0} = \tilde{a}_{11,1} = \tilde{a}_{21,0} = \tilde{a}_{22,0} = \tilde{a}_{21,1} = 0$, $\tilde{a}_{12,0} = -1$ and $\tilde{a}_{11,2} = \tilde{a}_{22,1} = 1$, hence

$$\tilde{A}_0 = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix},\ \tilde{A}_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ \tilde{A}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.3.54}$$

Determining $\tilde{b}$ from (3.3.42) yields

$$\tilde{B}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \tilde{B}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.3.55}$$

Finally from (3.3.39) we obtain

$$a(z) = \begin{pmatrix} z^2 & 0 \\ 0 & z \end{pmatrix} \begin{pmatrix} z^{-2} & -1 \\ 0 & z^{-1} \end{pmatrix} = \begin{pmatrix} 1 & -z^2 \\ 0 & 1 \end{pmatrix} \text{ and} \tag{3.3.56}$$

$$b(z) = \begin{pmatrix} z^2 & 0 \\ 0 & z \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} z^{-1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.3.57}$$

■

The next theorem summarizes the properties of the reversed (singular) echelon realization.

Theorem 3.3.4. *Let $k(z)$ be a rational $(s \times q)$ transfer function $(s \geq q)$, analytic for $|z| \leq 1$ and with $\operatorname{rk} k(z) = q$ for $|z| \leq 1$. Let the corresponding Kronecker indices be $\alpha = (r_1, \ldots, r_s)$ with $\sum_i r_i = r \geq s$ and $r_i \geq 1, i = 1 \ldots, s$. Then a unique ARMA realization in the backward shift z of the form*

$$k(z) = a^{-1}(z)b(z) \tag{3.3.58}$$

is defined, where

$$a(z) = A_0 - \sum_{j=1}^{P} A_j z^j,\ A_j \in \mathbb{R}^{s\times s},\ b(z) = \sum_{j=0}^{Q} B_j z^j,\ B_j \in \mathbb{R}^{s\times q},$$

are determined by (3.3.43), (3.3.42) and (3.3.39). This MFD has the following properties

(i)

$$a_{ii}(z) = 1 - \sum_{u=1}^{r_i} a_{ii,u} z^u,\ i = 1, \ldots, s$$

$$a_{ij}(z) = -\sum_{u=r_i-r_{ij}+1}^{r_i} a_{ij,u} z^u,\ i, j = 1, \ldots, s,\ i \neq j,$$

where r_{ij} has been defined in (3.3.44). (A_0 is lower triangular and all its diagonal elements are equal to 1.)

(ii)

$$b_{ij}(z) = \sum_{u=0}^{r_i-1} b_{ij,u} z^u, \ i = 1, \ldots, s, \ j = 1, \ldots, q$$

(iii) (a, b) *is left prime.*

Proof. Properties (i) - (ii) are immediate consequences of (3.3.43), (3.3.42) and (3.3.39) and the properties of $(\tilde{a}(z), \tilde{b}(z))$ discussed above. Property (iii) follows from the fact that $(\tilde{a}, \tilde{b})$ is left prime and A_0 has full rank s (see Hannan and Deistler (1988), proof of Lemma 2.4.2). ■

Hence, in general the degree of the i-th row of $a(z)$ is equal to r_i: $v(a_{ij}) \leq r_i$, $i, j = 1, \ldots, s$ and the degree of the i-th row of $b(z)$ is equal to $r_i - 1$: $v(b_{ij}) = r_i - 1$, $i = 1, \ldots, s$, $j = 1, \ldots, q$. Then, in general for the ARMA orders we have $P = \max(r_1, \ldots, r_s)$ and $Q = \max(r_1, \ldots, r_s) - 1$. Indeed some of the free parameters may be zero and hence either P or Q may actually be smaller than the above values (see the example above). The number of free parameters in the (reversed) singular echelon realization, d_α say, is determined by the Kronecker indices α and equals

$$d_\alpha = r(1 + q) + \sum_{i \neq j} r_{ij}$$

as can be seen from Theorem 3.3.4 (i)-(ii). Further, also the positioning of the free parameters is determined by the Kronecker indices. The following example will illustrate the role of the Kronecker indices.

Example 3.3.4. For dimensions $q = 1, s = 2$ and Kronecker indices $\alpha = (3, 1)$ given, we get $r = 4$, $r_{12} = 1, r_{21} = 2$ and thus there are 11 free parameters ($= d_\alpha$) to appear in the singular echelon ARMA form (3.3.59) and the reversed singular echelon ARMA form (3.3.60), respectively, and the ARMA orders are $P = 3, Q = 2$:

$$\begin{aligned}
\tilde{a}(z) &= \begin{pmatrix} \tilde{a}_{11,0} & \tilde{a}_{12,0} \\ \tilde{a}_{21,0} & \tilde{a}_{22,0} \end{pmatrix} + \begin{pmatrix} \tilde{a}_{11,1} & 0 \\ \tilde{a}_{21,1} & 1 \end{pmatrix} z + \begin{pmatrix} \tilde{a}_{11,2} & 0 \\ 0 & 0 \end{pmatrix} z^2 + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} z^3 \\
\tilde{b}(z) &= \begin{pmatrix} \tilde{b}_{1,0} \\ \tilde{b}_{2,0} \end{pmatrix} + \begin{pmatrix} \tilde{b}_{1,1} \\ 0 \end{pmatrix} z + \begin{pmatrix} \tilde{b}_{1,2} \\ 0 \end{pmatrix} z^2 \qquad (3.3.59) \\
a(z) &= \begin{pmatrix} z^3 & 0 \\ 0 & z \end{pmatrix} \tilde{a}(z^{-1}) \\
&= \begin{pmatrix} 1 & 0 \\ \tilde{a}_{21,1} & 1 \end{pmatrix} + \begin{pmatrix} \tilde{a}_{11,2} & 0 \\ \tilde{a}_{21,0} & \tilde{a}_{22,0} \end{pmatrix} z + \begin{pmatrix} \tilde{a}_{11,1} & 0 \\ 0 & 0 \end{pmatrix} z^2 + \begin{pmatrix} \tilde{a}_{11,0} & \tilde{a}_{12,0} \\ 0 & 0 \end{pmatrix} z^3 \\
b(z) &= \begin{pmatrix} z^3 & 0 \\ 0 & z \end{pmatrix} \tilde{b}(z^{-1}) z^{-1} = \begin{pmatrix} \tilde{b}_{1,2} \\ \tilde{b}_{2,0} \end{pmatrix} + \begin{pmatrix} \tilde{b}_{1,1} \\ 0 \end{pmatrix} z + \begin{pmatrix} \tilde{b}_{1,0} \\ 0 \end{pmatrix} z^2, \qquad (3.3.60)
\end{aligned}$$

whereas the corresponding unrestricted standard ARMA(3,1)-realization (see below for a definition) involves 18 free parameters.

For $\alpha = (2,2)$ leaving q and s unchanged, we get that $r = 4$ as before, and since $r_{12} = r_{21} = 2$ there are 12 free parameters to appear in the singular echelon ARMA form (3.3.61) and the reversed singular echelon ARMA form (3.3.62), respectively, and for the ARMA orders we have $P = 2, Q = 1$:

$$\begin{aligned}
\tilde{a}(z) &= \begin{pmatrix} \tilde{a}_{11,0} & \tilde{a}_{12,0} \\ \tilde{a}_{21,0} & \tilde{a}_{22,0} \end{pmatrix} + \begin{pmatrix} \tilde{a}_{11,1} & \tilde{a}_{12,1} \\ \tilde{a}_{21,1} & \tilde{a}_{22,1} \end{pmatrix} z + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} z^2 \\
\tilde{b}(z) &= \begin{pmatrix} \tilde{b}_{1,0} \\ \tilde{b}_{2,0} \end{pmatrix} + \begin{pmatrix} \tilde{b}_{1,1} \\ \tilde{b}_{2,1} \end{pmatrix} z \qquad (3.3.61) \\
a(z) &= \begin{pmatrix} z^2 & 0 \\ 0 & z^2 \end{pmatrix} \tilde{a}(z^{-1}) \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \tilde{a}_{11,1} & \tilde{a}_{12,1} \\ \tilde{a}_{21,1} & \tilde{a}_{22,1} \end{pmatrix} z + \begin{pmatrix} \tilde{a}_{11,0} & \tilde{a}_{12,0} \\ \tilde{a}_{21,0} & \tilde{a}_{22,0} \end{pmatrix} z^2 \\
b(z) &= \begin{pmatrix} z^2 & 0 \\ 0 & z^2 \end{pmatrix} \tilde{b}(z^{-1}) z^{-1} = \begin{pmatrix} \tilde{b}_{1,1} \\ \tilde{b}_{2,1} \end{pmatrix} + \begin{pmatrix} \tilde{b}_{1,0} \\ \tilde{b}_{2,0} \end{pmatrix} z, \qquad (3.3.62)
\end{aligned}$$

thus here the unrestricted standard ARMA(2,1)-realization and the reversed singular echelon realization coincide (which is a consequence of all Kronecker indices being equal). ■

In an analogous way to standard echelon realizations, it can be shown that the reversed singular echelon form is identifiable, i.e. if an ARMA realization is in reversed singular echelon form the realization is unique within the class of reversed singular echelon forms (which follows mainly from the fact that the singular echelon form makes use of the uniquely defined Kronecker indices). Notice that if actually $s = q$, that is if k is square, then the reversed singular echelon form and the reversed echelon form coincide (although in the standard ARMA context Kronecker indices as well as Hankel matrices are usually defined in a different way).

Remark 3.3.2. If $s = r$, i.e. if the minimum state dimension equals the minimum static factor dimension, all Kronecker indices are equal to 1, implying that the canonical MFD $a^{-1}b$ from (3.3.43), (3.3.42) and (3.3.39) reduces to an AR(1) realization implying further that the singular reversed echelon ARMA realization coincides with the transition equation of a minimum state space realization.

It is immediate, that any MFD $k(z) = a^{-1}(z)b(z)$, where (a, b) is left prime, and thus the canonical MFD given in (3.3.43), (3.3.42) and (3.3.39) has to satisfy

- The stability condition: $\det a(z) \neq 0$ for $|z| \leq 1$, since otherwise $k(z)$ would not be analytic on $|z| \leq 1$.
- The miniphase (or invertibility) condition: $\operatorname{rk} b(z) = q$ for $|z| \leq 1$, since otherwise $k(z)$ would not be of rank q on $|z| \leq 1$.

Left multiplying $(a(z), b(z))$ from Theorem 3.3.4 by A_0^{-1}, we get an ARMA realization of $k(z)$ that is in *standard form*, i.e. the new AR coefficient at lag 0, $\bar{A}_0$ say, equals I_s. Further, if the realization is in standard form, the MA coefficient at lag 0, $\bar{B}_0$ say, equals K_0. Clearly, this realization preserves the left-primeness and the ARMA orders (P, Q) of the echelon parametrization. However requiring only, that $\bar{A}_0 = I_s$ is in general not enough to describe a unique set of parameters. An important special case in which the unrestricted standard form is equal to the reversed echelon form and is thus identifiable, occurs if all Kronecker indices are equal, as can be easily seen from 3.3.4 (i)-(ii). In any case, for some purposes (e.g. for prediction) it is more convenient to use the standard form.

3.3.5 The quasi-static representation of the GDFM

Summarizing, we have shown that under Assumptions 1.1 and 3.1 there always exists a (quasi)-static representation of (2.1.1) of the form

$$y_t^n = \bar{\Lambda}^n f_t + u_t^n, \tag{3.3.63}$$

where $\bar{\Lambda}^n$ is an $(n \times s)$ dimensional static factor loading matrix[6] and f_t is an s-dimensional static factor ($s \geq q$), where for sufficiently large n, $\operatorname{rk} \bar{\Lambda}^n = s$ holds and where f_t solves an ARMA(P, Q) system,

$$a(z) f_t = b(z) \varepsilon_t, \tag{3.3.64}$$

with the following properties:

- $a(z) = I_s - A_1 z - \ldots - A_P z^P$ and $b(z) = B_0 + B_z + \ldots + B_Q z^Q$ are $(s \times s)$ and $(s \times q)$ dimensional polynomial matrices and (a, b) is left prime, and where
- The stability condition ($\det a(z) \neq 0$ for $|z| \leq 1$) and the miniphase condition ($\operatorname{rk} b(z) = q$ for $|z| \leq 1$) hold.
- The q-dimensional white noise (ε_t) are the linear innovations of (χ_t^n) and f_t, and $\Gamma_\varepsilon = I_q$, and may be seen as the q-dimensional dynamic factor process corresponding to the Wold representation of χ_t^n (fundamental shocks).
- Denoting the transfer function corresponding to the Wold representation of f_t by $k = a^{-1} b$ and its Kronecker indices by $\alpha = (r_1, \ldots, r_s)$, then the order P of $a(z)$ is less or equal to $\max(r_1, \ldots, r_s)$ and the order Q of $b(z)$ is strictly less than $\max(r_1, \ldots, r_s)$.

Concerning identifiability resuming the main results, we have that

- Under the assumption that $\Gamma_f = I_s$ and for given Γ_χ^n, $\bar{\Lambda}^n$ and f_t in (3.3.63) are identifiable up to static rotations. Hence identification consists in choosing a specific rotation (see below).

[6] To avoid confusion, notice, that in the context of state space realizations $\bar{\Lambda}^n$ was denoted by C_1^n. The renaming shall emphasize that it is a static factor loading matrix.

- For given (f_t), then $f_t = k(z)\varepsilon_t$ and $k(z)$ and ε_t are identifiable up to static rotations. Again identification consists in choosing a specific rotation. One possibility to do this is to impose $q(q-1)/2$ zero restrictions on K_0 since orthonormality of ε_t already implicates $q(q+1)/2$ restrictions, for instance by requiring the upper triangular part of K_0 to be 0. Then $k(z)$ is identifiable up to sign changes of its columns.
- For given $k(z)$, the ARMA realization (3.3.64) in standard form is in general not identifiable. One possibility to reach identifiability is the derivation via the singular echelon form as described above, but of course other identification conditions are also conceivable (see for instance Hannan and Deistler (1988) for a discussion of structural identifiability).

3.3.6 AR-representation of the static factor and zeroless transfer function case

In the sequel we are going to analyze the autoregressive representation of the static factors with special emphasis on the case in which the spectral factor w^n in (3.3.2) is zeroless, i.e. $\operatorname{rk} w^n = q$ for all $z \in \mathbb{C}$ (see also Anderson and Deistler (2008)). Recall that a regular ARMA system (in the sense that b is square) that satisfies the miniphase assumption $\det b(z) \neq 0$ for $|z| \leq 1$ can be expressed as an AR system of in general infinite order. Quite obviously, this AR system is finite if and only if b is unimodular. In the singular ARMA case the requirement for the existence of a finite AR realization then translates to the condition that b is zeroless. It is easy to see that b is zeroless if w^n is zeroless, which is, as we will see, a generic property in the GDFM framework, where $n >> q$. Therefore consider the intuitive example of a polynomial w^n with $q = 1$. Then $\operatorname{rk} w^n(z_0) = 0$ for some $z_0 \in \mathbb{C}$ if and only if z_0 is a common root of all n polynomial entries of w^n. A zero of w^n thus requires the coefficients to satisfy $n-1$ constraints − being zeroless is therefore a generic property.

Let us however start with the general case, where we only require $\operatorname{rk} w^n(z) = q$ for $|z| \leq 1$, which as stated above implies that $k(z)$ in (3.3.32) and thus also $b(z)$ in (3.3.64) have full column rank q for $|z| \leq 1$ (miniphase assumption). Recalling that $b(z)$ can be written as

$$b = ulv, \tag{3.3.65}$$

where the $(s \times s)$ and $(q \times q)$ polynomial matrices u and v are unimodular, and where the $(s \times q)$ polynomial matrix l, i.e. the *Smith form* of b, is diagonal and displays the zeros of b (see e.g. Hannan and Deistler (1988), Lemma 2.2.3) - hence l does not have any zeros for $|z| \leq 1$. Thus defining b^- as

$$b^- = v^{-1}(l'l)^{-1}l'u^{-1}, \tag{3.3.66}$$

wee see, that, because l does not have any zeros for $|z| \leq 1$, by its diagonality the same holds for $l'l$, b^- is analytic in an open disc containing the unit disc and can be extended in a convergent power series in the same region. Thus, b^- is a one-sided (causal) left inverse of b, implying that (ε_t) can be expressed as a causal linear transformation of (f_t),

$$\varepsilon_t = b^-(z)a(z)f_t. \tag{3.3.67}$$

Writing equation (3.3.64) as

$$a(z)f_t = (b(z) - b(0))\varepsilon_t + b(0)\varepsilon_t \tag{3.3.68}$$

and substituting (3.3.67), we obtain that f_t solves the possibly infinite AR system

$$\underbrace{\left[I_s - (b(z) - b(0))b^-(z)\right] a(z)}_{\phi(z)} f_t = b(0)\varepsilon_t \tag{3.3.69}$$

where $\Phi_0 = \phi(0) = I_s$, $\phi(z) = I_s - \sum_{k=1}^{\infty} \Phi_k z^k$ with $\sum_{k=0}^{\infty} \|\Phi_k\|^2 < \infty$ and $b(0) = B_0 = K_0$.

Getting back to the zeroless case, it follows immediately that if $w^n(z)$ is zeroless, then $k(z)$ in (3.3.32) and thus also $b(z)$ in (3.3.64) are zeroless. Hence the Smith-form l of b in (3.3.65) is equal to $l = (I_q, 0)'$ and b^- reduces to

$$b^- = v^{-1}l'u^{-1}, \tag{3.3.70}$$

which is polynomial, implying that $\phi(z)$ in (3.3.69) is polynomial. Consequently, in the zeroless case the quasi-static factor f_t always allows of a finite-order AR representation.

In any case, denoting the Hilbert space spanned by all past and present values $f_{i\tau}$, $i = 1, \ldots, s, \tau \leq t$ by $\mathbb{H}_f(t^-)$ and in an analogous way the Hilbert space spanned by all past and present values $\varepsilon_{i\tau}$, $i = 1, \ldots, q, \tau \leq t$ by $\mathbb{H}_\varepsilon(t^-)$, then

$$\mathbb{H}_f(t^-) = \mathbb{H}_\varepsilon(t^-).$$

Writing the AR representation of f_t, (3.3.69) as

$$f_t = \sum_{k=1}^{\infty} \Phi_k f_{t-k} + B_0\varepsilon_t,$$

it follows that the two terms on the right hand side are orthogonal. In fact, the first term on the right is the projection of f_t onto $\mathbb{H}_f((t-1)^-)$ and is thus the best linear least squares one-step ahead predictor of f_t given information up to time $t-1$ and the second term is the corresponding prediction error (see Chapter 4). By the projection theorem we know that this projection exists and is unique. On the other hand we get from equation (3.3.65) that, if $q < s$, the $(s-q)$ bottom rows of $u^{-1}(z)b(z)$ only consist of zeros, and since u^{-1} is polynomial and of full rank, this implies that there exist s-dimensional polynomial vectors $p(z)$, $p(z) \neq 0$, such that $p'(z)b(z) = 0$ and hence $p'(z)a(z)f_t = 0$. Consequently, there exist a vector a and some integer R, such that

$$a' \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-R} \end{pmatrix} = 0.$$

Hence, if $q < s$, the random variables $f_{i\tau}$, $i = 1, \ldots, s$, $t - R - 1 \leq \tau \leq t - 1$ are linearly dependent and thus $f_{i\tau}$, $i = 1, \ldots, s$, $\tau \leq t - 1$ form no basis of $\mathbb{H}_f((t-1)^-)$, such that in

general the coefficients Φ_k, $k = 1, 2 \ldots$, in (3.3.69) are not unique.

In the zeroless case and if $q < s$, the finite AR representation is in general not unique in two respects: first, the AR order is not unique, since left multiplying (ϕ, B_0) by any polynomial $p(z)$, satisfying $p(0) = I_s$ and $p(z)B_0 = B_0$ yields another standard AR representation with possibly different and arbitrarily high order, and second even for fixed AR order, the coefficients Φ_j are in general not uniquely identifiable by the arguments pointed out above. The following lemma provides a criterion for uniqueness of the AR coefficients in the zeroless case.

Lemma 3.3.1. *Let $\phi(z) = I_s - \Phi_1 z - \ldots - \Phi_p z^p$ and (ϕ, B_0) be left prime. For p fixed, the coefficients Φ_j, $j = 1, \ldots, p$ are unique if and only if $\mathrm{rk}(\Phi_p, B_0) = s$. In this case, p is minimal, i.e. there does not exist any observationally equivalent AR representation $(\tilde{\phi}, \tilde{B}_0)$ with lower order than p.*

Proof. Since (ϕ, B_0) is left prime, all observationally equivalent AR representations $(\tilde{\phi}, \tilde{B}_0)$ with $\tilde{\phi}(0) = I_s$ have to satisfy $(\tilde{\phi}, \tilde{B}_0 = B_0) = u(\phi, B_0)$, where the $(s \times s)$ polynomial matrix u, is such that $u(0) = I_s$ and $u(z)B_0 = B_0$. Hence $u = I_s + \tilde{u}$, where $\tilde{u}(0) = 0$ and $\tilde{u}(z)B_0 = 0$. Let u be of order m, then for fixed p a comparison of coefficients yields $\tilde{U}_m \Phi_p = 0$, thus combining the zero requirements $\tilde{U}_m$ has to solve $\tilde{U}_m(\Phi_p, B_0) = 0$. If $\mathrm{rk}(\Phi_p, B_0) = s$, it follows that $\tilde{U}_m = 0$, implying that $u(z) = I_s$.

Conversely, if $\mathrm{rk}(\Phi_p, B_0) < s$, let $\tilde{u}(z) = \tilde{U}_1 z$. From a comparison of coefficients we get that $\tilde{U}_1 \Phi_p = 0$ and there exist non trivial $\tilde{U}_1$ satisfying $\tilde{U}_1(\Phi_p, B_0) = 0$.

Finally, the proof that p is minimal if $\mathrm{rk}(\Phi_p, B_0) = s$ holds, follows from exactly the same arguments as used in the first part of this proof. ■

An important case in which the conditions of Lemma 3.3.1 are satisfied occurs if $r = s$, i.e. if the minimum state dimension coincides with the minimum static factor dimension. In this case, the non-zero Kronecker indices corresponding to w^n and hence all Kronecker indices corresponding to k are equal to 1, and the reversed echelon ARMA realization of k, $k = a^{-1}b$ reduces to an AR(1) realization in standard form,

$$f_t = A_1 f_{t-1} + B_0 \varepsilon_t, \tag{3.3.71}$$

where (a, B_0) is left prime. Furthermore, (3.3.71) coincides with the transition equation of a minimal state space realization of w^n, thus in particular the pair (A_1, B_0) is controllable, i.e. $\mathrm{rk}\,\mathscr{C}_r = \mathrm{rk}(B_0, A_1 B_0, \ldots, A_1^{r-1} B_0) = r$. Now, let us assume that $\mathrm{rk}(A_1, B_0) < r$, then there exists a regular $(r \times r)$ matrix T, such that the bottom rows of $T(A_1, B_0)$ only contain zeros, implying that the bottom rows of $T\mathscr{C}_r$ only contain zeros too, which is in contradiction to controllability.

Chapter 4

Prediction in the GDFM

Prediction is one of the the main objectives in dealing with factor models (see for instance Deistler and Hamann (2005), Stock and Watson (2005) and Forni *et al.* (2005a)). If the objective is the prediction of a high-dimensional time series (e.g. in forecasting the asset returns of a large portfolio), then as stated in Tiao (2001) multivariate modeling may improve the forecast accuracy when there is information on one component contained in the historical data of another one. Then a factor model approach may be advantageous compared to traditional multivariate time series models, since it allows of a sparse parametrization even if the cross-sectional dimension is large relative to sample size. If the objective is the prediction of only a single or a few time series, relevant information on these series may be comprised in a huge number of potential "inputs" (e.g. in macroeconomic forecasting, see Stock and Watson (1998)). Then again, modeling inputs and outputs jointly by using a factor model approach might be advantageous compared to traditional modeling, in that it (partly) avoids the input selection problem without disregarding potentially useful information, while at the same time the parametrization remains quite sparse. Thus here we will be concerned with the prediction of future values of observations y_t^n that allow for a dynamic factor model representation (2.1.1) and satisfy Assumptions 1.1, 2.1 and 3.1.

In general, the problem of *prediction* is to find an "optimal" approximation of the future value y_{t+h}^n, $h > 0$, by a (measurable) function of present and past values y_τ^n, $\tau \leq t$. Here we will only consider *linear* approximation functions and optimality of the prediction will be understood in *least-squares* sense. Then it follows by the *projection theorem*, that the best linear predictor of y_{t+h}^n, given information up to t, $y_{t+h|t}^n$ say, is the projection of y_{t+h}^n onto the Hilbert space spanned by present and past values of y_t^n, i.e.

$$y_{t+h|t}^n = \text{proj}(y_{t+h}^n | \overline{\text{span}}(y_{i\tau} | i = 1, \dots, n,\ \tau \leq t)). \tag{4.0.1}$$

Notice that in general (i.e. when the approximation function may be any measurable function) the best least squares approximation of y_{t+h}^n is the conditional expectation,

$$\tilde{y}_{t+h|t}^n = \mathbb{E}(y_{t+h}^n | y_\tau^n, \tau \leq t).$$

In general the determination of the conditional expectation is of rather complex nature and requires further assumptions on the underlying distributions, whereas the projections onto $\mathbb{H}_y(t^-) = \overline{\text{span}}(y_{i\tau}|i = 1, \ldots, n,\ \tau \leq t)$ only depend on the second moments of the variables, which is the main reason for restricting the approximation functions to be linear. See for instance Brockwell and Davis (1987) for a more detailed description of the prediction of stationary processes.

Since by Assumption 1.1 b) the latent variable and the noise are mutually orthogonal and since they are both contained in $\mathbb{H}(y)$ (see Remark 3.1.5), we may split the prediction of y_{t+h}^n into two separate prediction problems, i.e. the prediction of χ_{t+h}^n and u_{t+h}^n, respectively, and obtain the forecast of y_{t+h}^n as the sum of the two separate predictions (see Forni *et al.* (2005a)). Since by Assumption 2.1 the noise process (u_t^n) is only weakly correlated, it is well justified to predict the components of u_{t+h}^n by means of univariate or low-dimensional time series models, such as e.g. univariate or low dimensional AR models. Thus here we will focus on the prediction of the latent variable.

From representation (3.3.63) (i.e. $y_t^n = \bar{\Lambda}^n f_t + u_t^n$) and since $\operatorname{rk} \bar{\Lambda}^n = s$ for n sufficiently large implies equality of the Hilbert spaces $\mathbb{H}_\chi^n(t^-) = \overline{\text{span}}(\chi_{i\tau}|i = 1, \ldots, n,\ \tau \leq t)$ and $\mathbb{H}_f(t^-) = \overline{\text{span}}(f_{i\tau}|i = 1, \ldots, s,\ \tau \leq t)$, respectively, we obtain

$$\chi_{t+h|t}^n = \bar{\Lambda}^n f_{t+h|t}, \tag{4.0.2}$$

implying further that we may even focus on the prediction of the static factor f_t, which will be performed using its ARMA representation (3.3.64).

It is emphasized that throughout this chapter f_t, $t \in \mathbb{Z}$, is treated as if it was observed, and it is assumed that its ARMA representation is fully known. Of course, in practice f_t and the corresponding ARMA model have to be estimated. However the framework and methods presented below serve as an important step towards prediction commencing from data.

4.1 Prediction from an infinite past in the singular ARMA case

We commence from the standard ARMA representation of f_t,

$$a(z)f_t = b(z)\varepsilon_t,$$

where $a(z) = I_s - A_1 z - \ldots - A_P z^P$ and $b(z) = B_0 + B_1 z + \ldots + B_Q z^Q$ are $(s \times s)$ and $(s \times q)$ dimensional polynomial matrices $(s \geq q)$ and where $\det a(z) \neq 0$ for $|z| \leq 1$ and $\operatorname{rk} b(z) = q$ for $|z| \leq 1$.

Then, as shown in Section 3.3.6 the stability and miniphase assumptions imply that the Hilbert space spanned by all past and present values $f_{i\tau}$, $i = 1, \ldots, s, \tau \leq t$, denoted by $\mathbb{H}_f(t^-)$, and the

Hilbert space spanned by all past and present values $\varepsilon_{i\tau}$, $i = 1, \ldots, q, \tau \leq t$, denoted by $\mathbb{H}_\varepsilon(t^-)$, are equal, and that further f_t may be expressed by a possibly infinite autoregression, i.e.

$$\phi(z) f_t = B_0 \varepsilon_t, \tag{4.1.1}$$

where $\phi(z) = I_s - \sum_{k=1}^{\infty} \Phi_k z^k$ and the coefficients Φ_k are determined using $a(z)$ and $b(z)$ as in (3.3.69). Notice, that (3.3.69) is only one particular choice for defining an autoregressive relation of f_t. However, this formulation has the convenient property, that $\phi(z)$ is polynomial in the generic case of a zeroless transfer function k (see Section 3.3.6). We thus have

$$f_{t+1} = \sum_{k=1}^{\infty} \Phi_k f_{t+1-k} + B_0 \varepsilon_{t+1} \tag{4.1.2}$$

and the best linear least-squares one-step ahead predictor for f_{t+1}, i.e. $f_{t+1|t} = \mathrm{proj}(f_{t+1}|\mathbb{H}_f(t^-))$ by what was said in the introduction of this chapter, can be expressed as

$$f_{t+1|t} = \sum_{k=1}^{\infty} \Phi_k f_{t+1-k}, \tag{4.1.3}$$

since $f_{t+1|t} \in \mathbb{H}_f(t^-)$ and since the prediction error $f_{t+1} - f_{t+1|t} = B_0 \varepsilon_{t+1}$ is orthogonal to $\mathbb{H}_f(t^-)$ because of the equivalence of the latter to $\mathbb{H}_\varepsilon(t^-)$.

The two-step ahead predictor is given by

$$f_{t+2|t} = \Phi_1 f_{t+1|t} + \sum_{k=2}^{\infty} \Phi_k f_{t+2-k}, \tag{4.1.4}$$

since $f_{t+2|t} \in \mathbb{H}_f(t^-)$ and since the prediction error $f_{t+2} - f_{t+2|t} = B_0 \varepsilon_{t+2} + \Phi_1 B_0 \varepsilon_{t+2}$ is orthogonal to $\mathbb{H}_f(t^-)$ by the same argument as above. In a completely analogous way, any h-step, $h \geq 1$ predictor may be determined.

Eventually, two remarks seem to be indicated. First, as analyzed in Section 3.3.6, the AR coefficients Φ_k, $k = 1, 2, \ldots$ are in general not unique. However, the predictor $f_{t+h|t}$, $h \geq 1$ is unique since it is a projection. Second, also as analyzed in Section 3.3.6, in the zeroless transfer function case the AR-representation of (f_t) is finite, implying of course that the predictor $f_{t+h|t}$, $h \geq 1$ only depends on a finite number of past and present f_t's. In the general non-zeroless case, since in practice f_t is not available for $t < 1$, the last expressions are of rather theoretical interest but cannot be computed. Of course a feasible solution is to truncate the infinite sums at $k = t + h$, which yields an approximation of $f_{t+h|t}$ that quite obviously converges as $t \to \infty$, but this is still not the best linear prediction from a finite past. With the latter we will be concerned in the next section.

4.2 Prediction from a finite past in the singular ARMA case

In this section we will develop a recursive algorithm for the one-step predictors for f_{t+1} given a finite number of past and present realizations, f_τ, $\tau = 1, \ldots, t$,

$$\hat{f}_{t+1} = \mathrm{proj}(f_{t+1}|\overline{\mathrm{span}}(f_{i\tau}|i = 1, \ldots, s, \tau = 1, \ldots, t)),\ t \geq 1. \tag{4.2.1}$$

Notice, that the hat in $\hat{f}_{t+1}$ is used to distinguish this predictor from the previously defined quantity $f_{t+1|t}$, which depends on the infinite past, f_τ, $\tau \leq t$. Furthermore, as far as one-step predictors are concerned, for sake of simplicity we omit the term $|t$ in the subscript of $\hat{f}_{t+1}$. Then, as $t \to \infty$, quite evidently

$$\hat{f}_{t+1} \to f_{t+1|t}, \tag{4.2.2}$$

and thus $\hat{f}_{t+1}$ is an approximation for the (theoretical) quantity $f_{t+1|t}$. Clearly, in the zeroless case, where f_t allows of a finite AR representation of order $\bar{P}$ say, for $t \geq \bar{P}$, $\hat{f}_{t+1} = f_{t+1|t} = \sum_{k=1}^{\bar{P}} \Phi_k f_{t-k+1}$ and there is no need for a recursive approximation. Hence the subsequent algorithms have been developed rather for the non-zeroless case, but will hold in general.

The reason for considering recursive algorithms is that for large t the direct determination of $\hat{f}_{t+1}$ from (4.2.1) requires the solution of a large system of linear equations, which can be avoided using the recursive algorithm going to be discussed below. Furthermore the latter has the property that the predictor based on $t+1$ realizations uses the predictors based on t realizations, implying that if the sample size is increased the procedure does not have to be repeated, but may simply be continued. Hence here the term "recursive" could be replaced by "real-time" or "on-line".

First we will present a generalization of the *multivariate innovations algorithm* (see Brockwell and Davis (1987), Proposition 11.4.2), in that it allows for singular processes (z_t) (in the sense that the variance matrix of the linear innovations, $(z_{t+1} - \hat{z}_{t+1})$, may be singular), and second we will show how to apply this algorithm to a possibly singular ARMA process (f_t).

4.2.1 The multivariate innovations algorithm for singular innovations

Let us consider an s-dimensional process (z_t), $z_t \in L_2(\Omega, \mathcal{F}, P)$ for $t \in \mathbb{Z}$, where $(\Omega, \mathcal{F}, P)$ is the underlying probability space and let us define the Hilbert space spanned by present and past realizations of (z_t), starting with $t = 1$, as $\mathbb{K}_z(t) = \overline{\text{span}}(z_{j\tau} | j = 1, \ldots, s, \tau = 1, \ldots, t)$. Then

$$\hat{z}_{t+1} = \text{proj}(z_{t+1} | \mathbb{K}_z(t)) \ \in \mathbb{K}_z(t)$$

and thus it is clear that,

$$(z_{t+1} - \hat{z}_{t+1}) \in \mathbb{K}_z(t+1) \setminus \mathbb{K}_z(t),$$

implying that

$$\mathbb{K}_z(t) = \overline{\text{span}}(z_{j\tau} - \hat{z}_{j\tau} | j = 1, \ldots, s, \tau = 1, \ldots, t),$$

such that we may express $\hat{z}_{t+1}$ as a linear combination of its linear innovations,

$$\hat{z}_{t+1} = \sum_{j=1}^{t} \Theta_{tj}(z_{t+1-j} - \hat{z}_{t+1-j}), \tag{4.2.3}$$

where $\Theta_{tj} \in \mathbb{R}^{s \times s}$, $j = 1, \ldots, t$. The $(s \times s)$-dimensional prediction error covariance matrices will be denoted as

$$V_t = \mathbb{E}(z_{t+1} - \hat{z}_{t+1})(z_{t+1} - \hat{z}_{t+1})'. \tag{4.2.4}$$

If $\operatorname{rk} V_t = q$, $q \leq s$, we may express V_t by means of its canonical decomposition as

$$V_t = O_t \Omega_t O_t',$$

where Ω_t denotes the diagonal matrix containing the q non-zero eigenvalues of V_t in its diagonal and O_t is the matrix of corresponding (normalized) eigenvectors. Then $\hat{z}_{t+1}$ may as well be expressed as a linear combination of the first q principal components of its linear innovations,

$$\hat{z}_{t+1} = \sum_{j=1}^{t} \tilde{\Theta}_{tj} O_{t-j}' (z_{t+1-j} - \hat{z}_{t+1-j}), \tag{4.2.5}$$

where $\tilde{\Theta}_{tj} \in \mathbb{R}^{s \times q}$.

The next theorem will show how to recursively compute the coefficient matrices Θ_{tj} and the prediction error covariance matrices V_t, respectively. Notice that we do not require stationarity of (z_t), but only the existence of its second moments.

Theorem 4.2.1. *Let (z_t), $t \in \mathbb{Z}$, be an s-dimensional process with $\mathbb{E}\, z_t = 0$ for all t and with covariance function $K(i,j) = \mathbb{E}(z_i z_j')$. If for every $t \geq 1$, the prediction error covariance matrix V_t from (4.2.4) is of rank $q \leq s$ then the one-step predictors $\hat{z}_{t+1}$, $t \geq 0$ and V_t, $t \geq 1$ are given as*

$$\hat{z}_{t+1} = \begin{cases} 0 & \text{if } t = 0 \\ \sum_{j=1}^{t} \Theta_{tj} (z_{t+1-j} - \hat{z}_{t+1-j}) & \text{if } t > 0 \end{cases} \tag{4.2.6}$$

and

$$\begin{aligned} V_0 &= K(1,1) \\ V_k &= O_k \underbrace{\Omega_k}_{q \times q} O_k' + 0, \ \textit{(eigenvalue decomposition)} \\ \tilde{\Theta}_{t,t-k} &= \left(K(t+1, k+1) - \sum_{j=0}^{k-1} \Theta_{t,t-j} V_j \Theta_{k,k-j}' \right) O_k \Omega_k^{-1} \\ \Theta_{t,t-k} &= \tilde{\Theta}_{t,t-k} O_k', \quad k = 0, \ldots, t-1 \\ V_t &= K(t+1, t+1) - \sum_{j=0}^{t-1} \Theta_{t,t-j} V_j \Theta_{t,t-j}' \end{aligned} \tag{4.2.7}$$

and the recursions are solved in the order V_0; $\tilde{\Theta}_{1,1}$, $\Theta_{1,1}$, V_1; $\tilde{\Theta}_{2,2}$, $\Theta_{2,2}$, $\tilde{\Theta}_{2,1}$, $\Theta_{2,1}$, V_2; ...

Proof. Let $i < j$, then $z_i - \hat{z}_i \in \mathbb{K}_z(i) \subseteq \mathbb{K}_z(j-1)$ and since $z_j - \hat{z}_j$ is orthogonal to $\mathbb{K}_z(j-1)$, we have

$$(z_i - \hat{z}_i) \perp (z_j - \hat{z}_j), \text{ if } i \neq j. \tag{4.2.8}$$

Hence post multiplication of (4.2.6) by $(z_{k+1}-\hat{z}_{k+1})'$ for $k=0,1,\ldots,t$, and taking expectations, yields

$$\begin{aligned}\mathbb{E}\,\hat{z}_{t+1}(z_{k+1}-\hat{z}_{k+1})' &= \sum_{j=1}^{t}\Theta_{tj}\,\mathbb{E}(z_{t+1-j}-\hat{z}_{t+1-j})(z_{k+1}-\hat{z}_{k+1})' \\ &= \Theta_{t,t-k}V_k \end{aligned} \tag{4.2.9}$$

and since $\mathbb{E}(z_{t+1}-\hat{z}_{t+1})(z_{k+1}-\hat{z}_{k+1})'=0$ implies $\mathbb{E}\,z_{t+1}(z_{k+1}-\hat{z}_{k+1})'=\mathbb{E}\,\hat{z}_{t+1}(z_{k+1}-\hat{z}_{k+1})'$, we also have

$$\mathbb{E}\,z_{t+1}(z_{k+1}-\hat{z}_{k+1})'=\Theta_{t,t-k}V_k \tag{4.2.10}$$

and hence

$$\Theta_{t,t-k}V_k = K(t+1,k+1)-\mathbb{E}\,z_{t+1}\hat{z}_{k+1}'. \tag{4.2.11}$$

Replacing $\hat{z}_{k+1}$ in (4.2.11) by the representation from (4.2.6), we obtain with (4.2.10)

$$\begin{aligned}\Theta_{t,t-k}V_k &= K(t+1,k+1)-\sum_{j=0}^{k-1}\mathbb{E}\,z_{t+1}(z_{j+1}-\hat{z}_{j+1})'\Theta_{k,k-j}' \\ &= K(t+1,k+1)-\sum_{j=0}^{k-1}\Theta_{t,t-j}V_j\Theta_{k,k-j}'. \end{aligned} \tag{4.2.12}$$

Since $V_k=O_k\Omega_kO_k'$ and since with (4.2.7) $\Theta_{t,t-k}O_k=\tilde{\Theta}_{t,t-k}$, post multiplying (4.2.12) by $O_k\Omega_k^{-1}$, yields

$$\Theta_{t,t-k}O_k=\tilde{\Theta}_{t,t-k}=\left(K(t+1,k+1)-\sum_{j=0}^{k-1}\Theta_{t,t-j}V_j\Theta_{k,k-j}'\right)O_k\Omega_k^{-1}. \tag{4.2.13}$$

Then we may write

$$z_{t+1}=z_{t+1}-\hat{z}_{t+1}+\sum_{j=0}^{t-1}\Theta_{t,t-j}(z_{j+1}-\hat{z}_{j+1}), \tag{4.2.14}$$

and since the set $(z_j-\hat{z}_j)$, $j=1,\ldots,t+1$, is orthogonal, taking the covariances of both sides of (4.2.14) finally yields

$$K(t+1,t+1)=V_t+\sum_{j=0}^{t-1}\Theta_{t,t-j}V_j\Theta_{t,t-j}'. \tag{4.2.15}$$

■

4.2.2 The multivariate innovations algorithm in the singular ARMA case

Next we will consider the case of the (singular) causal ARMA(P,Q) process (3.3.64), i.e.

$$a(z)f_t=b(z)\varepsilon_t,$$

where $a(z) = I - A_1 z - \ldots - A_P z^P$, $A_j \in \mathbb{R}^{s\times s}$, $b(z) = B_0 + B_1 z + \ldots B_Q z^Q$, $B_j \in \mathbb{R}^{s\times q}$ and (ε_t) is a q-dimensional white noise with $\Gamma_\varepsilon = I_q$, and show how to apply the algorithm given in Theorem 4.2.1, see also Brockwell and Davis (1987), Section 11.4. To facilitate computations, instead of applying the algorithm to (f_t), we will first apply it to the transformed process

$$w_t = \begin{cases} f_t & 1 \le t \le \max(P,Q) \\ a(z) f_t & t > \max(P,Q), \end{cases} \tag{4.2.16}$$

that has the advantage that its covariance function $K(t,u) = \mathbb{E}(w_t w_u')$ vanishes for $|t-u| > q$, $t, u > \max(P,Q)$. Precisely, let $\Gamma_f(h) = \mathbb{E}(f_t f_{t-h}')$ be the covariance function of (f_t), then $K(t,u)$ is given as

$$K(t,u) = \begin{cases} \Gamma(t-u) & \text{if } 1 \le t \le u \le \max(P,Q), \\ \Gamma(t-u) - \sum_{j=1}^P A_j \Gamma(t-u+j) & \text{if } 1 \le t \le \max(P,Q) < u \le 2\max(P,Q), \\ \sum_{j=0}^Q B_j B_{j+u-1}' & \text{if } \max(P,Q) < t \le u \le t+Q, \\ 0 & \text{if } \max(P,Q) < t \text{ and } u > t+Q, \\ K(t,u)' & \text{if } u > t, \end{cases} \tag{4.2.17}$$

where we used $B_j = 0$ for $j > Q$. Moreover, since (3.3.64) satisfies the causality assumption the Hilbert spaces spanned by present and past values of f_t and w_t are identical, $\mathbb{K}_f(t) = \mathbb{K}_w(t)$. Now applying Theorem 4.2.1 to (w_t), we get

$$\hat{w}_{t+1} = \begin{cases} \sum_{j=1}^t \Theta_{tj}(w_{t+1-j} - \hat{w}_{t+1-j}) & \text{if } 1 \le t < \max(P,Q) \\ \sum_{j=1}^Q \Theta_{tj}(w_{t+1-j} - \hat{w}_{t+1-j}) & \text{if } t > \max(P,Q), \end{cases} \tag{4.2.18}$$

where the coefficients Θ_{tj} and the prediction error covariances V_t are determined as in (4.2.7) with covariance function K as in (4.2.17) and it is easily verified that when both $t > \max(P,Q)$ and $j > Q$, Θ_{tj} vanishes.

To obtain the predictor $\hat{f}_{t+1}$ from (4.2.18), we project both sides of (4.2.16) onto $\mathbb{K}_f(t-1) = \mathbb{K}_w(t-1)$, i.e.

$$\hat{w}_t = \begin{cases} \hat{f}_t & 1 \le t \le \max(P,Q) \\ \hat{f}_t - A_1 f_{t-1} - \ldots - A_P f_{t-P} & t > \max(P,Q). \end{cases} \tag{4.2.19}$$

Further we observe by combining (4.2.16) and (4.2.19), that the linear innovations of f_t and w_t are identical, i.e.

$$(w_t - \hat{w}_t) = (f_t - \hat{f}_t), \tag{4.2.20}$$

such that we obtain for $\hat{f}_{t+1}$,

$$\hat{f}_{t+1} = \begin{cases} \sum_{j=1}^t \Theta_{tj}(f_{t+1-j} - \hat{f}_{t+1-j}) & \text{if } 1 \le t < \max(P,Q) \\ A_1 f_t + \ldots + A_P f_{t+1-P} + \sum_{j=1}^Q \Theta_{tj}(f_{t+1-j} - \hat{f}_{t+1-j}) & \text{if } t > \max(P,Q). \end{cases} \tag{4.2.21}$$

and

$$\mathbb{E}(f_{t+1} - \hat{f}_{t+1})(f_{t+1} - \hat{f}_{t+1})' = V_t. \tag{4.2.22}$$

Example 4.2.1. Consider the case of a (singular) AR(P) process $f_t = A_1 f_{t-1} + \ldots A_P f_{t-P} + B_0 \varepsilon_t$. Then all coefficients Θ_{tj}, $j = 1, 2, \ldots$, are 0 and with (4.2.21) we obtain for $t \geq P$,

$$\hat{f}_{t+1} = A_1 f_t + \ldots + A_P f_{t+1-P},$$

and from (4.2.22),

$$\mathbb{E}(f_{t+1} - \hat{f}_{t+1})(f_{t+1} - \hat{f}_{t+1})' = B_0 B_0'.$$

■

Remark 4.2.1. It can be shown, that, since (3.3.64) satisfies the miniphase assumption, asymptotically for $t \to \infty$ the Hilbert spaces spanned by ε_t and the first q principal components of $(f_{t+1} - \hat{f}_{t+1})$, respectively, coincide, and that $V_t \to B_0 B_0'$ for $t \to \infty$.

4.2.3 h-step prediction of a singular ARMA process

Let (z_t) be a process satisfying the conditions of Theorem 4.2.1. Then the best linear predictor of z_{t+h}, $h \geq 1$, given values z_τ, $\tau = 1, \ldots, t$, is defined as

$$\hat{z}_{t+h|t} = \text{proj}(z_{t+h} | \mathbb{K}_z(t)), \tag{4.2.23}$$

which, since for $h \geq 1$, $\mathbb{K}_z(t) \subseteq \mathbb{K}_z(t+h-1)$, equals

$$\begin{aligned} \hat{z}_{t+h|t} &= \text{proj}\left(\text{proj}(z_{t+h} | \mathbb{K}_z(t+h-1)) | \mathbb{K}_z(t)\right) \\ &= \text{proj}\left(\hat{z}_{t+h} | \mathbb{K}_z(t)\right). \end{aligned} \tag{4.2.24}$$

Replacing $\hat{z}_{t+h}$ in the last expression with its representation in (4.2.6), we thus obtain

$$\hat{z}_{t+h|t} = \text{proj}\left(\sum_{j=1}^{t+h-1} \Theta_{t+h-1,j} (z_{t+h-j} - \hat{z}_{t+h-j}) | \mathbb{K}_z(t) \right), \tag{4.2.25}$$

which, since $(z_{t+h-j} - \hat{z}_{t+h-j}) \perp \mathbb{K}_z(t)$ for $j < h$, yields

$$\hat{z}_{t+h|t} = \sum_{j=h}^{t+h-1} \Theta_{t+h-1,j} (z_{t+h-j} - \hat{z}_{t+h-j}). \tag{4.2.26}$$

For the ARMA(P, Q) process (f_t) given in (3.3.64) it then follows with (4.2.21), that

$$\hat{f}_{t+h|t} = \begin{cases} \sum_{j=h}^{t+h-1} \Theta_{t+h-1,j} (f_{t+h-j} - \hat{f}_{t+h-j}) & 1 \leq h \leq (\max(P,Q) - t) \\ \sum_{i=1}^{P} A_i \hat{f}_{t+h-i|t} + \sum_{j=h}^{Q} \Theta_{t+h-1,j} (f_{t+h-j} - \hat{f}_{t+h-j}), & h > (\max(P,Q) - t), \end{cases} \tag{4.2.27}$$

which can be solved recursively for fixed t for $\hat{f}_{t+1}$, $\hat{f}_{t+2|t}$, ...

Chapter 5

Estimation of the latent variables and the factor space

In the sequel we will be concerned with the estimation of the GDFM variables, that is in the first place the latent variables and (a certain rotation of) the factors and the factor loadings. In some sense, our purpose here is thus to extract the signal and to average-out the noise.

In view of Section 3.1 (see remark 3.1.3) it seems most natural to consider the dynamic PC model to estimate the GDFM, since, as has been shown above, asymptotically these two models coincide. The dynamic PC estimators that have been first proposed and analyzed by Forni *et al.* (2000) will be discussed in detail in Section 5.1. We will see, that these estimators, albeit consistent, possess a severe disadvantage, that is the involvement of two-sided and hence non-causal filters, which obstructs their use in e.g. forecasting. For this reason, estimators based on the quasi-static representation of the GDFM (i.e. equations (3.3.63) and (3.3.64)) that use static PC estimators (Stock and Watson (1998)) or generalized static PC estimators (Bai and Ng (2002), Forni *et al.* (2005a)) have been proposed. Since the corresponding transformations are static, these estimators overcome the problem of two-sidedness occurring in dynamic PCA and may be used for forecasting purposes. We will discuss these estimators and their asymptotic properties in Section 5.2.

As distinct from the previous chapters, where we dealt with the observed process (i.e. infinitely many observations) or the population second moments respectively, the data will now comprise a finite number of observations, (y_t^n), $t = 1, \dots, T$, $n \in \mathbb{N}$ fixed. Notice, that in the sequel quantities based on a finite number of observations, hence estimators, will be provided with a hat $\hat{}$, e.g. if μ_j^n denotes the j-th largest eigenvalue of Γ_y^n, then the corresponding estimate, i.e. the j-th largest eigenvalue of the sample covariance $\hat{\Gamma}_y^n$ will be denoted by $\hat{\mu}_j^n$. Notice further, that throughout this chapter the factor dimensions q and s, respectively, are assumed to be known.

5.1 Estimation by dynamic PCA

Let us recall the dynamic PC model that has been introduced in section 1.2. Commencing from the eigenvalue decomposition of the spectral density f_y^n (1.2.10), i.e.

$$f_y^n(\lambda) = O_1^n(e^{-i\lambda})\Omega_1^n(\lambda)O_1^n(e^{-i\lambda})^* + O_2^n(e^{-i\lambda})\Omega_2^n(\lambda)O_2^n(e^{-i\lambda})^*, \tag{5.1.1}$$

we have defined the n-th order PC model as $y_t^n = \tilde{\chi}_t^n + \tilde{u}_t^n$, where

$$\tilde{\chi}_t^n = O_1^n(z)\psi_t^n, \quad \psi_t^n = O_1^n(z)^* y_t^n, \quad \tilde{u}_t^n = O_2^n(z)O_2^n(z)^* y_t^n. \tag{5.1.2}$$

Lemmas 3.1.10 and 3.1.11, that were parts of the proof of Theorem 3.1.1, have shown that the n-th order PC variables $\tilde{\chi}_t^n$ and $\tilde{u}_t^n$ converge to the GDFM variables χ_t^n and u_t^n as n goes to infinity, precisely for every $i \in \mathbb{N}$ we have that

$$\lim_{n\to\infty} \tilde{\chi}_{it}^n = \chi_{it} \text{ and } \lim_{n\to\infty} \tilde{u}_{it}^n = u_{it}$$

in mean square.

To gain some intuition about this result consider the following simple example.

Example 5.1.1. Let us assume a one-factor model, where one half of the observations is loading contemporaneously and the other half is loading with lag one. Precisely, let ξ_t be white noise with $\text{var}(\xi_t) = 1$, u_t^n be orthogonal white noise with covariance matrix equal to I_n and let $y_t^n = \chi_t^n + u_t^n$ with

$$\chi_{it} = \begin{cases} \xi_t & \text{for } i = 1, 3, 5, \ldots, n-1 \\ \xi_{t-1} & \text{for } i = 2, 4, 6, \ldots, n. \end{cases},$$

and for simplicity assume that the cross-sectional dimension n is even and that it is always increased by an even number.

Then the spectral density of (y_t^n) can be written as

$$f_y^n(\lambda) = \frac{1}{2\pi}\left((1, e^{-i\lambda}, 1, e^{-i\lambda}, \ldots)'(1, e^{i\lambda}, 1, e^{i\lambda}, \ldots) + I_n\right),$$

and one easily verifies that the normalized eigenvector corresponding to the largest eigenvalue of $f_y^n(\lambda)$, $o_1^n(e^{-i\lambda})$ say, equals

$$o_1^n(e^{-i\lambda}) = \frac{1}{\sqrt{n}}(1, e^{-i\lambda}, 1, e^{-i\lambda}, \ldots)',$$

(corresponding to the eigenvalue $\omega_1^n(\lambda) = \frac{n+1}{2\pi}$). The corresponding latent PC-variables, $\tilde{\chi}_t^n = o_1^n(z)o_1^n(z)^* y_t^n$ are thus given as

$$\begin{aligned} \tilde{\chi}_{it}^n &= \begin{cases} \frac{1}{n}(y_{1,t} + y_{2,t+1} + y_{3,t} + \ldots) \\ \frac{1}{n}(y_{1,t-1} + y_{2,t} + y_{3,t-1} + \ldots) \end{cases} \\ &= \begin{cases} \xi_t + \frac{1}{n}(u_{1,t} + u_{2,t+1} + u_{3,t} + \ldots) & \text{for } i = 1, 3, \ldots, n-1 \\ \xi_{t-1} + \frac{1}{n}(u_{1,t-1} + u_{2,t} + u_{3,t-1} + \ldots) & \text{for } i = 2, 4, \ldots, n, \end{cases} \end{aligned}$$

where the latter terms in the last expression converge to 0 in mean square as n goes to infinity, since $\text{var}(\frac{1}{n}(u_{1,t} + u_{2,t+1} + u_{3,t} + \ldots + u_{n,t-1})) = \frac{1}{n}$ and hence the latent PC-variables do indeed converge to the true GDFM variables. ■

This example also demonstrates that the filters involved in dynamic PCA (i.e. $O_1(z)O_1(z)^*$ and $O_2(z)O_2(z)^*$) may be two-sided and that the definition of $\tilde{\chi}_t^n$ and $\tilde{u}_t^n$ may thus require values $y_\tau^n, \tau > t$. Indeed, since the PCA filters are always of the form $O(z)O(z)^*$, they can only be two-sided or static. Also, if $O(z)O(z)^*$ is not static, any transformation that causes the principal components to be a causal transformation of the data (see Molenaar (1987) for this type of transformations), has to result in purely non-causal loadings and vice versa, and is thus futile if the purpose is prediction.

Getting back to estimation, the population spectral density $f_y^n(\lambda)$ in (5.1.1) has to be replaced by an estimator. To prove consistency of the PC estimates we will strengthen Assumption 1.1 a) in that the autocovariances of y_t^n decay fast enough to satisfy

Assumption 5.1 (Consistency of the spectral estimate.)**.**
For all $n \in \mathbb{N}$*,*

$$\sum_{s=-\infty}^{\infty} |s| \|\Gamma_y^n(s)\| < \infty.$$

The spectral estimates we consider here are *lag window* estimates (also known as *covariogram smoothing* estimates) and are of the form

$$\hat{f}_y^n(\lambda) = \frac{1}{2\pi} \sum_{|k| \leq m_T} w(\frac{k}{m_T}) \hat{\Gamma}_y^n(k) e^{-i\lambda k}, \tag{5.1.3}$$

where $\hat{\Gamma}_y^n(k)$ is the sample autocovariance, $w(x)$ is a positive even, piecewise continuous weight function, satisfying $w(0) = 1$, $|w(x)| \leq 1$ for all x and $w(x) = 0$ for $|x| > 1$ and m_T is the truncation parameter defining the lag window size and depending on the sample size T.

Remark 5.1.1. In practice, $\hat{f}_y^n$ cannot be evaluated at $\lambda \in [-\pi, \pi]$, but only at discrete frequencies λ_j and (5.1.3) becomes

$$\hat{f}_y^n(\lambda_j) = \frac{1}{2\pi} \sum_{k \leq m_T} w_{k,T} \hat{\Gamma}_y^n(k) e^{-i\lambda_j k},$$

where, choosing an equally spaced grid of frequencies

$$\lambda_j = \frac{2\pi j}{2m_T + 1}, j = 0, \pm 1, \ldots, \pm m_T.$$

Under Assumption 5.1 any lag-window estimator $\hat{f}_y^n(\lambda)$ with appropriate smoothing weights can be shown to satisfy

$$\lim_{T \to \infty} \text{P}\left(\sup_{\lambda \in [-\pi, \pi]} |\hat{f}_{ij}(\lambda) - f_{ij}(\lambda)| > \varepsilon \right) = 0, \tag{5.1.4}$$

for any $\varepsilon > 0$, where $\hat{f}_{ij}(\lambda)$ and $f_{ij}(\lambda)$ denote the (i,j) element of $\hat{f}_y^n(\lambda)$ and $f_y^n(\lambda)$ respectively (for a reference see e.g. Brockwell and Davis (1987)).

Recall that eigenvalues and normalized eigenvectors are continuous functions of the corresponding matrix elements (see proof of Lemma 1.2.1) and that f_y^n is uniformly continuous in λ (see Brillinger (1981), page 23), implying that the eigenvalues and normalized eigenvectors of f_y^n are themselves uniformly continuous in λ. Denote the eigenvalues and corresponding eigenvectors of $\hat{f}_y^n(\lambda)$ by $\hat{\omega}_j^n(\lambda)$ and $\hat{o}_j^n(\lambda)$ respectively, $j = 1, \ldots, n$, then Slutsky's Theorem together with the uniform continuity imply that for $j = 1, \ldots, n$ and $\varepsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left(\sup_{\lambda\in[-\pi,\pi]} |\hat{\omega}_j^n(\lambda) - \omega_j^n(\lambda)| > \varepsilon \right) = 0 \text{ and} \tag{5.1.5}$$

$$\lim_{T\to\infty} \mathrm{P}\left(\sup_{\lambda\in[-\pi,\pi]} \|\hat{o}_j^n(\lambda) - o_j^n(\lambda)\| > \varepsilon \right) = 0. \tag{5.1.6}$$

The last result indicates convergence of $\hat{\chi}_t^n = \hat{O}_1(z)\hat{O}_1(z)^* y_t^n$ to the n-th order PC variable $\tilde{\chi}_t^n$ as $T \to \infty$ and hence with Lemma 3.1.10 to the GDFM-variable χ_t^n as $n \to \infty$. However, since the filters involved may be two-sided and of infinite length whereas for $t < 1$ and $t > T$ there are no observations available, the filters may have to be truncated. As a consequence of the truncation of $\hat{O}_1(z)\hat{O}_1(z)^*$ convergence can only be granted for a "central" part of the sample, whereas for fixed t the estimators may never be consistent. Let the central part of an increasing sample consist of observations $y_{t^*(T)}^n$, where the $t^*(T)$ are such that there exist real numbers a and b and

$$0 < a \leq \liminf_{T\to\infty} \frac{t^*(T)}{T} \leq \limsup_{T\to\infty} \frac{t^*(T)}{T} \leq b < 1. \tag{5.1.7}$$

We may then state the following result (see Forni *et al.* (2000)):

Theorem 5.1.1. *Under Assumptions 1.1, 2.1 and 5.1, let $\varepsilon > 0, \eta > 0$, then there exist $n^* = n^*(\varepsilon,\eta) \in \mathbb{N}$ and $T^* = T^*(\varepsilon,\eta)$, such that*

$$P(|\hat{\chi}_{it}^n - \chi_{it}| > \varepsilon) \leq \eta,$$

for all $T \geq T^*$, $n \geq n^*$ and t satisfying (5.1.7).

Proof. For a proof see Forni *et al.* (2000), Proposition 3. ■

Consequently the resulting estimates may get worse when approaching the ends of the sample (since then truncation of the filters may have more severe effects). Since for forecasting purposes obviously the most recent data points are of great relevance, the dynamic PC estimators may not be optimal. Estimators that overcome the shortcomings of the dynamic PC estimates will be discussed in the next section.

5.2 Estimation by static or generalized static PCA

First recall, that under Assumptions 1.1 and 3.1, as it has been shown in Section 3.3 there always exists a quasi-static representation of the GDFM (2.1.1), which in general can be achieved at the cost of a higher dimensional quasi-static factor. This representation has been given in (3.3.63) and (3.3.64).

Our purpose here, is the estimation of the (quasi-) static factors or to be more precise the static factor space and the estimation of the latent variable as a projection onto that space. Once again the essential idea is to average out the noise term, but now averaging will be based on representation (3.3.63) and the corresponding variance decomposition,

$$\Gamma_y^n = \bar{\Lambda}^n \bar{\Lambda}^{n'} + \Gamma_u^n, \tag{5.2.1}$$

and will be restricted to static transformations of y_t^n.

To obtain asymptotic identifiability of (5.2.1) we need to assume that the s-largest eigenvalue of $\Gamma_\chi^n = \bar{\Lambda}^n \bar{\Lambda}^{n'}$, $\mu_{\chi,s}^n$ say[1], diverges as $n \to \infty$ (see Chamberlain and Rothschild (1983)), which rules out the case in which some of the elements of f_t are only loaded by a finite number of observations. This is indeed a restriction of generality, since as it is easily seen it is not implied by Assumption 2.1 b). Consider for instance the example given in Forni *et al.* (2005a), i.e. the GDFM, where $\chi_{1t} = \varepsilon_{t-1}$ and $\chi_{jt} = \varepsilon_t$, $j > 1$, hence $q = 1$ and $s = 2$ and Assumption 2.1 b) is clearly satisfied. However $\mu_{\chi,2}^n = 1$ and thus bounded. Notice that on the other hand, Assumption 2.1 a) does imply that the eigenvalues of Γ_u^n, $\mu_{u,j}^n$ say, are bounded from above as $n \to \infty$, since

$$\Gamma_u^n = \int_{-\pi}^{\pi} f_u^n(\lambda) d\lambda \leq \int_{-\pi}^{\pi} \omega_{u,1}^n(\lambda) I_n d\lambda \leq 2\pi \ \bar{\omega}_{u1} \ I_n,$$

where $\bar{\omega}_{u1} = \operatorname{ess\,sup}_n(\omega_{u,1}^n)$.

Furthermore we will assume that the eigenvalues of Γ_χ^n, grow linearly in n and that they are distinct.

Assumption 5.2 (Eigenvalues of Γ_χ^n)**.**
For $j = 1, \ldots, s$ there exist positive real constants $\bar{c}_j$, $\underline{c}_j$ such that

$$0 \leq \underline{c}_s \leq \liminf_{n\to\infty} n^{-1} \mu_{\chi,s}^n \leq \bar{c}_s < \ldots < \underline{c}_1 \leq \liminf_{n\to\infty} n^{-1} \mu_{\chi,1}^n \leq \bar{c}_1 < \infty.$$

Indeed linear growth does not mean much loss of generality and is of rather technical nature, since there is no ordering in the cross-sectional dimension and hence there is no obvious reason why the divergence should decelerate or accelerate as n gets larger.

In the following we will consider estimates based on:

[1]Notice that in the sequel we will denote eigenvalues corresponding to covariance matrices by the letter μ to distinguish them from eigenvalues corresponding to spectral densities that have been denoted by ω.

- The static PC model as proposed and analyzed by Stock and Watson (1998).
- A generalized (or weighted) static PC model as considered by Boivin and Ng (2006).
- A generalized static PCA based on the covariance matrices resulting from the dynamic PC model as proposed and analyzed by Forni *et al.* (2005a).

Let us first present these three methods. Subsequently we will then discuss their asymptotic properties.

5.2.1 The static PC estimates

The static (or quasi-static) PC model has been discussed in detail in section 1.2. Recall however, that commencing from the objective function

$$\min_{\substack{B^n\in\mathbb{R}^{s\times n}\\ C^n\in\mathbb{R}^{n\times s}}} \operatorname{tr}\mathbb{E}(y_t^n - C^nB^ny_t^n)(y_t^n - C^nB^ny_t^n)', \tag{5.2.2}$$

which is minimized (for fixed s) via the canonical representation of Γ_y^n (1.2.3), i.e.

$$\Gamma_y^n = O_1^n\Omega_1^nO_1^{n\prime} + O_2^n\Omega_2^nO_2^{n\prime}, \tag{5.2.3}$$

if $B^n = O_1^{n\prime}$, $C^n = O_1^n$. Hence we defined the static PC variables as

$$\tilde{\chi}_t^{n,PCA} = O_1^n\Omega_1^{n^{1/2}}\psi_t^{n,PCA},\ \psi_t^{n,PCA} = \Omega_1^{n^{-1/2}}O_1^{n\prime}y_t^n,\ \tilde{u}_t^{n,PCA} = O_2^nO_2^{n\prime}y_t^n,\quad t\in\mathbb{Z}, \tag{5.2.4}$$

where here we have rescaled the principal components $\psi_t^{n,PCA}$ to have unit variance. Notice, that due to orthogonality of O_1^n and O_2^n, the latent PC variables $\tilde{\chi}_t^{n,PCA}$ are the projections of y_t^n onto $\overline{\text{span}}(\psi_{it}^{n,PCA}|i=1,\dots,s)$.

Remark 5.2.1. Notice, that rescaling of the principal components is not necessary for estimation (since it does not change the latent PC variables) but will simplify the proof of consistency, since, as will be shown below, each row of $\Omega_1^{n^{-1/2}}O_1^{n\prime}$ is a DAS implying that $\psi_t^{n,PCA}$ converges to the factor space as n increases.

Let us continue example 5.1.1 to gain some intuition about the static PC method.

Example 5.2.1. The one-factor GDFM may be written in static form with 2-dimensional stacked factor $f_t = (\xi_t, \xi_{t-1})'$ and corresponding factor loading matrix

$$\bar{\Lambda}^n = \begin{pmatrix} 1 & 0\\ 0 & 1\\ 1 & 0\\ \vdots & \vdots \end{pmatrix}.$$

The covariance matrix of y_t^n is equal to

$$\Gamma_y^n = \bar{\Lambda}^n\bar{\Lambda}^{n\prime} + I_n = \begin{pmatrix} 1 & 0 & 1 & 0 & \dots\\ 0 & 1 & 0 & 1 & \dots\\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} + I_n,$$

and one may easily check that the 2 largest eigenvalues equal $\frac{n}{2}+1$, where

$$o_1^n = \sqrt{\frac{2}{n}}(1,0,1,0,\ldots)' \text{ and}$$

$$o_2^n = \sqrt{\frac{2}{n}}(0,1,0,1,\ldots)'$$

are corresponding normalized eigenvectors. The latent static PC variables are thus given as

$$\begin{aligned}\tilde{\chi}_{it}^{n,PCA} &= \begin{cases} \frac{2}{n}(y_{1,t}+y_{3,t}+\ldots+y_{n-1,t}) \\ \frac{2}{n}(y_{2,t}+y_{4,t}+\ldots+y_{n,t}) \end{cases} \\ &= \begin{cases} \xi_t + \frac{2}{n}(u_{1,t}+u_{3,t}+\ldots+u_{n-1,t}) & \text{for } i=1,3,\ldots,n-1 \\ \xi_{t-1} + \frac{2}{n}(u_{2,t}+u_{4,t}+\ldots+u_{n,t}) & \text{for } i=2,4,\ldots,n. \end{cases}\end{aligned}$$

Again the latter terms in the last expressions vanish in mean square as n goes to infinity, since $\text{var}(\frac{2}{n}(u_{1,t}+u_{3,t}+\ldots+u_{n-1,t})) = \frac{2}{n}$ and hence the latent static PC variables do converge to the true GDFM variables. However the rate of convergence is two times slower than it is for the dynamic PC variables, which is explained by the fact that here only half of the variables are exploited to average out the noise term. ■

For estimation Γ_y^n in (5.2.3) is replaced by the sample covariance

$$\hat{\Gamma}_y^n = \frac{1}{T}\sum_{t=1}^{T} y_t^n y_t^{n'}, \tag{5.2.5}$$

and the static PC estimates are then using an obvious notation given as

$$\hat{\chi}_t^{n,PCA} = \hat{O}_1^n \hat{\Omega}_1^{n\,1/2} \hat{\psi}_t^{n,PCA},\ \hat{\psi}_t^{n,PCA} = \hat{\Omega}_1^{n\,-1/2} \hat{O}_1^{n'} y_t^n,\ \hat{u}_t^{n,PCA} = \hat{O}_2^n \hat{O}_2^{n'} y_t^n,\quad t=1,\ldots,T. \tag{5.2.6}$$

Remark 5.2.2. Notice that, since by Lemma A.1.1, $o_1^n,\ldots,o_s^n$ are the solution of

$$\begin{aligned} o_i^n \qquad &= \arg\max_{b\in\mathbb{R}^{n\times 1}} b'\Gamma_y^n b \\ \text{s.t.}\quad & b'b=1,\ b'o_j^n=0,\ j=1,\ldots,i-1, \end{aligned}$$

(where for $i=1$ only the first constraint applies), the principal components, are (apart from rescaling) those static, orthogonal transformations of the data, that have maximum variance.

5.2.2 The generalized static PC estimates

The generalized (or weighted) static PC model is related to classical PCA in an analogous way as generalized least squares (GLS) is related to ordinary least squares in the regression context. Recall, that instead of equally weighting all residuals GLS weights the residuals proportional to the inverse of the square root of their variance matrix resulting in an estimator (for the

conditional mean) that puts lower weight on observations with higher residual variance and is such (linearly) efficient. In the multivariate, case a similar principle may be applied to the cross-sectional dimension. Thus instead of the least squares objective function (5.2.2) we may consider the generalized least squares problem

$$\min_{\substack{B\in\mathbb{R}^{s\times n}\\ C\in\mathbb{R}^{n\times s}}} \operatorname{tr}\mathbb{E}(y_t^n - C^nB^ny_t^n)'\Gamma_u^{n^{-1}}(y_t^n - C^nB^ny_t^n). \tag{5.2.7}$$

Roughly speaking, here lower weight is put on observations with higher noise variance that are therefore less reliable. Observing that (5.2.7) is equal to

$$\min_{\substack{B\in\mathbb{R}^{s\times n}\\ C\in\mathbb{R}^{n\times s}}} \operatorname{tr}\left((\Gamma_u^{n^{-1/2}}\Gamma_y^{n^{1/2}} - \Gamma_u^{n^{-1/2}}C^nB^n\Gamma_y^{n^{1/2}})(\Gamma_u^{n^{-1/2}}\Gamma_y^{n^{1/2}} - \Gamma_u^{n^{-1/2}}C^nB^n\Gamma_y^{n^{1/2}})'\right), \tag{5.2.8}$$

and Lemma A.1.2 imply that a minimum is achieved if

$$\Gamma_u^{n^{-1/2}}C^nB^n\Gamma_y^{n^{1/2}} = \bar{V}_1^n\Delta_1^{n^{1/2}}\bar{V}_1^{n'}. \tag{5.2.9}$$

Here Δ_1^n is the $(s\times s)$ diagonal matrix containing the s largest eigenvalues of $\Gamma_u^{n^{-1/2}}\Gamma^n{}_y\Gamma_u^{n^{-1/2}}$ (i.e. the s largest generalized eigenvalues of the couple of matrices(Γ_y^n,Γ_u^n), see section A.2) in its diagonal, and $\bar{V}_1^n$ is the matrix of corresponding eigenvectors. Minimizers B^n, C^n of (5.2.7) thus have to satisfy

$$C^nB^n = \Gamma_u^{n^{1/2}}\bar{V}_1^n\Delta_1^{n^{1/2}}\bar{V}_1^{n'}\Gamma_y^{n^{-1/2}}. \tag{5.2.10}$$

And since a little algebra shows, that

$$\Gamma_y^{n^{-1/2}} = \bar{V}^n\Delta_1^{n^{-1/2}}\bar{V}^{n'}\Gamma_u^{n^{-1/2}}, \tag{5.2.11}$$

we obtain, that in the minimum,

$$C^nB^n = \Gamma_u^{n^{1/2}}\bar{V}_1^n\bar{V}_1^{n'}\Gamma_u^{n^{-1/2}}, \tag{5.2.12}$$

such that we can choose $B^n = \bar{V}_1^{n'}\Gamma_u^{n^{-1/2}}$, $C^n = \Gamma_u^{n^{1/2}}\bar{V}_1^n$.

In view of section A.2, denoting the matrix of generalized eigenvectors of (Γ_y^n,Γ_u^n) corresponding to the s largest generalized eigenvalues by V_1^n, we have $V_1^n = \Gamma_u^{n^{-1/2}}\bar{V}_1^{n'}$ and hence we may alternatively write $B^n = V_1^{n'}$, $C^n = \Gamma_u^nV_1^n$. We then may define the generalized principal components $\tilde{\psi}_t^{n,GPCA}$ as $\tilde{\psi}_t^{n,GPCA} = V_1^{n\prime}y_t^n$ with covariance matrix equal to Δ_1^n. Hence rescaling the generalized principal components to have unit variance, we define the generalized static PC variables as

$$\tilde{\chi}_t^{n,GPCA} = \Gamma_u^nV_1^n\Delta_1^{n^{1/2}}\psi_t^{n,GPCA},\ \psi_t^{n,GPCA} = \Delta_1^{n^{-1/2}}V_1^{n\prime}y_t^n,\ \ \tilde{u}_t^{n,GPCA} = \Gamma_u^nV_2^nV_2^{n'}y_t^n,\ \ t\in\mathbb{Z}. \tag{5.2.13}$$

Again, since

$$\mathbb{E}(y_t^n\psi_t^{n,GPCA'}) = \Gamma_y^nV_1^n\Delta_1^{n^{-1/2}} = \Gamma_u^nV_1^n\Delta_1^{n^{1/2}},$$

the latent generalized PC variable $\tilde{\chi}_t^{n,GPCA}$ is equal to the projection of y_t^n onto the space spanned by the scalar components of $\psi_t^{n,GPCA}$.

For estimation, Γ_y^n and Γ_u^n in (5.2.8) have to be replaced by estimators: again Γ_y^n will be replaced by the sample covariance, however the estimation of Γ_u^n is more delicate, since the estimate has to be invertible. Taking e.g. the sample error covariance matrix from the corresponding unweighted PC model is infeasible, since it is by construction singular (with rank $n-s$). A feasible alternative is the diagonal matrix containing the diagonal elements of the sample covariance of the unweighted PC residuals in its diagonal as proposed in Boivin and Ng (2006). Clearly this estimator disregards a potential correlation structure among the noise part. Another estimator is the estimate based on dynamic PCA that has been proposed in Forni *et al.* (2005a) and that will be discussed below. Assuming that $\hat{\Gamma}_u^n$ is a full rank estimator for Γ_u^n, the estimators corresponding to the generalized static PC model are, using an obvious notation, given as

$$\hat{\chi}_t^{n,GPCA} = \hat{\Gamma}_u^n \hat{V}_1^n \hat{V}_1^{n\prime} y_t^n,\ \hat{\psi}_t^{n,GPCA} = \hat{\Delta}_1^{n\,-1/2} \hat{V}_1^{n\prime} y_t^n,\ \hat{u}_t^{n,GPCA} = \hat{\Gamma}_u^n \hat{V}_2^n \hat{V}_2^{n\prime} y_t^n,\quad t = 1, \ldots, T. \tag{5.2.14}$$

Remark 5.2.3. Notice, that as discussed in section 3.2 under the assumption that Γ_u^n is diagonal, the generalized static PC estimators are equal to the maximum likelihood estimators in the classical factor analytic model with orthogonal noise (see Lawley and Maxwell (1971), Chapter 4).

Remark 5.2.4. In correspondence to remark 5.2.2 notice, that by Lemma A.2.1, the generalized eigenvectors $v_1^n, \ldots, v_s^n$ are the solution of

$$\begin{aligned} v_j^n \quad &= \quad \arg\max_{b \in \mathbb{R}^n} b' \Gamma_y^n b \\ \text{s.t.} \quad &\quad b' \Gamma_u b = 1,\quad b' \Gamma_u^n v_i^n = 0, i = 1, \ldots, j-1. \end{aligned} \tag{5.2.15}$$

(where for $j = 1$ only the first constraint applies). And hence the generalized principal components are those static, orthogonal transformations of the data, that have maximal variance under the constraint that the contributing residual variance is restricted.

5.2.3 Generalized static PC estimates of FHLR

Another estimator based on a generalized eigenvalue problem has been proposed in Forni *et al.* (2005a), FHLR henceforth. It differs from the generalized PC estimators discussed above mainly in that the covariance matrices involved are $\hat{\Gamma}_\chi^n$ and $\hat{\Gamma}_u^n$ (instead of $\hat{\Gamma}_y^n$ and $\hat{\Gamma}_u^n$) resulting from a first-step frequency domain PCA. Hence roughly speaking, compared to the last section the idea is, not only to put lower weight on observations with higher noise variance, but also to put higher weight on observations with high signal variance.

In the first step, the spectral density matrix $f_y^n(\lambda)$ of y_t^n is decomposed into $\tilde{f}_\chi^n(\lambda)$ and $\tilde{f}_u^n(\lambda)$ according to the dynamic PC model (5.1.2), hence $\tilde{f}_\chi^n(\lambda)$ and $\tilde{f}_u^n(\lambda)$ respectively are the matrices

on the right hand side of equation (5.1.1). Then the covariance matrices of the n-th order dynamic PC variables are given as

$$\tilde{\Gamma}_{\chi}^{n} = \int_{-\pi}^{\pi} \tilde{f}_{\chi}^{n}(\lambda) d\lambda \quad \text{and} \quad \tilde{\Gamma}_{u}^{n} = \int_{-\pi}^{\pi} \tilde{f}_{u}^{n}(\lambda) d\lambda. \tag{5.2.16}$$

Convergence of the n-th order dynamic PC variables (5.1.2) to the corresponding GDFM variables in mean square as $n \to \infty$ (as shown in Lemma 3.1.10) implies that every element of $\tilde{f}_{\chi}^{n}(\lambda)$ and $\tilde{f}_{u}^{n}(\lambda)$ respectively converges to the corresponding element of the GDFM spectral densities $f_{\chi}^{n}(\lambda)$ and $f_{u}^{n}(\lambda)$ λ-a.e. in $[-\pi, \pi]$ as $n \to \infty$ (see Lemma 3.1.4). Since furthermore each of the elements of $\tilde{f}_{\chi}^{n}(\lambda)$ and $\tilde{f}_{u}^{n}(\lambda)$ are bounded from above by the corresponding elements of $f_{y}^{n}(\lambda)$ and since the latter are bounded from above by Assumption 1.1, it follows by Lebesgue's dominated convergence theorem, that the elements of $\tilde{\Gamma}_{\chi}^{n}$ and $\tilde{\Gamma}_{u}^{n}$ converge to the corresponding GDFM quantities as $n \to \infty$. Replacing $f_{y}^{n}(\lambda)$ with a consistent lag-window estimator $\hat{f}_{y}^{n}(\lambda)$ (5.1.3), yields, since as argued before, eigenvectors and eigenvalues are uniformly continuous, consistent estimates $\hat{f}_{\chi}^{n}(\lambda)$ and $\hat{f}_{u}^{n}(\lambda)$ of $\tilde{f}_{\chi}^{n}(\lambda)$ and $\tilde{f}_{u}^{n}(\lambda)$ respectively as $T \to \infty$. Concluding,

$$\hat{\Gamma}_{\chi}^{n} = \int_{-\pi}^{\pi} \hat{f}_{\chi}^{n}(\lambda) d\lambda \quad \text{and} \quad \hat{\Gamma}_{u}^{n} = \int_{-\pi}^{\pi} \hat{f}_{u}^{n}(\lambda) d\lambda \tag{5.2.17}$$

are consistent estimates for $\tilde{\Gamma}_{\chi}^{n}$ and $\tilde{\Gamma}_{u}^{n}$ respectively, as $T \to \infty$.

The second step consists of the estimation of the factor space and the latent variables as projections onto that space.

For the estimation of the factor space, consider the generalized eigenvalue decomposition of $\tilde{\Gamma}_{\chi}^{n}$ with respect to $\tilde{\Gamma}_{u}^{n}$. Then let D_{1}^{n} and D_{2}^{n} denote the diagonal matrices containing the s largest and $n - s$ smallest generalized eigenvalues respectively in their diagonals, and let W_{1}^{n} and W_{2}^{n} denote the corresponding matrices of generalized eigenvectors, normalized such that $W_{j}^{n\prime} \tilde{\Gamma}_{u}^{n} W_{j}^{n} = I_{n}, j = 1, 2$.

The aggregates[2], $\psi_{t}^{n,FHLR}$ say, normalized to have unit variance, are then of the form

$$\psi_{t}^{n,FHLR} = (I_{s} + D_{1}^{n^{1/2}})^{-1} W_{1}^{n\prime} y_{t}^{n}, \tag{5.2.18}$$

where the normalizing factor $(I_{s} + D_{1}^{n^{1/2}})^{-1}$ results from the fact that

$$W_{1}^{n\prime} \Gamma_{y}^{n} W_{1}^{n} = W^{n\prime}{}_{1} (\tilde{\Gamma}_{\chi}^{n} + \tilde{\Gamma}_{u}^{n}) W_{1}^{n} = D_{1}^{n} + I_{s}.$$

Hence the aggregates $\psi_{t}^{n,FHLR}$ differ from $\psi_{t}^{n,GPCA}$ in that in the first case the weight matrix consists of the generalized eigenvectors of $(\tilde{\Gamma}_{\chi}^{n}, \tilde{\Gamma}_{u}^{n})$, whereas in the latter case it consists of the generalized eigenvectors of $(\Gamma_{y}^{n}, \tilde{\Gamma}_{u}^{n})$. Notice, that as distinct to the methods discussed above,

[2]Notice, that we will show below that the components of $\psi_{t}^{n,\cdot}$ are indeed aggregates in the sense of Definition 3.1.1, however for the moment the term may as well be understood in its colloquial sense.

the estimates here are neither the solution of a least squares problem (5.2.2) nor a generalized least squares problem (5.2.7). However, applying Lemma A.2.1, the generalized eigenvectors $w_1^n, \ldots, w_s^n$ solve the following maximization problem

$$\begin{aligned} w_j^n &= \arg\max_{b\in\mathbb{R}^n} b'\Gamma_\chi^n b \\ \text{s.t.} \qquad & b'\Gamma_u b = 1, \quad b'\Gamma_u^n w_i^n = 0, i = 1, \ldots, j-1, \end{aligned} \tag{5.2.19}$$

(where the second constraint only applies for $j > 1$). Hence in correspondence to remark 5.2.4, we may argue that the aggregates $\psi_t^{n,FHLR}$, are such that the part of their variance that is affiliated to the latent variable is maximized under the constraint that the part of their variance that is affiliated to the noise is restricted.

The latent variables are then defined as the projections onto the space spanned by $\psi_t^{n,FHLR}$. Notice, that trivially,

$$\begin{aligned} \chi_t^n &= \text{proj}(y_t^n|\text{span}(f_{it}|i = 1, \ldots, s)) \\ &= \text{proj}(\chi_t^n|\text{span}(f_{it}|i = 1, \ldots, s)) + \text{proj}(u_t^n|\text{span}(f_{it}|i = 1, \ldots, s)) \\ &= \text{proj}(\chi_t^n|\text{span}(f_{it}|i = 1, \ldots, s)), \end{aligned}$$

where

$$\text{proj}(y_t^n|\text{span}(f_{it}|i = 1, \ldots, s)) = \mathbb{E}(y_t^n f_t')(\mathbb{E}(f_t f_t'))^{-1} f_t \tag{5.2.20}$$

and

$$\text{proj}(\chi_t^n|\text{span}(f_{it}|i = 1, \ldots, s)) = \mathbb{E}(\chi_t^n f_t')(\mathbb{E}(f_t f_t'))^{-1} f_t. \tag{5.2.21}$$

Hence replacing f_t by the aggregates $\psi_t^{n,FHLR}$, the projections (5.2.20) and (5.2.21) are approximated corresponding to (5.2.20) as

$$\tilde{\chi}_t^n = \Gamma_y^n W_1^n (I_s + \hat{D}_1^{n^{1/2}})^{-1} \psi_t^{n,FHLR} \tag{5.2.22}$$

or, as proposed in Forni *et al.* (2005a), corresponding to (5.2.21), as

$$\tilde{\chi}_t^{n,FHLR} = \Gamma_\chi^n (I_s + \hat{D}_1^{n^{1/2}})^{-1} \psi_t^{n,FHLR}. \tag{5.2.23}$$

Lastly, for estimation, $\tilde{\Gamma}_\chi^n$ and $\tilde{\Gamma}_u^n$ are replaced by their estimators (5.2.17).

Comparing the PCA and GPCA method on the one hand and the FHLR method on the other hand, one may argue that since $\hat{\Gamma}_\chi^n$ is based on the dynamic PC model it exploits the dynamic correlation structure and may thus yield more efficient estimates. However general statements about efficiency gains are difficult and depend largely on the underlying model structure. At least from a theoretical point of view, the question of (asymptotic) efficiency gain is not yet answered. In Forni *et al.* (2005a) and Boivin and Ng (2006) Monte Carlo experiments are described showing that the FHLR and GPCA estimators are better than the PCA estimators in certain cases, whereas the results of an empirical study in the context of macroeconomic forecasting

(see Stock and Watson (2005)) do not show much difference. For the moment, let us get back to our simple example (see examples 5.1.1 and 5.2.1) at least to gain some intuition about the method.

Example 5.2.2. From the dynamic PCA of y_t^n (see example 5.1.1) we conclude that the spectral density of the latent PC variable, $\tilde{f}_\chi^n(\lambda) = o_1^n(\lambda)\omega_1^n(\lambda)o_1^n(\lambda)^*$, is of the form

$$\tilde{f}_\chi^n(\lambda) = \frac{1}{2\pi}\frac{n+1}{n}\begin{pmatrix} 1 & e^{i\lambda} & 1 & e^{i\lambda} & \dots \\ e^{-i\lambda} & 1 & e^{-i\lambda} & 1 & \dots \\ 1 & e^{i\lambda} & 1 & e^{i\lambda} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

and hence we have for the respective covariance matrix

$$\begin{aligned} \tilde{\Gamma}_\chi^n &= \int_\pi^{-\pi} \tilde{f}_\chi^n(\lambda)d\lambda \\ &= \left(1+\frac{1}{n}\right)\begin{pmatrix} 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & 1 & \dots \\ 1 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \end{aligned}$$

Comparing $\tilde{\Gamma}_\chi^n$ to the covariance matrix of the static PC variables (see example 5.2.1) , i.e.

$$\tilde{\Gamma}_\chi^{n,PCA} = \left(1+\frac{2}{n}\right)\begin{pmatrix} 1 & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & 1 & \dots \\ 1 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

we see that its convergence to Γ_χ is faster as n increases which is due to the fact, that roughly speaking, in this example, the static PCA disregards half of the variables, whereas the dynamic PCA by aggregating lead and lagged variables exploits the whole panel.

Then, since $\tilde{\Gamma}_\chi^n + \tilde{\Gamma}_u^n = \Gamma_y$, we have

$$\tilde{\Gamma}_u^n = \frac{1}{n}\begin{pmatrix} n-1 & 0 & -1 & 0 & \dots \\ 0 & n-1 & 0 & -1 & \dots \\ -1 & 0 & n-1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The largest generalized eigenvalues of the pair $(\tilde{\Gamma}_\chi^n, \tilde{\Gamma}_u^n)$ are $\nu_1^n = \nu_2^n = n+1$ and

$$w_1^n = \frac{2}{\sqrt{n}}(1,0,1,0,\dots)'$$

and

$$w_2^n = \frac{2}{\sqrt{n}}(0,1,0,1,\ldots)'$$

are corresponding generalized eigenvectors (that have been normalized to satisfy ${w^n}_j'\tilde{\Gamma}_u^n w_j^n = 1$).

Determining $\tilde{\chi}_t^{n,FHLR}$ according to (5.2.23) we have, with $W_1^n = (w_1^n, w_2^n)$

$$\psi_t^{n,FHLR} = {W^n}_1' y_t = \frac{2}{\sqrt{n}} \begin{pmatrix} 1 & 0 & 1 & 0 & \ldots \\ 0 & 1 & 0 & 1 & \ldots \end{pmatrix} y_t,$$

$$({W^n}_1' \Gamma_y W_1^n) = \begin{pmatrix} n+2 & 0 \\ 0 & n+2 \end{pmatrix},$$

and

$$\tilde{\Gamma}_\chi^n W_1^{n'} = \begin{pmatrix} \frac{n+1}{\sqrt{n}} & 0 \\ 0 & \frac{n+1}{\sqrt{n}} \\ \frac{n+1}{\sqrt{n}} & 0 \\ \ldots & \ldots \end{pmatrix}.$$

Summarizing we then have,

$$\begin{aligned} \tilde{\chi}_{it}^{n,FHLR} &= \begin{cases} \frac{2}{n}\frac{n+1}{n+2}(y_{1,t} + y_{3,t} + \ldots + y_{n-1,t}) \\ \frac{2}{n}\frac{n+1}{n+2}(y_{2,t} + y_{4,t} + \ldots + y_{n,t}) \end{cases} \\ &= \begin{cases} \frac{n+1}{n+2}\xi_t + \frac{2}{n}\frac{n+1}{n+2}(u_{1,t} + u_{3,t} + \ldots + u_{n-1,t}) & \text{for } i = 1,3,\ldots,n-1 \\ \frac{n+1}{n+2}\xi_{t-1} + \frac{2}{n}\frac{n+1}{n+2}(u_{2,t} + u_{4,t} + \ldots + u_{n,t}) & \text{for } i = 2,4,\ldots,n. \end{cases} \end{aligned}$$

We observe, that $\tilde{\chi}_{it}^{n,FHLR}$ albeit consistent (for $n \to \infty$) it is biased, while at the same time the error term vanishes slightly faster than in conventional PCA. ■

5.2.4 Asymptotic properties

In this subsection we will be concerned with consistency results for the three estimates presented above. Here we will largely follow the proofs given in Forni *et al.* (2005a) for the FHLR-estimates and adapt them for the PCA- and GPCA-estimates, respectively. First we will show that, as n increases, the spaces spanned by the three aggregates $\psi_t^{n,\cdot}$ presented above converge to the static factor space, where convergence of spaces is understood in the sense that the perpendiculars of a projection from one space to the other converge to 0 in mean square. Second we will show, that the projections $\tilde{\chi}_t^{n,\cdot}$ onto these aggregates converge to the latent GDFM variable. Notice that these two issues still concern population results. And third, we will see, that as the sample size T goes to infinity, the three estimates $\hat{\chi}_t^{n,\cdot}$ converge in probability to their population counter parts and are thus consistent estimates for the GDFM variable χ_t^n. Besides we will show that, as n and T go to infinity, the (static) factor space can be consistently estimated by $\hat{\psi}_t^{n,\cdot}$

Notice that throughout this section, we will always assume w.l.o.g. that the smallest eigenvalue of Γ_u^n, $\mu_{u,n}^n$ say, is bounded away from 0 for all $n \in \mathbb{N}$ (see the the discussion below (3.1.4) for

details) and that $\Gamma_f = I_s$. Furthermore we will assume that the latent variables (χ_t^n) as well as the noise processes (u_t^n) satisfy the following conditions in addition to Assumption 1.1.

Assumption 5.3.
The latent variable process (χ_t^n) and the noise processes (u_t^n), $n \in \mathbb{N}$, are such that

- *For all $i, j \in \mathbb{N}$ the processes $(u_{it}u_{jt})$ are stationary with absolutely summable autocovariances.*
- *For all $i, j \in \mathbb{N}$ the processes $(\chi_{it}\chi_{jt})$ are stationary with absolutely summable autocovariances.*

Assumption 5.3 implies that an analogous statement holds true for the processes $(y_{it}y_{jt})$ and ensures that all covariance matrices involved may be consistently estimated for fixed n as $T \to \infty$ (see e.g. Brockwell and Davis (1987), 7.2).

To proof consistency in case of the GPCA and FHLR estimates, we will need the following lemma (see Forni *et al.* (2005a), Lemma 7.1) providing some asymptotic properties of the generalized eigenvalues of a couple of matrices.

Lemma 5.2.1. *Given an integer $s > 0$, consider sequences Γ^n and Σ^n of real, symmetric positive semi-definite $(n \times n)$-matrices, $n = s, s+1, \ldots$ and assume that*

- *The s-th largest eigenvalue of Γ^n, μ_s^n say, diverges as $n \to \infty$, whereas $\sup_n \mu_{s+1}^n$ is bounded.*
- *Σ^n's smallest eigenvalue, σ_n^n say, is bounded away from 0, i.e. $\sigma_{inf} = \inf_n \sigma_n^n > 0$ and its largest eigenvalue is bounded from above, i.e. $\sigma_{sup} = \sup_n \sigma_1^n < \infty$.*

and denote by ν_j^n, $j = 1, \ldots, n$, the generalized eigenvalues of (Γ^n, Σ^n) in descending order of magnitude. Then the s-th largest generalized eigenvalue diverges as $n \to \infty$, i.e. $\sup_n \nu_s^n = \infty$ and ν_{s+1}^n is bounded for all $n \in \mathbb{N}$.

Proof. Since the generalized eigenvalues of (Γ^n, Σ^n) are the eigenvalues of $\Sigma^{n-1/2}\Gamma^n\Sigma^{n-1/2}$, applying the Courant-Fisher Theorem (Lemma A.1.1) yields for a given $(n \times (j-1))$ complex matrix D_j of full column rank and for v such that $D_j'v = 0$

$$\begin{aligned} \nu_j^n &= \max_{v'v=1} v'\Sigma^{n^{-1/2}}\Gamma^n\Sigma^{n^{-1/2}}v \\ &\leq \mu_j^n v'\Sigma^{n^{-1}}v \leq \frac{\mu_j^n}{\sigma_{inf}}, \end{aligned} \tag{5.2.24}$$

where for $j \geq s+1$ the last expression is bounded.

Let $v_1^n, \ldots, v_{s-1}^n$ denote the (normalized) generalized eigenvectors of (Γ^n, Σ^n) corresponding to $\nu_1^n, \ldots, \nu_{s-1}^n$. Further let $O_1^n = (o_1^n, \ldots, o_s^n)$ denote the matrix consisting of the normalized eigenvectors of Γ^n corresponding to the s largest eigenvalues and Ω_1^n the diagonal matrix containing

the s largest eigenvalues in its diagonal. Consider the $s-1$ linear equations in the s-dimensional unknown x:

$$v_j^{n'}\Sigma^n O_1^n x = 0,\ j = 1, \ldots, s-1. \tag{5.2.25}$$

Let x_0 denote a solution of (5.2.25) and define $q^n = O_1^n x_0$, hence $v_j^{n'}\Sigma^n q^n = 0$. And since Σ^n is positive definite with smallest eigenvalue bounded away from 0, we can rescale x_0 such that $q^{n'}\Sigma^n q^n = 1$. From

$$q^{n'}q^n = x_0' O_1^{n'} O_1^n x_0 = x_0' x_0 =: \alpha^2 \tag{5.2.26}$$

and

$$\frac{q^{n'}}{\alpha}\Sigma^n \frac{q^n}{\alpha} = \frac{1}{\alpha^2} \leq \sigma_{sup} \tag{5.2.27}$$

we get, that

$$x_0' x_0 = \alpha^2 \geq \frac{1}{\sigma_{sup}} > 0. \tag{5.2.28}$$

Hence

$$\begin{aligned} \nu_s^n &\geq q^{n'}\Gamma^n q^n = x_0' O_1^{n'}\Gamma^n O_1^n x_0 = \\ &= x_0'\Omega_1^n x_0 \geq \mu_s^n x_0' x_0 \geq \frac{\mu_s^n}{\sigma_{sup}}, \end{aligned} \tag{5.2.29}$$

where the last expression diverges as $n \to \infty$. ■

Now we can show that the space spanned by the population aggregates $\psi_t^{n,\cdot}$ converges to the factor space (see also Stock and Watson (1998) and Forni *et al.* (2005a)).

Theorem 5.2.1. *Suppose Assumptions 1.1, 2.1 and 5.2 hold. Considering the projection equations,*

$$\psi_t^{n,\cdot} = \text{proj}(\psi_t^{n,\cdot}|f_{1t}, \ldots, f_{st}) + \gamma_t^{n,\cdot} \tag{5.2.30}$$

and

$$f_t = \text{proj}(f_t|\psi_{1t}^{n,\cdot}, \ldots, \psi_{st}^{n,\cdot}) + \eta_t^{n,\cdot}, \tag{5.2.31}$$

where . replaces PCA and GPCA, then $\gamma_t^{n,\cdot}$ and $\eta_t^{n,\cdot}$ converge to 0 in mean square as $n \to \infty$. If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then the same holds true for . replaced by FHLR.

Proof. To start with, we will show that the aggregates $\psi_t^{n,\cdot}$ converge to the space spanned by the factors f_t as n tends to infinity. All aggregates $\psi_t^{n,\cdot}$ are of the form

$$\psi_t^{n,\cdot} = B^{n,\cdot} y_t^n = B^{n,\cdot}\chi_t^n + B^{n,\cdot}u_t^n \tag{5.2.32}$$

where the terms on the right hand side are orthogonal, since $\chi_t^n \perp u_t^n$ and where obviously $B^{n,\cdot}\chi_t^n \in \text{span}(f_t)$. Hence $B^{n,\cdot}u_t^n$ are the perpendiculars of the orthogonal projection of $\psi_t^{n,\cdot}$ onto the static factor space and we will have to show that $B^{n,\cdot}u_t^n$ converges to 0 in mean square as $n \to \infty$ for either weight matrix $B^{n,\cdot}$ presented above. Since by Assumption 2.1 (u_t^n) is weakly correlated, we thus have to show that the rows of either weight matrix $B^{n,\cdot}$ are DAS (albeit

static), since then the second term on the right hand side of (5.2.32) vanishes as $n \to \infty$.

Static PCA
We have

$$\psi_{jt}^{n,PCA} = \frac{1}{\sqrt{\mu_j^n}} o_j^{n\prime} y_t^n,$$

$j = 1, \ldots, s$, where μ_j^n is the j-th largest eigenvalue of Γ_y^n and o_j^n is a corresponding normalized eigenvector. From Corollary A.1.1 we know that $\mu_j^n \geq \mu_{\chi,j}^n$ and by Assumption 5.2, $\mu_{\chi,j}^n \to \infty$, $j = 1, \ldots, s$, implying that $\mu_j^n \to \infty$, $j = 1, \ldots, s$. Since furthermore $o_j^{n\prime} o_j^n = 1$ holds, we obtain that $\frac{1}{\mu_j^n} o_j^{n\prime} o_j^n = \frac{1}{\mu_j^n}$ converges to 0 for $j = 1, \ldots, s$ as $n \to \infty$ and hence $\frac{1}{\sqrt{\mu_j^n}} o_j^{n\prime}$, $j = 1, \ldots, s$, is a DAS.

Generalized Static PCA
We have

$$\psi_{jt}^{n,GPCA} = \frac{1}{\sqrt{\nu_j^n}} v_j^{n\prime} y_t^n,$$

$j = 1, \ldots, s$, where ν_j^n is the j-th largest generalized eigenvalue of (Γ_y^n, Γ_u^n) and v_j^n is a corresponding generalized eigenvector. First recall, that the eigenvalues of Γ_u^n are bounded from above as $n \to \infty$, since

$$\begin{aligned} \Gamma_u^n &= \int_{-\pi}^{\pi} f_u^n(\lambda) d\lambda \leq \int_{-\pi}^{\pi} \omega_{u,1}^n(\lambda) I_n d\lambda \\ &\leq 2\pi \underbrace{\operatorname{ess\,sup}_n(\omega_{u,1}^n)}_{\bar{\omega}_{u1}} I_n, \end{aligned}$$

and that the eigenvalues of Γ_u^n are bounded away from 0. Hence we can apply Lemma 5.2.1 implying that ν_j^n, $j = 1, \ldots, s$ diverges as $n \to \infty$ whereas ν_j^n, $j = s+1, \ldots, n$ is bounded. Furthermore v_j^n, $j = 1, \ldots, n$ is bounded in modulus, since $1 = v_j^{n\prime} \Gamma_u^n v_j^n \geq \mu_{u,n}^n v_j^{n\prime} v_j^n$ and $\mu_{u,n}^n$ is bounded away from 0. Hence $\frac{1}{\sqrt{\nu_j^n}} v_j^{n\prime}$, $j = 1, \ldots, s$, is a DAS.

Generalized Static PCA based on dynamic PCA second moments - FHLR
We have

$$\psi_{jt}^{n,FHLR} = \frac{1}{\sqrt{\tilde{\nu}_j^n + 1}} w_j^{n\prime} y_t^n,$$

$j = 1, \ldots, s$, where $\tilde{\nu}_j^n$ is the j-th largest generalized eigenvalue of $(\tilde{\Gamma}_\chi^n, \tilde{\Gamma}_u^n)$ and w_j^n is a corresponding generalized eigenvector. First observe that the eigenvalues of $\tilde{\Gamma}_u^n$ are bounded from above as $n \to \infty$, since for any $(n \times 1)$ unit vector v

$$\begin{aligned} v'\tilde{\Gamma}_u^n v &= v' \left(\int_{-\pi}^{\pi} \tilde{f}_u^n(\lambda) d\lambda \right) v = \int_{-\pi}^{\pi} v' \tilde{f}_u^n(\lambda) v d\lambda \\ &\leq \int_{-\pi}^{\pi} \omega_{u1}^n(\lambda) d\lambda \leq 2\pi \bar{\omega}_{u1}, \end{aligned}$$

and that the eigenvalues of $\tilde{\Gamma}_u^n$ are bounded away from 0 by assumption. Second, observing that

$$A = \Gamma_u^n - \tilde{\Gamma}_u^n = \tilde{\Gamma}_\chi^n - \Gamma_\chi^n$$

and that for any $(n \times 1)$ unit vector v, since $0 < v'\tilde{\Gamma}_u^n v \leq 2\pi\bar{\omega}_{u1}$ and $0 < v'\Gamma_u^n v \leq 2\pi\bar{\omega}_{u1}$,

$$|v'Av| = |v'\Gamma_u^n v - v'\tilde{\Gamma}_u^n v| \leq 2\pi\bar{\omega}_{u1}$$

implies that

$$\tilde{\Gamma}_\chi^n + 2\pi\bar{\omega}_{u1} I_n = \Gamma_\chi^n + (A + 2\pi\bar{\omega}_{u1} I_n),$$

where the matrix in brackets is positive semi-definite. Applying corollary A.1.1 implies that the eigenvalues of the left hand side are larger than the eigenvalues of Γ_χ^n, hence the s largest eigenvalues of $\tilde{\Gamma}_\chi^n$ diverge as $n \to \infty$. We can thus apply Lemma 5.2.1, implying that the s largest generalized eigenvalues of $(\tilde{\Gamma}_\chi^n, \tilde{\Gamma}_u^n)$, i.e. $\tilde{\nu}_j^n$, $j = 1, \ldots, s$, diverge as $n \to \infty$. Furthermore w_j^n, $j = 1, \ldots, n$ is bounded in modulus, since $1 = v_j^{n\prime}\tilde{\Gamma}_u^n w_j^n$ and since the eigenvalues of $\tilde{\Gamma}_u^n$ are bounded away from 0. Hence $\frac{1}{\sqrt{\tilde{\nu}_j^n + 1}} w_j^{n\prime}$, $j = 1, \ldots, s$, is a DAS.

For the orthogonal projections of f_t onto the spaces spanned by $\psi_t^{n,\cdot}$, we have that, if

$$\begin{aligned} \psi_t^{n,\cdot} &= \text{proj}(\psi_t^{n,\cdot} | f_{1t}, \ldots, f_{st}) + \gamma_t^{n,\cdot} \\ &= A^{n,\cdot} f_t + \gamma_t^{n,\cdot} \end{aligned} \tag{5.2.33}$$

then

$$\begin{aligned} f_t &= \text{proj}(f_t | \psi_{1t}^{n,\cdot}, \ldots, \psi_{st}^{n,\cdot}) + \eta_t^{n,\cdot} \\ &= A^{n,\cdot\prime} \psi_t^{n,\cdot} + \eta_t^{n,\cdot}. \end{aligned} \tag{5.2.34}$$

Taking the covariances of (5.2.33) and (5.2.34) we obtain

$$I_s = A^{n,\cdot} A^{n,\cdot\prime} + \Gamma_\gamma^{n,\cdot} = A^{n,\cdot\prime} A^{n,\cdot} + \Gamma_\eta^{n,\cdot}, \tag{5.2.35}$$

where $\Gamma_\gamma^{n,\cdot}$ and $\Gamma_\eta^{n,\cdot}$ are the covariances of $\gamma_t^{n,\cdot}$ and $\eta_t^{n,\cdot}$ respectively. By taking the trace on both sides we get

$$\text{tr}(\Gamma_\gamma^{n,\cdot}) = \text{tr}(\Gamma_\eta^{n,\cdot}), \tag{5.2.36}$$

and since the left hand side vanishes as $n \to \infty$, the result follows. ■

Hence for each $t \in \mathbb{Z}$, the Hilbert space spanned by the scalar elements of $\psi_t^{n,\cdot}$ "converges" to the static factor space $\overline{\text{span}}(f_{jt} | j = 1, \ldots, s)$. Moreover since all $\psi_t^{n,\cdot}$ result from static transformations of y_t^n and have covariance equal to I_s, the last result may be restated in that, there exist orthogonal $(s \times s)$-dimensional matrices $R^{n,\cdot}$ with $R^{n,\cdot\prime} R^{n,\cdot} = R^{n,\cdot} R^{n,\cdot\prime} = I_s$, independent of t, such that

$$\|\psi_t^{n,\cdot} - R^{n,\cdot} f_t\|^2$$

converges to zero as $n \to \infty$.

The next lemma shows that any projections onto the space spanned by $\psi_t^{n,\cdot}$ will converge to the corresponding projections onto the static factor space. Notice the correspondence to the definition of Cauchy sequences of spaces (see definition 3.1.3, where we introduced a similar principle), but as distinct to definition 3.1.3 here we do not consider spaces spanned by a whole process ($t \in \mathbb{Z}$) but spaces spanned by single observations.

Lemma 5.2.2. *Let $\mathbb{H}$ denote the Hilbert space of 0-mean, square integrable random variables, $v = (v_1, \ldots, v_s)$ denote an s-tuple of independent elements of $\mathbb{H}$ and let $K = \text{span}(v_1, \ldots, v_s)$. Assume that $(v^n) = (v_j^n | j = 1, \ldots, s, n \in \mathbb{N})$ is a sequence of s-tuples of orthonormal elements of $\mathbb{H}$, such that $v_j^n - \text{proj}(v_j^n|K)$ converges to 0 in mean square for $j = 1, \ldots, s$ as $n \to \infty$ and let $K^n = \text{span}(v_1^n, \ldots, v_s^n)$. Then for any $x \in \mathbb{H}$, $\text{proj}(x|K^n) \to \text{proj}(x|K)$ in mean square as $n \to \infty$.*

Proof. For a proof see Forni *et al.* (2005a). ■

As a consequence we have the following lemma (see Stock and Watson (1998) and Forni *et al.* (2005a)):

Lemma 5.2.3. *Suppose Assumptions 1.1, 2.1 and 5.2 hold. Then every component of $\tilde{\chi}_t^{n,PCA}$ and $\tilde{\chi}_t^{n,GPCA}$ respectively converges to the corresponding component of χ_t^n in mean square as $n \to \infty$. If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then every component of $\tilde{\chi}_t^{n,FHLR}$ converges to the corresponding component of χ_t^n in mean square as $n \to \infty$.*

Proof. The result follows immediately from Lemma 5.2.2, since $\tilde{\chi}_t^{n,PCA}$ and $\tilde{\chi}_t^{n,GPCA}$ are the projections of y_t^n onto the space spanned by $\psi_t^{n,PCA}$ and $\psi_t^{n,GPCA}$ respectively and since $\tilde{\chi}_t^{n,FHLR}$ is the projection of χ_t^n onto $\psi_t^{n,FHLR}$. ■

So far, we have shown consistency for the population counter parts of the estimates as $n \to \infty$. Next, we are going to show, that the estimators themselves converge to their population counterparts as $T \to \infty$ and hence to the GDFM variables as both n and T go to infinity. Notice, that in the theorem and the corollaries below, T is not chosen independently of n. This is because for n fixed and T going to infinity under the assumptions imposed $\hat{\Gamma}_y^n$ converges to Γ_y^n in probability (and analogous statements hold true for $\hat{\Gamma}_\chi^n$ and $\hat{\Gamma}_u^n$ respectively). Here the asymptotics considered are sequential: first T goes to infinity for fixed n and then n goes to infinity. The proof follows the proof of Proposition 4.1 in Forni *et al.* (2005a).

Theorem 5.2.2. *Suppose Assumptions 1.1, 2.1, 5.3 and 5.2 hold. Then, for any $i \in \mathbb{N}$, $\varepsilon > 0$ and $\eta > 0$, there exist $N_0 = N_0(\varepsilon, \eta) \in \mathbb{N}$, $N_0 \geq i$ and $T_0 = T_0(n, \varepsilon, \eta)$, such that for all $n \geq N_0$ and all $T \geq T_0$,*

$$P(|\hat{\chi}_{it}^{n,PCA} - \chi_{it}| > \varepsilon) \leq \eta. \tag{5.2.37}$$

If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then analogously for for all $n \geq N_0$ and all $T \geq T_0$,

$$P(|\hat{\chi}_{it}^{n,PCA} - \chi_{it}| > \varepsilon) \leq \eta, \text{ and} \tag{5.2.38}$$

$$P(|\hat{\chi}_{it}^{n,PCA} - \chi_{it}| > \varepsilon) \leq \eta. \tag{5.2.39}$$

Proof. First for ease of notation, notice that each $\hat{\chi}_t^{n,\cdot}$ considered is of the form $\hat{\chi}_t^{n,\cdot} = \hat{K}^{n,\cdot} y_t^n$ (e.g. $\hat{K}^{n,PCA} = \hat{O}_1^n \hat{O}_1^{n\prime}$), denote the population analogues by $\tilde{\chi}_t^{n,\cdot} = \tilde{K}^{n,\cdot} y_t^n$. Each $\hat{K}^{n,\cdot}$ is a product of eigenvectors (PCA) or of eigenvectors and covariance matrices (GPCA, FHLR). Recall, that the sample covariance $\hat{\Gamma}_y^n$ from (5.2.5) and the matrices $\hat{\Gamma}_\chi^n$ and $\hat{\Gamma}_u^n$ from (5.2.17) converge in probability to Γ_y^n, $\tilde{\Gamma}_\chi^n$ and $\tilde{\Gamma}_u^n$, respectively as $T \to \infty$. Furthermore, since eigenvectors are continuous functions of the matrix elements, applying Slutsky's Theorem we have that given $n \in \mathbb{N}$, $\varepsilon > 0$ and $\eta > 0$, there exists $T_1 = T_1(n, \varepsilon, \eta)$, such that for $T \geq T_1$,

$$P\left(\sum_{j=1}^{n} |\hat{K}_{ij}^n - \tilde{K}_{ij}^n| > \varepsilon\right) < \eta. \tag{5.2.40}$$

Next, given n and $\eta > 0$, let $M(n, \eta) > 0$ such that

$$P\left(\max_{j=1,\ldots,n} |y_{jt}| \geq M(n,\eta)\right) < \eta.$$

Then we may write

$$\begin{aligned} P\left(\sum_{j=1}^{n} |(\hat{K}_{ij}^n - \tilde{K}_{ij}^n) y_{jt}| > \varepsilon\right) \leq \\ P\left(\sum_{j=1}^{n} |(\hat{K}_{ij}^n - \tilde{K}_{ij}^n) M(n,\eta/2)| > \varepsilon \text{ and } \max_{j=1,\ldots,n} |y_{jt}| < M(n,\eta/2)\right) \\ + \quad P\left(\max_{j=1,\ldots,n} |y_{jt}| \geq M(n,\eta/2)\right). \end{aligned}$$

Hence given $n \in \mathbb{N}$, $\varepsilon > 0$ and $\eta > 0$, there exists $T_2 = T_1(n, \varepsilon/M(n,\eta/2), \eta/2)$, such that for $T \geq T_2$,

$$P\left(|\hat{\chi}_{it}^{n,\cdot} - \tilde{\chi}_{it}^{n,\cdot}| > \varepsilon\right) = P\left(\sum_{j=1}^{n} |(\hat{K}_{ij}^n - \tilde{K}_{ij}^n) y_{jt}| > \varepsilon\right) < \eta. \tag{5.2.41}$$

Finally, Lemma 5.2.3 implies that given $\varepsilon > 0$ and $\eta > 0$, there exists $N_1(\varepsilon, \eta)$, such that for $n \geq N_1(\varepsilon, \eta)$,

$$P\left(|\chi_{it} - \tilde{\chi}_{it}^{n,\cdot}| > \varepsilon\right) < \eta. \tag{5.2.42}$$

Hence, defining $N_0(\varepsilon, \eta) = N_1(\varepsilon/2, \eta/2)$ and $T_0(n, \varepsilon, \eta) = T_2(n, \varepsilon/2, \eta/2)$, we have

$$P\left(|\chi_{it} - \hat{\chi}_{it}^{n,\cdot}| > \varepsilon\right) \leq P\left(|\chi_{it} - \tilde{\chi}_{it}^{n,\cdot}| > \varepsilon/2\right) + P\left(|\hat{\chi}_{it}^{n,\cdot} - \tilde{\chi}_{it}^{n,\cdot}| > \varepsilon/2\right) < \eta. \tag{5.2.43}$$

■

Quite obviously, also the static factor estimates $\hat{\psi}_t^{n,\cdot}$ converge in probability to their population counter parts $\psi_t^{n,\cdot}$ as $T \to \infty$.

Lemma 5.2.4. *Suppose Assumptions 1.1, 2.1, 5.3 and 5.2 hold. Then $\hat{\psi}_t^{n,PCA}$ converges in probability to $\psi_t^{n,PCA}$ as $T \to \infty$. If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then analogously $\hat{\psi}_t^{n,GPCA}$ and $\hat{\psi}_t^{n,FHLR}$ converge in probability to $\psi_t^{n,GPCA}$ and $\psi_t^{n,FHLR}$ respectively, as $T \to \infty$.*

Proof. The result follows immediately, since $\hat{\Gamma}_y^n$ from (5.2.5) and the matrices $\hat{\Gamma}_\chi^n$ and $\hat{\Gamma}_u^n$ from (5.2.17) converge in probability to Γ_y^n, $\tilde{\Gamma}_\chi^n$ and $\tilde{\Gamma}_u^n$, respectively as $T \to \infty$ and since eigenvalues and suitably normalized eigenvectors are continuous functions of the matrix elements. ■

Summing up the results of the last lemma and Theorem 5.2.1, we may state the following corollary, that claims that the static factors can be consistently estimated up to an orthogonal static transformation. This is of particular importance regarding the prediction problem, since as we have seen before, the prediction of the latent variable part can be reduced to the prediction of the static factors (see Section 4).

Corollary 5.2.1. *Suppose Assumptions 1.1, 2.1, 5.3 and 5.2 hold. Then, for any $i \in \mathbb{N}$, $\varepsilon > 0$ and $\eta > 0$, there exist $N_0 = N_0(\varepsilon, \eta) \in \mathbb{N}$, $N_0 \geq i$ and $T_0 = T_0(n, \varepsilon, \eta)$, such that for all $n \geq N_0$ and all $T \geq T_0$, there exist orthogonal $(s \times s)$ matrices R_1^n such that*

$$P(|\hat{\psi}_t^{n,PCA} - R_1^{n'} f_t| > \varepsilon) \leq \eta. \tag{5.2.44}$$

If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then analogously there exist regular $(s \times s)$ matrices R_2^n and R_3^n such that

$$P(|\hat{\psi}_t^{n,GPCA} - R_2^{n'} f_t| > \varepsilon) \leq \eta, \tag{5.2.45}$$

$$P(|\hat{\psi}_t^{n,FHLR} - R_3^{n'} f_t| > \varepsilon) \leq \eta. \tag{5.2.46}$$

Proof. As can be easily verified, the proof is essentially the same as the proof of Theorem 5.2.1 with $R_j^{n'} f_t$ in place of χ_t, $\psi_t^{n,\cdot}$ in place of $\tilde{\chi}_t^{n,\cdot}$ and $\hat{\psi}_t^{n,\cdot}$ in place of $\hat{\chi}_t^{n,\cdot}$. ■

Turning to the static factor loading matrix $\bar{\Lambda}^n$, the next corollary is an immediate consequence of Theorem 5.2.1 and Corollary 5.2.1. Here let $\hat{\Lambda}^{n,\cdot}$ denote the factor loadings estimate corresponding to $\psi_t^{n,\cdot}$ (e.g. $\hat{\Lambda}^{n,PCA} = \hat{O}_1^n \hat{\Omega}_1^{n\,1/2}$), and let $\bar{\Lambda}_i^n$ and $\hat{\Lambda}_i^{n,\cdot}$, $i = 1, \ldots, n$ denote the i-th row of the respective matrices.

Corollary 5.2.2. *Suppose Assumptions 1.1, 2.1, 5.3 and 5.2 hold. Then, for any $i \in \mathbb{N}$, $\varepsilon > 0$ and $\eta > 0$, there exist $N_0 = N_0(\varepsilon, \eta) \in \mathbb{N}$, $N_0 \geq i$ and $T_0 = T_0(n, \varepsilon, \eta)$, such that for all $n \geq N_0$ and all $T \geq T_0$, there exist orthogonal $(s \times s)$ matrices R_1^n such that*

$$P(\|\hat{\Lambda}_i^{n,PCA} - \bar{\Lambda}_i^n R_1^n\| > \varepsilon) \leq \eta. \tag{5.2.47}$$

If furthermore Assumption 5.1 holds and if the smallest eigenvalue of $\tilde{\Gamma}_u^n$ is bounded away from 0, then analogously there exist orthogonal $(s \times s)$ matrices R_2^n and R_3^n such that for any $i \in \mathbb{N}$,

$$P(\|\hat{\Lambda}_i^{n,GPCA} - \bar{\Lambda}_i^n R_2^n\| > \varepsilon) \leq \eta, \tag{5.2.48}$$

$$P(\|\hat{\Lambda}_i^{n,FHLR} - \bar{\Lambda}_i^n R_3^n\| > \varepsilon) \leq \eta. \tag{5.2.49}$$

In the sequel we are going to show that indeed consistency of the static factors can be achieved along any path (n, T) with both n and T tending to infinity (as has been done under different assumptions in Stock and Watson (1998) and Bai and Ng (2002)). Here, we will largely follow the proofs given in Doz *et al.* (2007) and present the results exemplarily for the PCA estimates.

In order to simplify the derivations, we will fix particular rotations R^n of the factors and the factor loadings in correspondence to the PCA estimates. Hence consider the canonical representation of $\Gamma_\chi^n = \bar{\Lambda}^n \bar{\Lambda}^{n\prime}$,

$$\Gamma_\chi^n = O_\chi^n \Omega_\chi^n O_\chi^{n\prime},$$

where Ω_χ^n is the diagonal matrix whose diagonal entries are the s non-zero eigenvalues of Γ_χ^n and O_χ^n is the matrix of corresponding normalized eigenvectors. Then $\bar{\Lambda}^n$ can be written as

$$\bar{\Lambda}^n = O_\chi^n \Omega_\chi^{n^{1/2}} R^{n\prime},$$

where R^n is the orthogonal $(s \times s)$ matrix of right-singular vectors of $\bar{\Lambda}^n$ (or equivalently the matrix of normalized eigenvalues of $\bar{\Lambda}^{n\prime}\bar{\Lambda}^n$) and is uniquely determined up to a sign change of its columns[3]. Hence defining

$$\tilde{\Lambda}^n = \bar{\Lambda}^n R^n, \quad \tilde{f}_t^n = R^{n\prime} f_t,$$

we have identified $\tilde{\Lambda}^n$ and $\tilde{f}_t^n$ up to a sign change of the columns of $\tilde{\Lambda}^n$ and the entries of $\tilde{f}_t^n$ respectively. Notice, that $\tilde{\chi}_t^n = \tilde{\Lambda}^n \tilde{f}_t^n = \chi_t^n$, but that since R^n is depending on n, $\tilde{\Lambda}^n$ is not stacked anymore and $\tilde{f}_t^n$ is a function of n. Hence, $\tilde{\Lambda}^n$ and $\tilde{f}_t^n$ do not satisfy the conditions of the factor model representation (3.3.63), however they may serve as facilities to derive the desired results.

We will need the following preliminary results (see Doz *et al.* (2007), Lemma 2 and Lemma 3). Let $\omega_1(A)$ denote the largest eigenvalue of a square matrix A and $\|C\| = (\omega_1(C'C))^{1/2}$ denote the spectral norm of an arbitrary matrix C. We use $O(.)$ to indicate that the order relation holds a.s., whereas $O_p(.)$ means that the order relation only holds in probability.

Lemma 5.2.5. *Under Assumptions 1.1, 2.1, 3.1, 5.2 and 5.3, the following properties hold as* $n, T \to \infty$*:*

$$(i) \quad \|\hat{\Gamma}_y^n - \Gamma_\chi^n\| = O(1) + O_p\left(\tfrac{n}{\sqrt{T}}\right)$$
$$(ii) \quad \|\hat{\Omega}_1^n - \Omega_\chi^n\| = O(1) + O_p\left(\tfrac{n}{\sqrt{T}}\right)$$

Proof. (i) We may write

$$\|\hat{\Gamma}_y^n - \Gamma_\chi^n\| \leq \|\hat{\Gamma}_y^n - \Gamma_y^n\| + \|\Gamma_y^n - \Gamma_\chi^n\|,$$

then the second term on the right equals $\|\Gamma_u^n\| = \mu_{u,1}^n$ and is thus bounded for all $n \in \mathbb{N}$, i.e. $O(1)$. For the first term consider

$$\|\hat{\Gamma}_y^n - \Gamma_y^n\|^2 \leq \mathrm{tr}(\hat{\Gamma}_y^n - \Gamma_y^n)^2 = \sum_{i=1}^n \sum_{j=1}^n (\hat{\gamma}_{y,ij} - \gamma_{y,ij})^2.$$

[3] R^n is unique (up to sign changes) since the eigenvalues of Γ_χ^n are supposed to be distinct.

Taking expectations,

$$\mathbb{E}\,\|\hat{\Gamma}_y^n - \Gamma_y^n\|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(\hat{\gamma}_{y,ij} - \gamma_{y,ij})^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} V(\hat{\gamma}_{y,ij}).$$

Under Assumption 5.3 (i), $V(\hat{\gamma}_{y,ij}) = O_p(1/T)$, thus $\mathbb{E}\,\|\hat{\Gamma}_y^n - \Gamma_y^n\|^2 = O_p(n^2/T)$ and the result follows from Markov's inequality.

(ii) We will employ an inequality that is due Weil: let A and E denote symmetric $(n \times n)$ matrices and let $\omega_j(.)$ denote the eigenvalues in descending order of magnitude, then

$$|\omega_j(A+E) - \omega_j(A)| \leq \sqrt{\omega_1(E^2)} \leq \sqrt{\mathrm{tr}(E^2)}.$$

Hence we obtain that

$$|\hat{\mu}_j^n - \mu_{\chi,j}^n| \leq \|\hat{\Gamma}_y^n - \Gamma_\chi^n\|,$$

which yields the result with (i). ■

Lemma 5.2.6. *Under Assumptions 1.1, 2.1, 3.1, 5.2 and 5.3, then there exist orthogonal* $(s \times s)$ *matrices* R^n *such that as* $n, T \to \infty$,

$$\hat{\Omega}_1^{n^{-1/2}}\hat{O}_1^{n'}\Lambda^n R^n = I_s + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof. Define $\hat{A}^n = \hat{\Omega}_1^{n^{-1/2}}\hat{O}_1^{n'}\tilde{\Lambda}^n$. First we are going to show that the off-diagonal elements of $\hat{A}^n$ go to zero.

Observing that

$$\hat{O}_1^n = \hat{\Gamma}_y^n\hat{O}_1^n\hat{\Omega}_1^{n^{-1}},$$

we may write

$$\hat{A}^n = \hat{\Omega}_1^{n^{-3/2}}\hat{O}_1^{n'}(\hat{\Gamma}_y^n - \Gamma_\chi^n)\tilde{\Lambda}^n + \hat{\Omega}_1^{n^{-3/2}}\hat{O}_1^{n'}\Gamma_\chi^n\tilde{\Lambda}^n. \tag{5.2.50}$$

Since

$$\tilde{\Lambda}^n = O_\chi^n\Omega_\chi^{n^{1/2}},$$

the first term on the right of (5.2.50) is

$$\hat{\Omega}_1^{n^{-3/2}}\hat{O}_1^{n'}(\hat{\Gamma}_y^n - \Gamma_\chi^n)O_\chi^n\Omega_\chi^{n^{1/2}},$$

and since $\hat{O}_1^{n'}\hat{O}_1^n = I_s$ and $O_\chi^{n'}O_\chi^n = I_s$,

$$\|\hat{\Omega}_1^{n^{-3/2}}\hat{O}_1^{n'}(\hat{\Gamma}_y^n - \Gamma_\chi^n)O_\chi^n\Omega_\chi^{n^{1/2}}\| \leq \mu_{\chi,1}^{n^{1/2}}\,\hat{\mu}_1^{n^{-3/2}}\|\hat{\Gamma}_y^n - \Gamma_\chi^n\|.$$

Now, $\mu_{\chi,1}^{n^{1/2}}\hat{\mu}_1^{n^{-3/2}}$ is $O_p(\frac{1}{n})$, hence it follows with Lemma 5.2.5(i) that the last expression is $O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$.

For the second term on the right of (5.2.50), notice that

$$\Gamma_\chi^n \tilde{\Lambda}^n = \Gamma_\chi^n O_\chi^n \Omega_\chi^{n^{1/2}} = O_\chi^n \Omega_\chi^{n^{3/2}},$$

such that the second term on the right of (5.2.50) is equal to

$$\hat{\Omega}_1^{n^{-3/2}} \hat{O}_1^{n'} O_\chi^n \Omega_\chi^{n^{3/2}}.$$

Furthermore, $\hat{A}^n$ is equal to

$$\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n'} O_\chi^n \Omega_\chi^{n^{1/2}},$$

hence it then follows, that

$$\hat{A}^n = O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}) + \hat{\Omega}_1^{n^{-1}} \hat{A}^n \Omega_\chi^n$$

and using Lemma 5.2.5(ii)

$$\hat{A}^n = O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}) + \Omega_\chi^{n^{-1}} \hat{A}^n \Omega_\chi^n.$$

Hence denoting the i, j element of $\hat{A}^n$ by $\hat{a}_{ij}^n$, we have for $i \neq j$,

$$\hat{a}_{ij}^n = \frac{\mu_{\chi,j}^n}{\mu_{\chi,i}^n} \hat{a}_{ij}^n + O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}),$$

and since we assumed that all eigenvalues are distinct, it follows that

$$\hat{a}_{ij}^n = O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}).$$

Next we want to show that the diagonal entries of $\hat{A}^n$ converge to 1. Therefore consider,

$$\begin{aligned} \hat{\Omega}_1^n &= \hat{O}_1^{n'} \hat{\Gamma}_y^n \hat{O}_1^n = \hat{O}_1^{n'} \Gamma_\chi^n \hat{O}_1^n + O_p(1) + O_p(\frac{n}{\sqrt{T}}) \\ &= \hat{O}_1^{n'} \tilde{\Lambda}^n \tilde{\Lambda}^{n'} \hat{O}_1^n + O_p(1) + O_p(\frac{n}{\sqrt{T}}), \end{aligned}$$

where we used Lemma 5.2.5 together with the fact that $\hat{O}_1^{n'} \hat{O}_1^n = I_s$. Hence pre and post multiplying both sides of the last expression by $\hat{\Omega}_1^{n^{-1/2}}$ yields

$$I_s = \hat{A}^n \hat{A}^{n'} + O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$$

and thus for $i = 1, \ldots, n$

$$1 = \sum_{j=1}^n \hat{a}_{ij}^{n^2} O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$$

and since $\hat{a}_{ij}^n = O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$ it follows, that

$$\hat{a}_{ii}^{n^2} = 1 + O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}).$$

Finally, since R^n has been defined up to a sign change of its columns, there always exists R^n, such that

$$\hat{a}_{ii}^n = 1 + O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}}).$$

■

We can now turn to the main result (for a reference, see Doz *et al.* (2007), Proposition 2).

Theorem 5.2.3. *Under Assumptions 1.1, 2.1, 3.1, 5.2 and 5.3, then there exist orthogonal* $(s \times s)$ *matrices* R^n *such that,*

$$\begin{array}{ll} (i) & \operatorname{plim}_{(n,T)\to\infty} \hat{\psi}_t^{n,PCA} - R^{n\prime} f_t = 0, \\ (ii) & \operatorname{plim}_{(n,T)\to\infty} \hat{\Lambda}_i^{n,PCA} - \bar{\Lambda}_i^n R^n = 0 \text{ for any } i \in \mathbb{N}. \end{array}$$

Proof. In the following it will be assumed that R^n is defined as described above.
(i) We may write

$$\begin{aligned} \hat{\psi}_t^{n,PCA} - \tilde{f}_t^n &= \hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} y_t^n - \tilde{f}_t^n \\ &= (\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \tilde{\Lambda}^n - I_s) \tilde{f}_t^n + \hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} u_t^n. \end{aligned} \tag{5.2.51}$$

For the first term of (5.2.51), we know from Lemma (5.2.6) that $(\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \tilde{\Lambda}^n - I_s) = O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$, and since $f_t = O(1)$ then the first term is $O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{T}})$.

Turning to the second term of (5.2.51), we have

$$\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} u_t^n u_t^{n\prime} \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}} \leq \frac{1}{\hat{\mu}_s^n} \hat{O}_1^{n\prime} u_t^n u_t^{n\prime} \hat{O}_1^n,$$

hence taking expectations,

$$\frac{1}{\hat{\mu}_s^n} \hat{O}_1^{n\prime} \Gamma_u^n \hat{O}_1^n \leq \frac{\mu_{u,1}^n}{\hat{\mu}_s^n} = O_p\left(\frac{1}{n}\right).$$

Thus it follows from Markov's inequality that

$$\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} u_t^n = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and thus we have

$$\hat{\psi}_t^{n,PCA} - \tilde{f}_t^n = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

(ii) Let e_i denote the i-th vector of the canonical basis of $\mathbb{R}^n$, hence e.g. $\hat{\Lambda}_i^{n,PCA} = e_i' \hat{\Lambda}^{n,PCA}$. Then using the equalities $\hat{\Lambda}^{n,PCA} = \hat{O}_1^n \hat{\Omega}_1^{n^{1/2}} = \hat{\Gamma}_y^n \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}}$ and $\tilde{\Lambda}^n) = \Gamma_\chi^n O_\chi^n \Omega_\chi^{n^{-1/2}}$, we may write

$$e_i'(\hat{\Lambda}^{n,PCA} - \tilde{\Lambda}^n) = e_i' \left(\hat{\Gamma}_y^n - \Gamma_\chi^n\right) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}} + e_i' \Gamma_\chi^n \left(\hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}} - O_\chi^n \Omega_\chi^{n^{1/2}}\right). \tag{5.2.52}$$

For the first term Lemma 5.2.5 implies

$$e_i' \left(\hat{\Gamma}_y^n - \Gamma_\chi^n\right) = O_p\left(\frac{\sqrt{n}}{\sqrt{T}}\right),$$

hence we obtain that

$$\|e_i' \left(\hat{\Gamma}_y^n - \Gamma_\chi^n\right) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}}\| \leq \hat{\mu}_s^{n^{-1/2}} \|e_i' \left(\hat{\Gamma}_y^n - \Gamma_\chi^n\right)\| \|\hat{O}_1^n\|$$

is $O_p\left(\frac{1}{\sqrt{T}}\right)$.

The second term of (5.2.52) may be written as

$$e_i'\tilde{\Lambda}^n\left(\tilde{\Lambda}^{n\prime}\hat{O}_1^n\hat{\Omega}_1^{n^{-1/2}} - \tilde{\Lambda}^{n\prime}O_\chi^n\Omega_\chi^n\right) = e_i'\tilde{\Lambda}^n\left(\tilde{\Lambda}^{n\prime}\hat{O}_1^n\hat{\Omega}_1^{n^{-1/2}} - I\right),$$

which is using Lemma 5.2.6 $O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$. Hence the result follows. ■

Chapter 6

Estimation of the ARMA model for the static factor

So far it has been shown that under the assumptions imposed the static factors (f_t) allow of an ARMA representation and that they can be consistently estimated (up to static rotations) by one of the methods described in the last section. The obvious next step is thus the estimation of the ARMA model for the static factors. Throughout this chapter we will assume that all relevant integer valued parameters, i.e. the Kronecker indices $\alpha = (r_1, \ldots, r_s)$ specifying the echelon ARMA representation of (f_t), and thus the ARMA orders P and Q, as well as the dynamic factor dimension q, are known. Hence estimation effects the coefficient matrices of the ARMA model under consideration.

The problem of multivariate ARMA estimation has been extensively studied and a variety of different methods is available. Important approaches are among others the *(quasi-) maximum likelihood* (ML) method (see e.g. Hannan and Deistler (1988), Reinsel (1993) or Lütkepohl (2005)), that is asymptotically efficient (see Hillmer and Tiao (1979)), the *prediction error* (PE) method (see e.g. Ljung (1987) or Findley *et al.* (2004)) and various *(iterative) regression* methods (see e.g. Hannan and Rissanen (1982) or Koreisha and Pukkila (1989)). However, in general multivariate ARMA estimation is still considered difficult compared to AR or ARX[1] estimation. Reasons for that are the comparatively more difficult specification of ARMA models and the fact that the (most widely used) ML and PE methods require numerical optimization, which is especially troublesome if the true model is nearly not-identifiable and in the presence of multiple local optima. Apart from these known problems, here we will have to deal with two additional difficulties: first, f_t may be dynamically singular (i.e. its $(s \times s)$ spectral density is of rank q and q may be smaller than s), and second f_t is not observed but has to be estimated in a prior step.

In the GDFM context, efficiency of the ARMA estimation may be subordinate considering that f_t is not observed and that the ARMA estimation will thus be applied to a (probably not

[1] Autoregressive with exogenous variables

efficient) estimate $\hat{\psi}_t^{n,\cdot}$. Therefore, here we will not consider ML-estimation, but the computationally simpler and consistent *autoregression-regression* approach (see for instance Mayne *et al.* (1984), Koreisha and Pukkila (1989) or Poskitt (1992)). First we will give a presentation of this estimation method adapted to the general case $q \leq s$ and report consistency results (as $T \to \infty$) when f_t, $t = 1, \ldots, T$, is treated as if it was observed. Second we will analyze the asymptotic properties of the estimated ARMA model when f_t is replaced by the estimates $\hat{\psi}_t^{n,\cdot}$ from Section 5.2 as both n and T tend to infinity.

6.1 Autoregression-regression approach

The autoregression-regression (or two-stages least squares) approach is probably the computationally simplest method for multivariate ARMA estimation. The idea is to derive estimates for the innovations series (ε_t) associated with the ARMA process through a long autoregression fit in the first place, and then to regress f_t onto lagged values of these estimates together with lagged values of f_t in order to estimate the ARMA coefficients. In the context of estimation the non-uniqueness of the ARMA realization becomes of importance. Thus here we will make use of the (singular) echelon ARMA realization introduced in Section 3.3.4 that provides a unique parametrization.

Let the (singular) reversed echelon ARMA representation of f_t be

$$A_0 f_t = A_1 f_{t-1} + \ldots + A_P f_{t-P} + B_0 \varepsilon_t + B_1 \varepsilon_{t-1} + \ldots + B_Q \varepsilon_{t-Q}, \tag{6.1.1}$$

where the coefficients satisfy the conditions of Theorem 3.3.4. A unique representation in standard form thus results from left multiplying (6.1.1) by A_0^{-1},

$$f_t = A_0^{-1} A_1 f_{t-1} + \ldots + A_0^{-1} A_P f_{t-P} + A_0^{-1} B_0 \varepsilon_t + A_0^{-1} B_1 \varepsilon_{t-1} \ldots + A_0^{-1} B_Q \varepsilon_{t-Q}, \tag{6.1.2}$$

where for the MA coefficient at lag 0 we have $A_0^{-1} B_0 = K_0$ (recalling that $k(z) = \sum_j K_j z^j$ is the transfer function corresponding to the Wold representation of (f_t)). Then, as shown in Section 3.3.6, if the miniphase assumption is satisfied, f_t may be expressed by a possibly infinite autoregression, i.e.

$$f_t = \sum_{j=1}^{\infty} \Phi_j f_{t-j} + K_0 \varepsilon_t. \tag{6.1.3}$$

Recall, that generically (that is in the zeroless transfer function case) (f_t) may always be represented by a finite order AR-system. Hence in general, approximating (6.1.3) by a finite autoregression[2] of order h say, where h is greater than P (and thus Q), yields estimates for $K_0 \varepsilon_t$. Let us consider an approximating model of order h,

$$f_t = \sum_{j=1}^{h} \Phi_{h,j} f_{t-j} + e_{h,t}, \tag{6.1.4}$$

[2]Notice that here the term autoregression is slightly misused since the innovations of an approximating finite model are not necessarily white noise as they should be in case of an AR process.

where the first term on the right is defined as the projection of f_t onto the Hilbert space spanned by $f_{i\tau}, i = 1, \ldots, s$, $\tau = t-h, \ldots, t-1$. Then the coefficients $\Phi_{h,j}, j = 1, \ldots, h$ solve the *Yule-Walker equations*, i.e.

$$\Gamma_f(k) = \sum_{j=1}^{h} \Phi_{h,j} \Gamma_f(k-j), \text{ for } k = 1, \ldots, h. \tag{6.1.5}$$

Defining $G_f(h)$ as the $(hs \times hs)$ block Toeplitz matrix, where the (i,j) block is $\Gamma_f(j-i)$, i.e.

$$G_f(h) = \begin{pmatrix} \Gamma_f(0) & \Gamma_f(1) & \ldots & \Gamma_f(h-1) \\ \Gamma_f(-1) & \Gamma_f(0) & \ldots & \Gamma_f(h-2) \\ \vdots & \vdots & \ddots & \\ \Gamma_f(-h+1) & \Gamma_f(-h+2) & \ldots & \Gamma_f(0) \end{pmatrix}, \tag{6.1.6}$$

then (6.1.5) can be written as

$$(\Phi_{h,1}, \ldots, \Phi_{h,h}) G_f(h) = (\Gamma_f(1), \ldots, \Gamma_f(h)). \tag{6.1.7}$$

From the projection theorem we know that for any $h \in \mathbb{N}$ the projection $\sum_{j=1}^{h} \Phi_{h,j} f_{t-j}$ exist and is unique. From the discussion in Section 3.3.6 we know that if $q < s$ and if h exceeds some minimum value, the random variables $f_{i\tau}, i = 1, \ldots, s$, $\tau = t-h, \ldots, t-1$ are linear dependent implying that $G_f(h)$ will be singular. If $G_f(h)$ is singular, then in principle two situations could arise. First, equation (6.1.7) has no solution, but this is ruled out by the projection theorem. Second, there are infinitely many solutions for $\Phi_{h,j}, j = 1, \ldots, h$, but although the coefficients are not unique, the projection is. Of course, in the zeroless case truncating (6.1.3) at h larger than some minimum lag length does not mean an approximation, but a true model representation.

Now, let the fitted model be

$$f_t = \sum_{j=1}^{h} \hat{\Phi}_{h,j} f_{t-j} + \hat{e}_{h,t}, \tag{6.1.8}$$

where $\hat{\Phi}_{h,j}$, $j = 1, \ldots, h$ solve the Yule-Walker equations using the sample autocovariances

$$\hat{\Gamma}_f(s) = \frac{1}{T} \sum_{t=s+1}^{T} f_t f_{t-s}'.$$

and

$$\hat{e}_{h,t} = f_t - \sum_{j=1}^{h} \hat{\Phi}_{h,j} f_{t-j}. \tag{6.1.9}$$

Then we propose to estimate ε_t (or to be correct a static rotation of the true ε_t, i.e. $H'\varepsilon_t$ where $H'H = HH' = I_q$) as the first q sample principal components of $\hat{e}_{h,t}$. Let $\hat{\Gamma}_{e,h} = 1/T \sum_{t=1}^{T} \hat{e}_{h,t} \hat{e}_{h,t}'$ be the sample covariance of $\hat{e}_{h,t}$, then

$$\hat{\varepsilon}_{h,t} = \hat{D}^{-1/2} \hat{P}' \hat{e}_{h,t}, \tag{6.1.10}$$

where $\hat{D}$ denotes the diagonal matrix containing the q largest eigenvalues of $\hat{\Gamma}_{e,h}$ in its diagonal and $\hat{P}$ denotes the matrix of corresponding normalized eigenvectors. Correspondingly, the estimate for the coefficient matrix K_0 (or indeed for $K_0 H$) is given as

$$\hat{K}_{h,0} = \hat{P}\hat{D}^{1/2}. \tag{6.1.11}$$

Notice, that if k is post multiplied by an orthogonal matrix, the linear dependence structure of the rows of the resulting transfer function will be the same as of k. Hence the Kronecker indices as well as the AR polynomial a of the reversed echelon form will remain unchanged, and only the MA polynomial b will be rotated.

In the next step, the estimates $\hat{\varepsilon}_{h,t}$ and their lagged values will be substituted into the ARMA(P,Q) model to set up a multivariate regression structure. Rearranging (6.1.1) we get,

$$f_t = (I_s - A_0)(f_t - K_0\varepsilon_t) + A_1 f_{t-1} + \ldots + A_P f_{t-P} + B_1\varepsilon_{t-1} + \ldots + B_Q\varepsilon_{t-Q} + K_0\varepsilon_t, \tag{6.1.12}$$

where we used the fact that $A_0 K_0 = B_0$. Substituting the estimates $\hat{e}_{h,t}$ and $\hat{\varepsilon}_{h,t-j}, j = 1,\ldots,Q$, respectively, yields,

$$f_t = (I_s - A_0)(f_t - \hat{e}_{h,t}) + A_1 f_{t-1} + \ldots + A_P f_{t-P} + B_1\hat{\varepsilon}_{h,t-1} + \ldots + B_Q\hat{\varepsilon}_{h,t-Q} + v_{h,t}, \tag{6.1.13}$$

where $v_{h,t}$ denotes the residual. Then we set up the regression model

$$\mathbf{f} = (I_s - A_0, A_1, \ldots, A_P, B_1, \ldots, B_Q)X_h + V_h \tag{6.1.14}$$

where $\mathbf{f} = (f_{P+1}, \ldots, f_T)$, $X_h = (x_{h,P+1}, \ldots, x_{h,T})$ with

$$x_{h,t} = \begin{pmatrix} f_t - \hat{e}_{h,t} \\ f_{t-1} \\ \vdots \\ f_{t-P} \\ \hat{\varepsilon}_{h,t-1} \\ \vdots \\ \hat{\varepsilon}_{h,t-Q} \end{pmatrix}$$

and $V_h = (v_{h,P+1}, \ldots, v_{h,T})$ are the regression residuals. Since (6.1.1) is in reversed singular echelon form, the coefficients satisfy a number of linear identifying restrictions (see Theorem 3.3.4). Concretely, let τ be the vector containing the elements of $(I_s - A_0, \ldots, A_P, B_1, \ldots, B_Q)$ arranged in a certain order, i.e. using the *vec operator*, that transforms an $(m \times n)$ matrix A into an mn vector by stacking its columns,

$$\tau = \mathrm{vec}(I_s - A_0, \ldots, A_P, B_1, \ldots, B_Q).$$

Then there exist a restriction matrix R_α and a vector τ_α containing the free parameters of $(I_s - A_0, \ldots, A_P, B_1, \ldots, B_Q)$ and depending on the Kronecker indices α, such that τ can be expressed as

$$\tau = R_\alpha \tau_\alpha. \tag{6.1.15}$$

Vectorizing (6.1.14) yields

$$\text{vec}(\mathbf{f}) = (X_h' \otimes I_s)\tau + \text{vec}(V_h), \tag{6.1.16}$$

where $\otimes$ denotes the *Kronecker product* (see for instance the appendix of Lütkepohl (2005) for a definition and basic rules) and thus

$$\text{vec}(\mathbf{f}) = (X_h' \otimes I_s)R_\alpha\tau_\alpha + \text{vec}(V_h). \tag{6.1.17}$$

Then the least-squares estimator of τ_α becomes

$$\hat{\tau}_{h,\alpha} = (R_\alpha'(X_h X_h' \otimes I_s)R_\alpha)^{-1}R_\alpha'(X_h \otimes I_s)\text{vec}(\mathbf{f}), \tag{6.1.18}$$

from which one may easily compute the estimated coefficients $\hat{A}_{h,j}, j = 0, \ldots, P$ and $\hat{B}_{h,j}, j = 0, \ldots, Q$ respectively, of the echelon form (6.1.1), where $\hat{B}_{h,0} = \hat{A}_{h,0}\hat{K}_{h,0}$.

The following example will illustrate the construction of τ_α and R_α.

Example 6.1.1. Consider the reversed echelon form from Example 3.3.4 given in (3.3.60), where $q = 1, s = 2$ and $\alpha = (3, 1)$. Then

$$\tau = \begin{pmatrix} 0 \\ a_{21,0} \\ 0 \\ 0 \\ a_{11,1} \\ a_{21,1} \\ 0 \\ a_{22,1} \\ a_{11,2} \\ 0 \\ 0 \\ 0 \\ a_{11,3} \\ 0 \\ a_{12,3} \\ 0 \\ b_{1,1} \\ 0 \\ b_{1,2} \\ 0 \end{pmatrix}, \ \tau_\alpha = \begin{pmatrix} a_{21,0} \\ a_{11,1} \\ a_{21,1} \\ a_{22,1} \\ a_{11,2} \\ a_{11,3} \\ a_{12,3} \\ b_{1,1} \\ b_{1,2} \end{pmatrix}, \text{ and } R_\alpha = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

6.1.1 Asymptotic properties for f_t known

Here, we will report consistency results for the estimation procedure described above (see for instance Poskitt (1992) for a reference treating the standard ARMA case). Quite evidently it

is most crucial in this method, that the innovations series is consistently estimated through the autoregression fit (6.1.8). Therefore we will need the following strengthened assumption on the innovations process (ε_t).

Assumption 6.1 (Consistency of ARMA estimates)**.**
The white noise process (ε_t) is an ergodic martingale difference process. Thus if $(\mathscr{F}_t)$ denotes the filtration generated by $(\varepsilon_\tau), \tau \leq t$, then

$$\mathbb{E}(\varepsilon_t | \mathscr{F}_{t-1}) = 0.$$

Moreover,

$$\mathbb{E}(\varepsilon_t \varepsilon_t' | \mathscr{F}_{t-1}) = I_q \text{ and } \mathbb{E}\,\varepsilon_{it}^4 < \infty \text{ for } i = 1, \ldots, q.$$

Notice, that the first part of Assumption 6.1 is equivalent to the statement that the best linear predictor of f_t equals the best predictor and hence is justified if linear prediction is reasonably accurate. The second part of Assumption 6.1 is a stronger restriction of generality and not as easy to be verified.

Theorem 6.1.1. *Let (f_t) have the ARMA representation (6.1.1), that is in reversed echelon form and satisfies the stability and the miniphase assumptions and where (ε_t) satisfies Assumption 6.1. Further let h_T be such that $\lim_{T\to\infty} h_T = \infty$ and $\lim_{T\to\infty} (\log(T)/T)^{1/2} h_T = 0$. Then*

$$\operatorname*{plim}_{T\to\infty} \hat{A}_{h_T,j} = A_j, \ j = 0, \ldots, P$$

$$\operatorname*{plim}_{T\to\infty} \hat{B}_{h_T,j} = B_j H, \ j = 0, \ldots, Q$$

hold, where H is a constant orthogonal $(q \times q)$ matrix.

The proof of Theorem 6.1.1 depends largely on the following result, that has been shown in Hannan and Kavalieris (1986), admittedly for the case that $q = s$, however since the proofs are not affected from possibly singular prediction errors, it still holds if $q < s$.

Lemma 6.1.1. *Under the assumptions of Theorem 6.1.1, then*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} (\hat{e}_{h_T,t} - K_0 \varepsilon_t)(\hat{e}_{h_T,t} - K_0 \varepsilon_t)' = 0 \ a.s., \tag{6.1.19}$$

and

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} (K_0 \varepsilon_t)(\hat{e}_{h_T,t} - K_0 \varepsilon_t)' = 0 \ a.s.. \tag{6.1.20}$$

Proof. See Hannan and Kavalieris (1986). ■

Obviously, the adequacy of the finite AR fit (6.1.8) depends on the choice of the lag length h. Notice, that for given T unlike in the case of exogenous regressors the fit does not necessarily improve by increasing h, because since lagged values of f_t are required, it has to be taken into account that the effective sample size becomes $T-h$. Hence by selecting a lag length that tends to infinity as $T \to \infty$ but not faster than $(\log(T)/T)^{-1/2}$ (for instance $\log(T)$) the regression errors $\hat{e}_{h_T,t}$ are strongly consistent estimates. The following corollary is a nearly immediate consequence.

Corollary 6.1.1. *Under the assumptions of Theorem 6.1.1 with $\hat{\varepsilon}_{h_T,t}$ as in (6.1.10) and $\hat{K}_{h_T,0}$ defined as in (6.1.11), then*

$$\operatorname*{plim}_{T\to\infty} \hat{\varepsilon}_{h_T,t} = H'\varepsilon_t$$

and

$$\operatorname*{plim}_{T\to\infty} \hat{K}_{h_T,0} = K_0 H,$$

hold, where H is a constant orthogonal $(q \times q)$ matrix.

Proof. Let us first consider the singular value decomposition of K_0,

$$K_0 = PD^{1/2}Q'.$$

Then we may set $H = Q$, hence $K_0H = PD^{1/2}$, $K_0K_0' = PDP'$ and $H'\varepsilon_t = P'D^{-1/2}K_0\varepsilon_t$. Second, it follows from (6.1.19) and (6.1.20) that

$$\lim_{T\to\infty} \hat{\Gamma}_{e,h} = \frac{1}{T}\sum_{t=1}^{T} K_0\varepsilon_t\varepsilon_t'K_0' \text{ a.s.}$$

Since the latter converges to K_0K_0' in probability, we also have

$$\operatorname*{plim}_{T\to\infty} \hat{\Gamma}_{e,h} = K_0K_0'.$$

Continuity of eigenvalues and normalized eigenvectors with respect to the matrix elements thus implies (for a suitable normalization condition) $\operatorname{plim}_T \hat{P} = P$ and $\operatorname{plim}_T \hat{D} = D$, the result follows. ■

Now let us define the following population quantities corresponding to $\hat{M}_{h,\alpha} = R_\alpha'(X_hX_h' \otimes I_s)R_\alpha$ and $\hat{m}_{h,\alpha} = R_\alpha'(X_h \otimes I_s)\mathrm{vec}(\mathbf{f})$. Hence, the matrix

$$M_\alpha = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathfrak{X}_\alpha(e^{-i\lambda})\mathfrak{X}_\alpha(e^{-i\lambda})^* d\lambda, \tag{6.1.21}$$

where

$$\mathfrak{X}_\alpha(z) = R_\alpha' \begin{pmatrix} k(z) - K_0 \\ (z^1,\ldots,z^P)' \otimes k(z) \\ (z^1,\ldots,z^Q)' \otimes H' \end{pmatrix} \otimes I_s, \tag{6.1.22}$$

and the vector

$$m_\alpha = \frac{1}{2\pi}\int_{-\pi}^{\pi} R'_\alpha \begin{pmatrix} \text{vec}(k(e^{-i\lambda})(k(e^{-i\lambda}) - K_0)^*) \\ (z^1, \ldots, z^P)' \otimes \text{vec}(k(e^{-i\lambda})k(e^{-i\lambda})^*) \\ (z^1, \ldots, z^Q)' \otimes \text{vec}(k(e^{-i\lambda})H) \end{pmatrix} d\lambda. \tag{6.1.23}$$

Then we have the following result.

Lemma 6.1.2. *Under the assumptions of Theorem 6.1.1,*

$$\operatorname*{plim}_{T\to\infty} \hat{M}_{h_T,\alpha} = M_\alpha$$

and

$$\operatorname*{plim}_{T\to\infty} \hat{m}_{h_T,\alpha} = m_\alpha.$$

Proof. See Poskitt (1992). ■

Proof. (Proof of Theorem 6.1.1)
By definition, $\hat{\tau}_{h_T,\alpha}$ is obtained as the solution of the normal equations $\hat{M}_{h_T,\alpha}\theta = \hat{m}_{h_T,\alpha}$. Now let $\tau = \text{vec}(I_s - A_0, \ldots, A_P, B_1 H, \ldots, B_Q H)$ and τ_α be the vector of free parameters in τ corresponding to α. From its definition we get that, $\tau_\alpha = M_\alpha^{-1} m_\alpha$ (see Poskitt (1992)). Thus it follows from Lemma 6.1.2, that $\text{plim}_T\, \hat{\tau}_{h_T,\alpha} = \tau_\alpha$. ■

6.1.2 Asymptotic properties for f_t estimated

Here we will assume that the GDFM is estimated using the static principal components method[3], and let $\hat{\psi}_t^n = \hat{\psi}_t^{n,PCA}$. Then we know from Theorem 5.2.3 that $\hat{\psi}_t^n - R^{n'} f_t \to 0$ in probability as $\min(n, T) \to \infty$, where $R^{n'} R^n = R^n R^{n'} = I_s$. Let $\tilde{f}_t^n = R^{n'} f_t$ denote the rotated factors where $\tilde{f}_t^n = R^{n'} k(z)\varepsilon_t = \tilde{k}^n(z)\varepsilon_t$ and let the reversed echelon ARMA representation of $\tilde{f}_t^n$ be

$$\tilde{A}_0^n \tilde{f}_t^n = \tilde{A}_1^n \tilde{f}_{t-1}^n + \ldots + \tilde{A}_P^n \tilde{f}_{t-P}^n + \tilde{B}_0^n \varepsilon_t + \ldots + \tilde{B}_Q^n \varepsilon_{t-Q}. \tag{6.1.24}$$

Then we propose to estimate the ARMA parameters $\tilde{A}_j^n$, $j = 0, \ldots, P$, $\tilde{B}_j^n$, $j = 0, \ldots, Q$ using the autoregression-regression method described above with $\hat{\psi}_t^n$ in place of f_t.

Consistency of these estimates follows from consistency of the ARMA estimates if $\tilde{f}_t^n$ was known together with the fact that $\hat{\psi}_t^n - \tilde{f}_t^n$ converges to 0 in probability as $(n, T) \to \infty$. Let us provide the previously defined estimators with a superscript n to indicate that they are based on $\hat{\psi}_t^n$. Then we have the following result.

Theorem 6.1.2. *If Assumptions 1.1, 2.1, 3.1, 5.2, 5.3 and 6.1 hold, and if h_T is such that $\lim_{T\to\infty} h_T = \infty$ and $\lim_{T\to\infty} (\log(T)/T)^{1/2} h_T = 0$, then*

[3]Of course, similar results as those to follow can also be obtained for the other two methods presented, since they all yield consistent estimates for the static factor space.

(i)

$$\operatorname*{plim}_{(n,T)\to\infty} \hat{\Gamma}^n_{e,h_T} - \tilde{K}^n_0 \tilde{K}^{n\prime}_0 = 0,$$

(ii)

$$\operatorname*{plim}_{(n,T)\to\infty} \hat{\varepsilon}^n_{e,h_T} - H'\varepsilon_t = 0,$$

(iii)

$$\operatorname*{plim}_{(n,T)\to\infty} \hat{A}^n_{h_T,j} - \tilde{A}^n_j = 0, \quad \operatorname*{plim}_{(n,T)\to\infty} \hat{B}^n_{h_T,j} - \tilde{B}^n_j H = 0,$$

where H is an orthogonal $(q \times q)$ matrix.

Proof. We will need the following preliminary results, that are due to Doz *et al.* (2007). Again $\|C\|$ will denote the spectral norm of a matrix C.

(A) Under the assumptions imposed

$$\|\hat{\Gamma}^n_y(h) - \Gamma^n_\chi(h)\| = O(1) + O_p\left(\frac{n}{\sqrt{T}}\right)$$

as $n, T \to \infty$. For $h = 0$ this has been shown in Lemma 5.2.5(i), the proof for arbitrary h works with the same arguments.

(B) Under the assumptions imposed

$$\operatorname*{plim}_{(n,T)\to\infty} \hat{\Gamma}^n_\psi(h) - \tilde{\Gamma}_f(h) = 0,$$

where $\tilde{\Gamma}_f(h) = R^{n\prime}\Gamma_f(h)R^n$ is the autocovariance function of $(\tilde{f}^n_t)$. We may write

$$\begin{aligned}\hat{\Gamma}_\psi(h) &= \frac{1}{T}\sum_{t=h+1}^{T} \hat{\psi}^n_t \hat{\psi}^{n\prime}_{t-h} = \hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \hat{\Gamma}^n_y(h) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}} \\ &= \hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} (\hat{\Gamma}^n_y(h) - \Gamma^n_\chi(h)) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}} + \hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \Gamma^n_\chi(h) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}}. \quad (6.1.25)\end{aligned}$$

Then the first term converges to 0 in probability as $n, T \to \infty$ with (A) and under Assumption 5.2, since

$$\|\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} (\hat{\Gamma}^n_y(h) - \Gamma^n_\chi(h)) \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}}\| \le \frac{1}{\hat{\mu}^n_s} \|\hat{O}_1^{n\prime} (\hat{\Gamma}^n_y(h) - \Gamma^n_\chi(h)) \hat{O}_1^n\| \le \frac{1}{\hat{\mu}^n_s} \|(\hat{\Gamma}^n_y(h) - \Gamma^n_\chi(h))\|$$

is $O_p(1/n) + O_p(1/\sqrt{T})$.

The second term of (6.1.25) equals

$$\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \tilde{\Lambda}^n \tilde{\Gamma}_f(h) \tilde{\Lambda}^{n\prime} \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}},$$

hence

$$\hat{\Gamma}^n_\psi(h) - \tilde{\Gamma}_f(h) = \left(\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n\prime} \tilde{\Lambda}^n - I_s\right) \tilde{\Gamma}_f(h) \left(\tilde{\Lambda}^{n\prime} \hat{O}_1^n \hat{\Omega}_1^{n^{-1/2}}\right) + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$$

and we know from Lemma 5.2.6 that $\hat{\Omega}_1^{n^{-1/2}} \hat{O}_1^{n'} \tilde{\Lambda}^n$ converges to I_s in probability as $(n, T) \to \infty$, which yields (B).

Next consider the autoregression fit

$$\hat{\psi}_t^n = \sum_{j=1}^{h_T} \hat{\Phi}_{h_T,j}^n \hat{\psi}_{t-j}^n + \hat{e}_{h_T,t}^n, \tag{6.1.26}$$

where $\hat{\Phi}_{h_T,j}^n$, $j = 1, \ldots, h_T$, are solutions to the Yule-Walker equations using the sample autocovariances $\hat{\Gamma}_\psi^n(h)$. Then it follows from (B) that $\text{plim}_{n,T\to\infty} \hat{e}_{h_T,t}^n - \tilde{e}_{h_T,t}^n = 0$ implying further that $\text{plim}_{n,T\to\infty} \hat{\Gamma}_{e,h_T}^n - \tilde{\Gamma}_{e,h_T}^n = 0$, where we have from Lemma 6.1.1, that the latter converges to $\tilde{K}_0^n \tilde{K}_0^{n'} = R^{n'} K_0 K_0' R^n$ as $T \to \infty$, which yields (i).

Defining $\hat{\varepsilon}_{h_T,t}^n$ as in (6.1.10) and observing that the singular value decomposition of $\tilde{K}_0^n$ is $\tilde{K}_0^n = (R^{n'} P) D^{1/2} Q'$, we may set $H = Q$ independent of n, and continuity of the eigenvalues and suitably normalized eigenvectors with respect to the matrix elements together with (i) and Corollary 6.1.1 imply (ii) .

Let us define $X_h^n = (x_{h,P+1}^n, \ldots, x_{h,T}^n)$ with

$$x_{h,t}^n = \begin{pmatrix} \hat{\psi}_t^n - \hat{e}_{h,t}^n \\ \hat{\psi}_{t-1}^n \\ \vdots \\ \hat{\psi}_{t-P}^n \\ \hat{\varepsilon}_{h,t-1}^n \\ \vdots \\ \hat{\varepsilon}_{h,t-Q}^n \end{pmatrix},$$

$\hat{M}_{h,\alpha(n)}^n = R_{\alpha(n)}'(X_h^n X_h^{n'} \otimes I_s) R_{\alpha(n)}$ and $\hat{m}_{h,\alpha(n)}^n = R_{\alpha(n)}'(X_h^n \otimes I_s)\text{vec}(\hat{\psi}^{\mathbf{n}})$, where $\alpha(n)$ denotes the Kronecker indices corresponding to $\tilde{k}^n$. These indices retain the same magnitude as the Kronecker indices α of k, but possibly in a different order. Thus combining (i), (ii), (B) and Lemma 6.1.2, it follows that

$$\underset{n,T\to\infty}{\text{plim}} \hat{M}_{h_T,\alpha(n)}^n - \tilde{M}_{\alpha(n)} = 0$$

and

$$\underset{n,T\to\infty}{\text{plim}} \hat{m}_{h_T,\alpha(n)}^n - \tilde{m}_{\alpha(n)} = 0,$$

where $\tilde{M}_{\alpha(n)}$ and $\tilde{m}_{\alpha(n)}$ are defined as in (6.1.21) - (6.1.23) with $\tilde{k}^n$ and $\tilde{K}_0$ respectively, in place of k and K_0.

Then $\hat{\tau}_{h_T,\alpha(n)}^n$ is obtained as the solution of the normal equations $\hat{M}_{h_T,\alpha(n)}^n \theta = \hat{m}_{h_T,\alpha(n)}^n$. Now let $\tau^n = \text{vec}(I_s - \tilde{A}_0^n, \ldots, \tilde{A}_P^n, \tilde{B}_1^n H, \ldots, \tilde{B}_Q^n H)$ and $\tau_{\alpha(n)}^n$ be the vector of free parameters in τ^n

corresponding to $\alpha(n)$. Thus $\tau^n_{\alpha(n)} = M^{-1}_{\alpha(n)} m_{\alpha(n)}$ and it follows from what was said above, that $\operatorname{plim}_{n,T} \hat{\tau}^n_{h_T,\alpha(n)} - \tau^n_{\alpha(n)} = 0$. ■

Chapter 7

Model selection in the GDFM

According to the results of Section 3.3 model selection in the GDFM comprises the determination of the numbers of dynamic and static factors, i.e. q and s, and the specification of an ARMA model for the static factor. As far as the first problem is concerned, a number of estimation methods has been proposed for the number of static factors (see Bai and Ng (2002)) as well as for the number of dynamic factors (see Bai and Ng (2007), Amengual and Watson (2007) and Hallin and Liska (2007)). These methods will be discussed in Sections 7.1 and 7.2 below.

The second problem, ARMA model selection, has been extensively studied in the literature and many different approaches exist (see for instance Hannan and Deistler (1988), Lütkepohl (2005), Reinsel (1993) or Tiao and Tsay (1989)). Practical procedures for estimating the Kronecker indices have been proposed and analyzed for instance by Poskitt (1992), Lütkepohl and Poskitt (1996) and Nsiri and Roy (1992). Although none of these approaches takes into account the particular situation occurring in the GDFM framework, where the output variable f_t is itself unobserved and may be dynamically singular, in Section 7.3 we will propose to estimate the Kronecker indices using the algorithm of Lütkepohl and Poskitt (1996), which seems justified by the subsequent simulation results (see Chapter 8).

Notice that as in the previous chapter, the methods will be presented exemplarily for GDFM estimation based on the static principal components method.

7.1 Estimation of the number of static factors

For the estimation of the number of static factors, i.e. s, a number of procedures have been proposed. For instance, Lewbel (1991) and Donald (1997) use the rank of a matrix to test for the numbers of factors. Here we will consider the *information criterion* (IC) approach, as proposed by Bai and Ng (2002). The information criteria considered here, in general consist of the sum of two quantities, the first being a measure of the goodness of fit and the second being a measure of the corresponding model complexity. More specifically, let the measure of goodness

of fit corresponding to k static factors be the trace of the estimated noise variance divided by n, i.e.

$$V_T^n(k) = \frac{1}{n} \operatorname{tr} \hat{\Gamma}_u^{n,k},$$

which, when the static PC estimates are used, equals the sum of the $n-k$ smallest eigenvalues of $\hat{\Gamma}_y^n$ divided by n, i.e.

$$V_T^n(k) = \frac{1}{n} \sum_{j=k+1}^{n} \hat{\mu}_j^n.$$

Further let the measure of corresponding model complexity be $kp(n,T)$, where $p(n,T)$ is a deterministic penalty function defining a certain trade-off between goodness of fit and complexity. Then a class of IC is defined as

$$IC_T^n(k) = \frac{1}{n} \sum_{j=k+1}^{n} \hat{\mu}_j^n + kp(n,T) \tag{7.1.1}$$

and estimation of s consists of selecting the minimizing argument of $IC_T^n(k)$, i.e.

$$\hat{s}^n = \underset{0 \le k \le s_{max}}{\arg\min}\ IC_T^n(k), \tag{7.1.2}$$

where s_{max} denotes some predefined upper bound for the actual s. In the sequel we are going to show (as has been done in Bai and Ng (2002) under slightly different assumptions) that this IC approach yields consistent estimates, if $p(n,T)$ is adequately chosen, hence if $p(n,T)$ satisfies certain limit conditions as n and T tend to infinity. Here we will give our own proofs, that largely lean on the proofs given in Hallin and Liska (2007) (in the course of estimating the number of dynamic factors q).

Theorem 7.1.1. *Suppose that Assumptions 1.1, 2.1, 3.1, 5.2 and 5.3 hold and let $p(n,T)$ be such that as $\min(n,T) \to \infty$,*

$$(i) \quad p(n,T) \to 0 \quad \textit{and} \quad (ii) \quad \min(n, T^{1/2}) p(n,T) \to \infty.$$

Then

$$\operatorname*{plim}_{\min(n,T)\to\infty} \hat{s}^n = s.$$

Proof. We will need some preliminary results.

(A) Under the assumptions imposed, there exists a positive real M, such that

$$\sup_n \max_{1 \le i,j \le n} \mathbb{E}\left(|\hat{\Gamma}_y^n - \Gamma_y^n|_{i,j}^2 \right) \le MT^{-1},$$

where the underscript i,j means that the quantities considered are the (i,j) elements of the respective matrices. The statement has been shown in Lemma 5.2.5(i).

(B) Under the assumptions imposed, for every $\varepsilon > 0$ there exist T_ε and B_ε such that for s_{max} fixed, $n \in \mathbb{N}$ and $T \geq T_\varepsilon$,

$$\max_{1 \leq k \leq s_{max}} P\left(T^{1/2} n^{-1} |\hat{\mu}_k^n - \mu_k^n| > B_\varepsilon) \leq \varepsilon\right).$$

This follows with the same arguments as in the proof of Lemma 5.2.5(ii).

For the result of the theorem we will have to show that $P(IC_T^n(s) < IC_T^n(k)) \to 1$ for $k \neq s$, $k \leq s_{max}$ as $(n, T) \to \infty$.

Let us first consider the case $k < s$. Then

$$IC_T^n(k) > IC_T^n(s) \tag{7.1.3}$$

if and only if

$$\frac{1}{n} \sum_{j=k+1}^{s} \hat{\mu}_j^n > (s-k) p(n,T),$$

that is with (B) if and only if

$$\sum_{j=k+1}^{s} \left(\frac{\mu_j^n}{n} + K_{1,T}^n\right) > (s-k) p(n,T),$$

where $K_{1,T}^n = O_p(T^{-1/2})$ uniformly in n. By Assumption 5.2 $\mu_{\chi,j}^n$, $j = 1, \ldots, s$ diverge linearly in n, implying that μ_j^n, $j = 1, \ldots, s$ also diverge linearly in n (using Corollary A.1.1a)), hence

$$\sum_{j=k+1}^{s} \left(\frac{\mu_j^n}{n}\right) \geq \underline{c}_j > 0.$$

Since $K_{1,T}^n$ converges to 0, a sufficient condition for (7.1.3) to hold with probability 1 as $n, T \to \infty$ is that $p(n,T) \to 0$.

Let us now consider the case $s < k$. Then

$$IC_T^n(k) > IC_T^n(s) \tag{7.1.4}$$

if and only if

$$(s-k) p(n,T) > \frac{1}{n} \sum_{j=s+1}^{k} \hat{\mu}_j^n,$$

that is with (B) if and only if

$$(s-k) p(n,T) > \frac{1}{n} \sum_{j=s+1}^{k} \left(\mu_j^n + K_{2,T}^n\right),$$

where $K_{2,T}^n = O_p(T^{-1/2})$ uniformly in n. Further we know from Assumption 5.2 and Corollary A.1.1b) that $n^{-1}\mu_j^n$, $j = s+1, \ldots, n$ is $O(n^{-1})$. Hence a sufficient condition for (7.1.4) to hold with probability 1 as $n, T \to \infty$ is that $\min(n, T^{1/2}) p(n,T) \to \infty$. ■

An alternative class of information criteria is defined as the logarithmic form of $IC_T^n(k)$ (see Bai and Ng (2002)), precisely as

$$IC_T^{n*}(k) = \log(V_T^n) + kp(n,T).$$

Then it can be shown using essentially the same arguments as in the last proof that the logarithmic IC has the same asymptotic properties as the original one.

Corollary 7.1.1. *Under the assumptions of Theorem 7.1.1, the class of criteria defined by*

$$IC_T^{n*}(k) = \log(V_T^n) + kp(n,T)$$

will also consistently estimate s.

In regard of Theorem 7.1.1 an obvious choice for $p(n,T)$ is

$$p(n,T) = \frac{\log(\min(n,T^{1/2}))}{\min(n,T^{1/2})}, \tag{7.1.5}$$

but of course there are infinitely many other possibilities (see Bai and Ng (2002) for more examples). Notice that in general, if $p(n,T)$ is an adequate penalty function in the sense of Theorem 7.1.1, then $cp(n,T)$, where c is an arbitrary positive real, also is. Bai and Ng (2002) propose to take $c = V_T^n(s_{max})$ in correspondence to Mallows's C_p criterion. Although asymptotically the estimated number of static factors will be the same for all corresponding information criteria, the results may differ in finite samples.

7.2 Estimation of the number of dynamic factors

Recently a number of different approaches for estimating the number q of dynamic factors have been suggested. Hallin and Liska (2007) propose the use of information criteria based on the dynamic PC model. This approach is analogous to the IC approaches of Bai and Ng (2002) discussed in the previous section, but is complicated by the fact that it requires a consistent estimate of the spectral density involving a number of additional parameters (i.e. the smoothing window and the window length). Amengual and Watson (2007) use an information criterion approach based on the static representation of the GDFM avoiding the estimation of the spectral density. More precisely, they suggest to substitute the AR-representation of the static factors (3.3.69) into the quasi-static representation (3.3.63), and obtain

$$\underbrace{y_t^n - \bar{\Lambda}^n \sum_{j=1}^{\infty} \Phi_j f_{t-j}}_{\bar{y}_t^n} = \bar{\Lambda}^n K_0 \varepsilon_t + u_t^n.$$

Hence, $(\bar{y}_t^n)$, has a static generalized factor model representation with q-dimensional static factor ε_t. Therefore, q may be estimated by means of the IC approach of Bai and Ng (2002) applied to

$(\bar{y}_t^n)$. Here we will follow Bai and Ng (2007), who propose to test the rank of the prediction error covariance in the AR representation of (f_t), since this approach matches especially well with the autoregression-regression method proposed to estimate the ARMA model. We will extend their results to the case where the AR representation of (f_t) is of infinite order (see Section 3.3.6).

Recall that under the Assumptions imposed f_t has a (possibly infinite) AR representation (3.3.69), where the one-step prediction errors $e_t = K_0\varepsilon_t$ have an $(s \times s)$ covariance matrix $\Gamma_e = K_0K_0'$ of rank $q \leq s$. Recall further, that f_t can be consistently estimated up to rotations as $n, T \to \infty$ by the first s sample principal components of y_t^n, $\hat{\psi}_t^{n,PCA}$. Then the idea is to fit an appropriately truncated long AR model for $\hat{\psi}_t^{n,PCA}$ and test the rank of the sample covariance of the corresponding prediction errors.

To determine the rank of a positive semidefinite $(s \times s)$ matrix A with eigenvalues $\omega_j(A)$, $j = 1, \ldots, s$, in descending order of magnitude, consider the following quantities:

$$D_1(k) = \left(\frac{\omega_{k+1}^2(A)}{\sum_{j=1}^s \omega_j^2(A)} \right)^{\frac{1}{2}}, \quad D_2(k) = \left(\frac{\sum_{l=k+1}^s \omega_l^2(A)}{\sum_{j=1}^s \omega_j^2(A)} \right)^{\frac{1}{2}}. \tag{7.2.1}$$

Obviously, if $\operatorname{rk} A = q$, then $D_1(k) = D_2(k) = 0$ for $k \geq q$.

As an intermediate step consider the finite autoregression fit of f_t (6.1.8), and recall that $\hat{\Gamma}_{e,h} = 1/T \sum_t \hat{e}_{h,t}\hat{e}_{h,t}'$ is the sample covariance of $\hat{e}_{h,t}$. Then by an appropriate choice of the truncation order h, we have $T^{1/2}(\hat{\Gamma}_{e,h} - \Gamma_e) = O_p(1)$. Defining $D_i(k)$ and $\hat{D}_i(k)$, $i = 1, 2$, as in (7.2.1) with Γ_e and $\hat{\Gamma}_{e,h}$ respectively in place of A, then continuity of the eigenvalues implies that $\hat{D}_i(k) = D_i k + O_p(T^{-1/2})$. Hence for $k \geq q$, $\hat{D}_i(k) = O_p(T^{-1/2})$ or equivalently for $k \geq q$ as $T \to \infty$, $\hat{D}_i(k) < m/T^{1/2-\delta}$ for some positive real m and $0 < \delta < 1/2$ with probability tending to 1. Defining $\mathcal{K}_i = \{k : \hat{D}_i(k) < m/T^{1/2-\delta}\}$, $i = 1, 2$, then $q \in \mathcal{K}_i$ for large T. On the other hand $q - 1$ does not belong to $\mathcal{K}_i$, since $\hat{D}_i(q-1) > c > 0$. Here, the threshold $m/T^{1/2-\delta}$ can be interpreted as the tolerated error induced by sampling variability from estimation of Γ_e.

We can now turn to the main result. Therefore consider the finite autoregression fit of $\hat{\psi}_t^{n,PCA}$ (6.1.26), let $\hat{\Gamma}_{e,h}^n = 1/T \sum_t \hat{e}_{h,t}^n \hat{e}_{h,t}^{n\prime}$ be the sample covariance of the residuals and let R^n be the orthogonal rotation of f_t corresponding to $\hat{\psi}_t^{n,PCA}$. Lastly, let us define $\hat{D}_i^n(k)$, $i = 1, 2$, as in (7.2.1) with $\hat{\Gamma}_{e,h}^n$ in place of A.

Theorem 7.2.1. *Suppose that Assumptions 1.1, 2.1, 3.1, 5.2, 5.3 and 6.1 hold and define for* $i = 1, 2$,

$$\mathcal{K}_i^n = \{k : \hat{D}_i^n(k) < m/\min(n^{1/2-\delta}, T^{1/2-\delta})\},$$

where $0 < m < \infty$ *and* $0 < \delta < 1/2$. *Then for* $i = 1, 2$, $\hat{q}_i^n = \min(k \in \mathcal{K}_i^n)$ *converges to* q *in probability as* $n, T \to \infty$.

Proof. With an appropriate choice of the truncation lag h, we have that $\|\hat{\Gamma}_{e,h}^n - R^{n\prime}\Gamma_e R^n\| =$

$O_p(n^{-1/2}) + O_p(T^{-1/2})$. Then the result follows with the same arguments as used above when f_t was treated as observed. ■

Clearly, in applications values for m and δ have to be chosen. As mentioned before $m/T^{1/2-\delta}$ defines the level of tolerated error, hence by choosing m and δ too large, in a finite sample it is likely that too many eigenvalues fall below the threshold and thus $\hat{q}$ will be smaller than q. Conversely by choosing m and δ too small, $\hat{q}$ is likely to be larger than q. Still, there exist no theoretic results that provide "good" choices for m and δ, however Bai and Ng (2007) propose $m = 1$ and $\delta = 0.1$, that yield good results in a simulation study.

7.3 Specification of the reversed echelon ARMA representation

As mentioned above ARMA model specification has been studied a lot in econometric literature and many methods have been proposed, ranging from graphical inspection or tests of the sample (partial) autocorrelation matrices to procedures that exploit a canonical structure like the echelon form and estimate the structural parameters by means of information criteria (see the references cited above for detailed discussions). Here we will consider the latter approach, more specifically we want to estimate the Kronecker indices $\alpha = (r_1, \dots, r_s)$ that specify the reversed echelon ARMA representation of (f_t). Taking into account that the overall number of possible model specifications ranging from $\alpha = (1, \dots, 1)$ to $\alpha = (P_{\max}, \dots, P_{\max})$, where $P_{\max}$ denotes some predefined upper bound for the AR order, is equal to $(P_{\max})^s$, an exhaustive search will often be prohibitive. Therefore we will consider the efficient search algorithm proposed by Lütkepohl and Poskitt (1996) marginally adapted to the reversed singular echelon representation, which is itself a modified version of the procedure originally proposed by Poskitt (1992). This algorithm exploits the property of reversed echelon forms that the restrictions of the i-th equation imposed by a set of Kronecker indices $\alpha = (r_1, \dots, r_s)$ depend only on the Kronecker indices $r_j \leq r_i$, but not on indices greater than r_i (see equations (3.3.43), (3.3.44)). Using this property, the Kronecker indices are estimated sequentially from smallest to largest using an information criterion. In the context of the GDFM where f_t is unobserved and where the one-step prediction error covariances may be singular, this is admittedly rather an "ad-hoc" approach.

The starting point of the algorithm is the long AR fit (6.1.8), from which the estimates for the one-step prediction errors and the innovations, $\hat{e}_t := \hat{e}_{h_T,t}$ and $\hat{\varepsilon}_t := \hat{\varepsilon}_{h_T,t}$ respectively, are obtained (see (6.1.9) and (6.1.10)). Then one proceeds in the following steps.

(i) For $p = 1, \dots, P_{\max}$ and $i = 1, \dots, s$, regress f_{it} on $(f_{jt} - \hat{e}_{jt})$, $j = 1, \dots, s$, $j \neq i$, and f_{t-h}, $\hat{\varepsilon}_{t-h}$, $h = 1, \dots, p-1$ and determine the residual sum of square $T\hat{\sigma}_i^2(p)$. For $i = 1, \dots, s$ evaluate an information criterion of the form

$$IC_i^T(p) = \log \hat{\sigma}_i^2(p) + d_i(p) C_T / T, \text{ for } p = 1, \dots, P_{\max}, \tag{7.3.1}$$

where $d_i(p)$ denotes the number of regression coefficients to be estimated and C_T will be specified below.

(ii) For $i = 1, \ldots, s$ set

$$\hat{r}_i' = \underset{1 \leq p \leq P_{\max}}{\arg\min}\, IC_i^T(p).$$

(iii) Set the estimate of the smallest Kronecker index to

$$\hat{r}_{i(1)} = \min_j(\hat{r}_j')$$

with $i(1) = \arg\min_j(\hat{r}_j')$.

(iv) For some $u \geq 2$, assume $\hat{r}_{i(1)} \leq \ldots \leq \hat{r}_{i(u-1)}$ are given. For $p = \hat{r}_{i(u-1)}, \ldots, P_{\max}$ and $i \notin \{i(1), \ldots, i(u-1)\}$ regress f_{it} on $(f_{jt} - \hat{e}_{jt})$, $j < i$, $j \notin \{i(1), \ldots, i(u-1)\}$ and $f_{j,t-h}$, $j \notin \{i(1), \ldots, i(u-1)\}$, $h = 1, \ldots, p$ and $\hat{\varepsilon}_{t-h}$, $h = 1, \ldots, p-1$ and $f_{j,t-h}$, $j \in \{i(1), \ldots, i(u-1)\}$, $h = p - \hat{r}_{i(j)} + 1, \ldots, p$ and determine the residual sum of square $T\hat{\sigma}_i^2(p)$. Then evaluate the information criterion

$$IC_i^T(p) = \log \hat{\sigma}_i^2(p) + d_i(p)C_T/T,$$

for $p = \hat{r}_{i(u-1)}, \ldots, P_{\max}$ and those $i \notin \{i(1), \ldots, i(u-1)\}$.

(v) Set the estimate of the u-th smallest Kronecker index to

$$\hat{r}_{i(u)} = \min_i(\arg\min_p IC_i^T(p)),$$

with

$$i(u) = \arg\min_i(\arg\min_p IC_i^T(p)).$$

(vi) Repeat steps (iv) and (v) for $u = 2, \ldots, s$.

In practice, values for C_T and $P_{\max}$ have to be chosen. For $P_{\max}$ Lütkepohl and Poskitt (1996) propose to choose $P_{\max} = \frac{1}{2}h_T$. C_T may be chosen such that the resulting information criterion becomes the AIC or BIC, hence $C_T = 2$ in the case of AIC and $C_T = \log T$ in the case of BIC. As has been shown in Lütkepohl and Poskitt (1996), in the standard ARMA case and for appropriate choices of C_T, this procedure yields consistent estimates for the Kronecker indices. In the GDFM framework of course, in practice the algorithm will not be applied to f_t but to an estimate $\hat{\psi}_t^n$, but providing a theoretical justification of the proposed selection criteria or a theoretically justified adaption goes beyond the scope of this thesis. However, the simulations performed in the next Chapter, provide empirical evidence, that at least from a forecasting perspective both AIC and BIC produce good results.

Chapter 8

Simulation study

In order to evaluate the performance of the estimators proposed, we run a Monte Carlo simulations study. Here we want to compare the predictions based on an ARMA model for the static factors, where the ARMA model is specified using the algorithm of Lütkepohl and Poskitt (1996) described in Section 7.3 and where the ARMA coefficients are estimated using the autoregression-regression method presented in Section 6.1, with the predictions based on the standard AR(1)-model. The simulated data will be generated by models satisfying the standard assumption that the minimal static factor is generated by an AR(1)-model as well as models where this assumption is violated.

The models from which we simulate are similar to those used in the literature (see e.g. Forni *et al.* (2005a) or Stock and Watson (2002a)). Indeed they only differ from standard models in that they allow for the case where f_t is not equivalent to the minimal state vector x_t (see Sections 3.3.2 and 3.3.3). Let us define them below.

Commencing from the static representation (3.3.63), i.e.

$$y_t^n = \bar{\Lambda}^n f_t + u_t^n = \chi_t^n + u_t^n,$$

we define f_t as the vector containing the first $s \leq r$ components of an r-dimensional state vector x_t, i.e. $x_t = (f_t', x_t^{(2)'})'$, and let x_t evolve as

$$x_t = Ax_{t-1} + B\varepsilon_t,$$

where ε_{it}, $i = 1, \ldots, q$ is drawn from a standard Normal distribution and where the elements of A and B are drawn from a uniform distribution with mean 0 and standard deviation 1, i.e.

$$\begin{aligned} A_{i,j} &\sim \mathcal{U}(-\sqrt{3}, \sqrt{3}),\ i,j = 1, \ldots, r \\ B_{i,j} &\sim \mathcal{U}(-\sqrt{3}, \sqrt{3}),\ i = 1, \ldots, r,\ j = 1, \ldots, q. \end{aligned}$$

Analogously the elements of the factor loading matrix $\bar{\Lambda}^n$ are generated as

$$\bar{\Lambda}_{i,j} \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}),\ i = 1, \ldots, n,\ j = 1, \ldots, s.$$

Notice that we only consider stable systems, hence if $\omega_{max}(A) \geq 1$, then A will be divided by some multiple of $\omega_{max}(A)$.

For the noise u_t^n we consider a quite general framework that allows for both a limited amount of serial and cross-sectional correlation, i.e. u_{it} is modeled as a stable AR(1)-process,

$$(1 - \alpha z)u_{it} = v_{it},$$

where $\alpha \sim \mathcal{U}(-0.99, 0.99)$ and v_{it} is allowed to be "groupwise" correlated. Hence let G_i denote the indices of the group v_{it} belongs to, then

$$v_{it} = \begin{cases} c\tilde{v}_{it} + \frac{1-c}{|G_i|-1} \sum_{j \in G_i \setminus i} \tilde{v}_{jt} & \text{if } |G_i| > 1 \\ \tilde{v}_{it} & \text{if } |G_i| = 1, \end{cases}$$

where $\tilde{v}_{it}$ i.i.d. $\mathcal{N}(0, 1)$. Hence the group size $|G_i|$ determines the amount of cross-correlation between noise components. If all $|G_i| = 1$, i.e. if each group only consists of one variable, then the noise components will be mutually uncorrelated.

In the models used for simulations, we fixed the number of dynamic factors as $q = 2$, the state dimension as $r = 6$ and the group size as $|G_i| = 5$ for all $i = 1, \ldots, n$. We generate data for different cross-sectional dimensions, $n = 20, 50, 100$, sample sizes, $T = 100, 200$ and minimal static factor dimensions $s = 6, 5, 4, 3, 2$ resulting in a total number of 30 experiments. For each experiment we indeed generate $T + 100$ data points, and use a rolling window of size T to calculate 100 out-of-sample one-step predictions each based on the preceding T observations, which corresponds to 100 Monte Carlo repetitions per generated model. Then we replicate each experiment 50 times, resulting in a total number of 5000 Monte Carlo repetitions for each experiment.

At each repetition all model parameters are estimated from the sample second moments of the preceding T observations. The static factors and factor loadings are estimated using the static PC method described in Section 5.2.1. The number of dynamic factors q is determined using the Bai-Ng procedure (see Section 7.2) with $m = 1$ and $\delta = 0.1$ and where the order of the long AR fit of the static factor estimates has been set to $h_T = \lfloor log(T) \rfloor$. The echelon ARMA form is specified (i.e. the Kronecker indices are estimated) using the algorithm described in Section 7.3 with $P_{\max} = h_T$ and with AIC and BIC as selection criteria and estimated using the autoregression-regression method presented in Section 6.1. With these estimators we then calculate one-step ahead predictors for the static factors and subsequently for the latent variable. Additionally, at each repetition an AR(1) model for the static factor estimates is fitted and the corresponding predictions are calculated.

To measure the forecast quality, we consider the out-of-sample *coefficient of determination* R^2, i.e. for $i = 1, \ldots, n$ we have

$$R^2(i) = 1 - \frac{\sum_t (\chi_{it} - \hat{\chi}_{it})^2}{\sum_t \chi_{it}^2},$$

which is averaged over i. For each experiment, average R^2-values across replications and their empirical standard deviations (in brackets) are reported in Tables 8.1 - 8.5.

The prediction results can be summarized as follows.

(1) For all values of s, and all methods considered, the forecasting quality improves as n and T increase.

(2) In the case $s = r$, for all n and T, ARMA models specified by means of both AIC and BIC yield a similar forecasting performance as the correctly specified AR(1)-method. Hence, allowing for an ARMA model for the static factors and specifying the structure by means of AIC or BIC does not seem to induce a risk of deteriorating the forecast quality compared to the standard AR(1) model in the case that the AR(1) model is the correctly specified one.

(3) In the case $s < r$, for all n and T, the more general ARMA approach constantly outperforms the standard AR(1) method. This is especially relevant, confirming intuition, the larger the difference between r and s. Comparing AIC and BIC selection, AIC often yields slightly better results, but the difference seems to be negligible. Indeed, the R^2 levels based on ARMA predictions highly depend on n and T, but decrease only slightly as s decreases.

In Tables 8.6 to 8.10 we report the average values across replications and their empirical standard deviations (in brackets) of the estimated AR orders ($\hat{P} = \max(\hat{r}_1, \dots, \hat{r}_s)$) according to AIC and BIC selection, respectively (the estimated MA orders are not reported, since by construction of the echelon form they equal the AR order minus 1) for each experiment. Not surprisingly, the orders estimated by means of AIC tend to be larger than the orders estimated by means of BIC. Further the estimated orders are relatively stable for different cross-sectional dimensions n, but they increase as T and the difference between r and s increase.

Finally, concerning the estimation of the dynamic factor dimension q we can affirm the results of Bai and Ng (2007), in that their testing procedure with $m = 1$ and $\delta = 0.1$ yields fairly accurate results. In our simulation, where q has been set to 2, the average value of $\hat{q}$ (across 150.000 experiments) was 2.03. The quality of the estimation shows low dependence on s, and improves as n, T increase. Tables 8.11 to 8.15 show the empirical distributions of the estimated values for each experiment.

	$T = 100$			$T = 200$		
	AR(1)	AIC	BIC	AR(1)	AIC	BIC
	0.4233	0.4394	0.4400	0.4760	0.5133	0.5083
$n = 20$	(0.0533)	(0.0502)	(0.0517)	(0.0531)	(0.0657)	(0.0586)
	0.5734	0.5477	0.5722	0.6171	0.6147	0.6180
$n = 50$	(0.0558)	(0.0565)	(0.0570)	(0.0452)	(0.0483)	(0.0445)
	0.6365	0.6240	0.6408	0.6556	0.6493	0.6574
$n = 100$	(0.0510)	(0.0529)	(0.0501)	(0.0542)	(0.0564)	(0.0546)

Table 8.1: Average (and standard deviation) of out-of-sample R^2 for $s = r = 6$.

	$T = 100$			$T = 200$		
	AR(1)	AIC	BIC	AR(1)	AIC	BIC
	0.3872	0.4727	0.4580	0.4096	0.5179	0.5138
$n = 20$	(0.0747)	(0.0711)	(0.0775)	(0.0679)	(0.0715)	(0.0782)
	0.5064	0.5729	0.5604	0.5017	0.6034	0.5929
$n = 50$	(0.0773)	(0.0773)	(0.0854)	(0.0582)	(0.0657)	(0.0661)
	0.5357	0.6120	0.5875	0.5385	0.6335	0.6010
$n = 100$	(0.0656)	(0.0520)	(0.0617)	(0.0550)	(0.0527)	(0.0714)

Table 8.2: Average (and standard deviation) of out-of-sample R^2 for $s = 5$.

	$T = 100$			$T = 200$		
	AR(1)	AIC	BIC	AR(1)	AIC	BIC
	0.3482	0.4835	0.4702	0.3435	0.5227	0.5154
$n = 20$	(0.0962)	(0.0859)	(0.0885	0.1181)	(0.0689)	(0.0597)
	0.4034	0.5374	0.5239	0.4345	0.6072	0.5939
$n = 50$	(0.0721)	(0.0652)	(0.0623)	(0.0793)	(0.0683)	(0.0747)
	0.3714	0.6024	0.5694	0.4200	0.6243	0.6002
$n = 100$	(0.0633)	(0.0729)	(0.0776)	(0.0904)	(0.0654)	(0.0733)

Table 8.3: Average (and standard deviation) of out-of-sample R^2 for $s = 4$.

	$T = 100$			$T = 200$		
	AR(1)	AIC	BIC	AR(1)	AIC	BIC
	0.2645	0.4107	0.3938	0.2910	0.5137	0.4857
$n = 20$	(0.0860)	(0.0811)	(0.0802)	(0.0970)	(0.0861)	(0.0857)
	0.3301	0.5572	0.5390	0.3314	0.5657	0.5574
$n = 50$	(0.0970)	(0.0724)	(0.0761)	(0.1153)	(0.0874)	(0.0903)
	0.3091	0.5946	0.5779	0.3579	0.5982	0.5842
$n = 100$	(0.0866)	(0.0750)	(0.0692)	(0.1160)	(0.0964)	(0.1080)

Table 8.4: Average (and standard deviation) of out-of-sample R^2 for $s = 3$.

	$T = 100$			$T = 200$		
	AR(1)	AIC	BIC	AR(1)	AIC	BIC
	0.1876	0.3969	0.3733	0.2191	0.4371	0.4149
$n = 20$	(0.1049)	(0.1183)	(0.1327)	(0.1131)	(0.1412)	(0.1412)
	0.2078	0.4582	0.4324	0.2605	0.4903	0.4672
$n = 50$	(0.0945)	(0.0777)	(0.0950)	(0.1312)	(0.1311)	(0.1387)
	0.2628	0.5315	0.4938	0.2874	0.5528	0.5353
$n = 100$	(0.1714)	(0.1643)	(0.1818)	(0.1218)	(0.1322)	(0.1420)

Table 8.5: Average (and standard deviation) of out-of-sample R^2 for $s = 2$.

	$T = 100$		$T = 200$	
	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$
	2.63	1.56	3.34	1.85
$n = 20$	(0.53)	(0.54)	(0.73)	(0.44)
	2.38	1.21	2.68	1.53
$n = 50$	(0.61)	(0.41)	(0.68)	(0.51)
	2.04	1.12	2.4	1.37
$n = 100$	(0.67)	(0.33)	(0.65)	(0.49)

Table 8.6: Average (and standard deviation) of estimated AR orders for $s = r = 6$.

	$T = 100$		$T = 200$	
	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$
	2.81	1.73	3.32	2.03
$n = 20$	(0.54)	(0.55)	(0.74)	(0.46)
	2.57	1.57	3.12	1.94
$n = 50$	(0.61)	(0.58)	(0.71)	(0.58)
	2.31	1.49	2.7	1.66
$n = 100$	(0.65)	(0.54)	(0.66)	(0.58)

Table 8.7: Average (and standard deviation) of estimated AR orders for $s = 5$.

	$T = 100$		$T = 200$	
	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$
	2.84	1.91	3.42	2.25
$n = 20$	(0.6)	(0.51)	(0.72)	(0.5)
	2.69	1.92	3.15	2.06
$n = 50$	(0.59)	(0.46)	(0.74)	(0.48)
	2.69	2.02	3.02	1.99
$n = 100$	(0.57)	(0.55)	(0.8)	(0.57)

Table 8.8: Average (and standard deviation) of estimated AR orders for $s = 4$.

	$T = 100$		$T = 200$	
	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$
	2.95	2.17	3.59	2.27
$n = 20$	(0.62)	(0.61)	(0.71)	(0.6)
	2.8	2.04	3.62	2.51
$n = 50$	(0.68)	(0.59)	(0.73)	(0.61)
	2.98	2.37	3.29	2.21
$n = 100$	(0.57)	(0.56)	(0.9)	(0.62)

Table 8.9: Average (and standard deviation) of estimated AR orders for $s = 3$.

	$T = 100$		$T = 200$	
	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$	$\hat{P}^{AIC}$	$\hat{P}^{BIC}$
	3.05	2.34	3.60	2.35
$n = 20$	(0.67)	(0.69)	(0.95)	(0.74)
	3.23	2.41	3.58	2.58
$n = 50$	(0.63)	(0.59)	(0.79)	(0.72)
	3.27	2.37	3.63	2.63
$n = 100$	(0.66)	(0.75)	(0.76)	(0.67)

Table 8.10: Average (and standard deviation) of estimated AR orders for $s = q = 2$.

	$T = 100$				$T = 200$			
	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$
$n = 20$	2.7	51.4	44.5	1.3	2.8	51.8	45.3	0.1
$n = 50$	–	90.2	9.8	–	–	100.0	–	–
$n = 100$	1.9	98.1	–	–	–	100.0	–	–

Table 8.11: Percentages of $\hat{q}$ equal to the indicated values for $s = r = 6$.

	$T = 100$				$T = 200$			
	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$
$n = 20$	8.9	71.0	20.1	–	8.2	75.0	16.8	–
$n = 50$	–	99.4	0.6	–	–	95.1	4.7	0.2
$n = 100$	5.2	94.7	0.1	–	–	99.8	0.2	–

Table 8.12: Percentages of $\hat{q}$ equal to the indicated values for $s = 5$.

	$T = 100$				$T = 200$			
	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$
$n = 20$	3.1	57.6	37.6	1.6	5.0	74.5	20.5	–
$n = 50$	7.6	87.4	5.0	–	5.3	88.0	6.7	–
$n = 100$	4.0	88.6	7.3	–	4.9	94.5	0.6	–

Table 8.13: Percentages of $\hat{q}$ equal to the indicated values for $s = 4$.

	$T = 100$				$T = 200$			
	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$
$n = 20$	7.3	79.3	13.3	–	19.2	79.8	0.9	–
$n = 50$	–	95.3	4.7	–	5.0	95.0	–	–
$n = 100$	10.0	74.8	15.3	–	–	100.0	–	–

Table 8.14: Percentages of $\hat{q}$ equal to the indicated values for $s = 3$.

	$T = 100$				$T = 200$			
	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$	$\hat{q} = 1$	$\hat{q} = 2$	$\hat{q} = 3$	$\hat{q} = 4$
$n = 20$	26.2	73.9	–	–	16.2	83.9	–	–
$n = 50$	7.4	92.5	–	–	6.3	93.7	–	–
$n = 100$	10.0	90.0	–	–	0.4	99.6	–	–

Table 8.15: Percentages of $\hat{q}$ equal to the indicated values for $s = 2$.

Chapter 9

Conclusion and future research

The main parts of this thesis are devoted to the modeling of high dimensional time series by means of generalized dynamic factor models. Thus here for the sake of clarity we want to briefly summarize the model building steps presented in this thesis. Below we will then give an outlook on potential future work concerning some open questions raised throughout this thesis.

Given an n-dimensional time series of observations, $y_t^n, t = 1, \ldots, T$, assumed to be generated by a GDFM, the model building steps proposed here are the following.

(i) *Remove the noise part.* Therefore, estimate the static factor dimension by means of an information criterion proposed by Bai and Ng (2002) and described in Section 7.1. Then estimate the latent variables, factor loadings and static factors by one of the methods presented in Section 5.2, e.g. by static PCA.

(ii) *Estimate the fundamental shocks.* On that account, fit a long autoregression to the static factor estimates in order to obtain estimates for the one-step prediction errors, see (6.1.8) in Section 6.1. Then use these estimates and their empirical covariance to determine the dynamic factor dimension by means of the criterion proposed by Bai and Ng (2007), that has been presented in Section 7.2. Extract the fundamental shocks (or the innovations) from the one-step prediction errors applying a static PCA, see (6.1.10) in Section 6.1.

(iii) *Specify and estimate an ARMA model for the static factor estimates.* First, use the static factor estimates and the innovation estimates to determine the Kronecker indices by means of the algorithm of Lütkepohl and Poskitt (1996) as described in Section 7.3. Second, set up a regression structure with the static factor estimates and the innovation estimates to identify the coefficients of the echelon ARMA form corresponding to the estimated Kronecker indices, see (6.1.17) in Section 6.1.

Subsequently to these model building steps, the estimated ARMA model may be used to generate predictions for the static factors, which in turn yield predictions for the latent variable, see Chapter 4. Alternatively and dependently on the purpose of the modeling, the estimated ARMA model may also be used for structural modeling, e.g. by means of an *impulse response*

analysis, describing how the fundamental shocks propagate to the observations, see for instance Forni *et al.* (2005b), which has not been explicitly treated in this thesis.

The model building process considered here is data driven and may be implemented as a fully automatic procedure, which simplifies its practical application. Furthermore it is more flexible than comparable procedures proposed in the literature, in that it does not require the static factor to follow an AR(1) process (see for instance Forni *et al.* (2005a) or Stock and Watson (1998)) or a higher order regular[1] AR process (see Doz *et al.* (2007)), but allows for general, possibly singular ARMA processes. Even when the factors have a finite order AR representation (see Section 3.3.6), estimating the echelon ARMA form may be advantageous compared to AR estimation in that it involves a smaller number of parameters to be estimated and may thus yield more precise forecasts (see for instance Lütkepohl and Poskitt (1996)). Therefore and as suggested by the simulation results, the proposed model building process may provide a substantial improvement in prediction accuracy when the static factors cannot be represented by an AR(1) process.

However regarding the framework presented in this thesis, there is still a number of open problems, some of which we want to point out in the sequel.

- We have proposed to estimate the ARMA model for the static factors using an autoregressive-regression approach. This approach is known to be inefficient, so one open problem is to find an efficient method, for instance using a ML-approach.
- The Kronecker indices specifying the ARMA model have been estimated using the algorithm of Lütkepohl and Poskitt (1996), for which so far no theoretical justification exists taking into account that the output variable is unobserved and may have singular one-step prediction errors.
- An alternative to the model building process described above, would be to directly estimate the state space representation of the latent variables, hence finding consistent (and efficient) estimates for (A, B, C^n) is another open problem.
- Finally, one could think about weakening Assumptions 1.1 e) and 3.1, in that the minimal static factor dimension and the minimal state dimension, respectively, are allowed to grow as the cross-sectional dimension increases.

[1] In the sense that the transfer function is square, i.e. $q = s$.

Appendix A

Some useful results on (generalized) eigenvalue problems

A.1 Eigenvalues and eigenvectors of complex matrices and related results

Let Γ denote a complex $(n \times n)$-dimensional square matrix. A complex number ω_j is called an *eigenvalue*, if there exists an n-dimensional vector $o_j \neq 0$, such that

$$\Gamma o_j = \omega_j o_j.$$

The vector o_j is then called an *eigenvector* of Γ. The eigenvalues of Γ are obtained as the (complex) roots of the polynomial in ω, $\det(\Gamma - \omega I_n)$. This is because, if ω_j is a root of $\det(\Gamma - \omega I_n)$, the columns of $(\Gamma - \omega_j I_n)$ are linearly dependent and there exists an $(n \times 1)$ vector o_j, such that $(\Gamma - \omega_j I_n)o_j = 0$, hence $\Gamma o_j = \omega_j o_j$. Obviously eigenvectors are not unique, since any non-zero scalar multiples are also eigenvectors. However, normalized eigenvectors, i.e. eigenvectors of length 1, corresponding to distinct eigenvalues are unique up to multiplication by $e^{i\lambda}$, $\lambda \in [-\pi, \pi]$.

Throughout this thesis the following "facts" about eigenvalues are frequently used (see for instance Lütkepohl (2005)).

a) $\det(\Gamma) = \prod_{j=1}^{n} \omega_j$.

b) $\mathrm{tr}(\Gamma) = \sum_{j=1}^{n} \omega_j$.

c) If Γ is Hermitian, then its eigenvalues are real numbers. If Γ is Hermitian and non-negative definite, then its eigenvalues are real and ≥ 0.

d) If ω_i and ω_j are distinct eigenvalues of Γ, then the corresponding eigenvectors o_i and o_j are orthogonal, i.e. $o_i^* o_j = 0$.

e) *Canonical decomposition* of a Hermitian matrix. If Γ is Hermitian, it can be decomposed as

$$\Gamma = O\Omega O^* = O \begin{pmatrix} \omega_1 & & 0 \\ & \ddots & \\ 0 & & \omega_n \end{pmatrix} O^*, \tag{A.1.1}$$

where ω_j are the eigenvalues of Γ and the columns of O are the corresponding normalized eigenvalues, hence $O^*O = OO^* = I_n$ and

$$O^*\Gamma O = \Omega.$$

f) *Square root* of a non-negative definite Hermitian matrix. If Γ is a non-negative definite Hermitian matrix, then we can define $\Gamma^{1/2}$ as

$$\Gamma^{1/2} = O\Omega^{1/2}O^* = O \begin{pmatrix} \sqrt{\omega_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\omega_n} \end{pmatrix} O^*, \tag{A.1.2}$$

and $\Gamma^{1/2}\Gamma^{1/2} = \Gamma$.

Next we are going to assemble some useful results. The following lemma, known as the *Courant-Fisher Theorem* (see Brillinger (1981), exercise 3.10.16), relates the eigenvalues of a matrix to an objective function.

Lemma A.1.1. *Let Γ be an $(n \times n)$ Hermitian matrix and let $\omega_j, j = 1, \ldots, n$, be its eigenvalues in descending order of magnitude. Then the eigenvalue ω_j is the solution of*

$$\begin{aligned} &\min_{D_j} \max_{o} o^*\Gamma o \\ s.t. \quad &o^*o = 1, D_j^*o = 0, \end{aligned} \tag{A.1.3}$$

where D_j is an $(n \times (j-1))$ complex matrix of full column rank for $j = 2, \ldots, n$ and the $(n \times 1)$ null matrix for $j = 1$ and o is an n vector.

The next corollary summarizes some important consequences of the Courant-Fisher Theorem, that are useful in dealing with the eigenvalues of sums of matrices.

Corollary A.1.1. *Let E and F be $(n \times n)$ Hermitian non-negative definite matrices and $\Gamma = E + F$, and let $\omega_{\Gamma j}$, ω_{Ej} and ω_{Fj}, $j = 1, \ldots, n$, be their eigenvalues in descending order of magnitude. Then for any $j = 1, \ldots, n$*

a) $\omega_{\Gamma j} \geq \omega_{Ej}, \quad \omega_{\Gamma j} \geq \omega_{Fj}$

b) $\omega_{\Gamma j} \leq \omega_{Ej} + \omega_{F1}, \quad \omega_{\Gamma j} \leq \omega_{E1} + \omega_{Fj}.$

Proof. Let an $(n \times (j-1))$ complex matrix D_j of full column rank be given, then for $D_j^* o = 0$ and $o^* o = 1$, applying the Courant-Fisher theorem we have

$$\begin{aligned} \max_o o^*\Gamma o &= \max_o o^*(E+F)o \geq \max_o o^* E o, \text{and} \\ \max_o o^*\Gamma o &= \max_o o^*(E+F)o \geq \max_o o^* F o, \end{aligned}$$

which proves statement a).

Furthermore we have

$$\begin{aligned} \max_o o^*\Gamma o &\leq \max_o o^* E o + \max_o o^* F o \leq \max_o o^* E o + \omega_{F1}, \text{and} \\ \max_o o^*\Gamma o &\leq \max_o o^* E o + \max_o o^* F o \leq \omega_{E1} + \max_o o^* F o, \end{aligned}$$

which proves b). ■

The following result (see Brillinger (1981), Theorem 3.7.4) shows how to approximate a matrix Γ by another matrix A of lower rank in order to minimize the eigenvalues of the squared residual matrix.

Lemma A.1.2. *Let Γ be an $(n \times n)$ Hermitian matrix with eigenvalues ω_j, $j = 1, \ldots, n$. The $(n \times n)$ matrix A that minimizes the j-th largest eigenvalue of*

$$(\Gamma - A)(\Gamma - A)^*$$

among all $(n \times n)$ matrices A with $rk(A) = q$ is given by

$$A = O_1 \Omega_1 O_1^*,$$

where Ω_1 denotes the diagonal matrix containing the q largest eigenvalues of Γ in its diagonal and O_1 denotes the $n \times q$ matrix of corresponding normalized eigenvectors. Then the minimum achieved is ω_{q+j}^2.

Proof. Let $B = \Gamma - A$ and let $\omega_j(BB^*)$ denote the j-th largest eigenvalue of BB^*. Applying the Courant-Fisher theorem, we get that

$$\omega_j(BB^*) = \min_D \max_{D^* x = 0} x^* B B^* x. \tag{A.1.4}$$

Hence for a fixed $(n \times (j-1))$ complex matrix D of full column rank (for $j = 1$, D is the $(n \times 1)$ null vector), we have

$$\begin{aligned} \omega_j(BB^*) &\geq \max_{D^* x = 0} x^* B B^* x \\ &\geq \max_{D^* x = 0, A^* x = 0} x^* B B^* x \\ &= \max_{D^* x = 0, A^* x = 0} x^* \Gamma \Gamma^* x \\ &\geq \omega_{q+j}^2, \end{aligned} \tag{A.1.5}$$

since we have at most $q+j-1$ zero restrictions. Setting $A = O_1 \Omega_1 O_1^*$ and denoting by Ω_1 and Ω_2 the diagonal matrices containing the q largest and $(n-q)$ smallest eigenvalues of Γ, respectively,

arranged in descending order of magnitude and $O_1 = (o_1 \dots o_q)$ and $O_2 = (o_{q+1} \dots o_n)$ are the $(n \times q)$- and $n \times (n-q)$-dimensional orthogonal matrices, respectively, of corresponding normalized eigenvectors, we get that

$$(\Gamma - A)(\Gamma - A)^* = \Gamma\Gamma^* - O_1\Omega_1^2 O_1^* = O_2\Omega_2^2 O_2^*, \tag{A.1.6}$$

and hence equality in (A.1.5) is indeed achieved. ■

A.2 Generalized eigenvalues and eigenvectors

Let Γ and Σ denote complex $(n \times n)$-dimensional square matrices. A complex number ν_j is called a *generalized eigenvalue* of Γ with respect to Σ, if there exists an n-dimensional vector $v_j \neq 0$, such that

$$\Gamma v_j = \nu_j \Sigma v_j.$$

The vector v_j is then called a *generalized eigenvector* of the couple of matrices (Γ, Σ). The generalized eigenvalues of (Γ, Σ) are obtained as the (complex) roots of the polynomial in ν, $\det(\Gamma - \nu\Sigma)$; since if ν_j is a root of $\det(\Gamma - \omega\Sigma)$, the columns of $(\Gamma - \nu_j\Sigma)$ are linearly dependent and there exists an $(n \times 1)$ vector v_j, such that $(\Gamma - \nu_j\Sigma)v_j = 0$ and hence $\Gamma v_j = \nu_j \Sigma v_j$. Again generalized eigenvectors are not unique, but usually they are rescaled, such that $v_j^*\Sigma v_j = 1$, and hence, if the corresponding generalized eigenvalues are distinct, v_j is uniquely determined up to multiplication by $e^{i\lambda}$, $\lambda \in [-\pi, \pi]$. Obviously, if $\Sigma = I_n$, then the generalized eigenvalues and eigenvectors are equal to the eigenvalues and eigenvectors of Γ.

Generalized eigenvalue decomposition of Hermitian (semi)-positive definite matrices

Suppose Γ and Σ are Hermitian semi-positive definite and positive definite $(n \times n)$-dimensional matrices and let Δ denote the $(n \times n)$ diagonal matrix containing the generalized eigenvalues of (Γ, Σ) in its diagonal and $V = (v_1, \dots, v_n)$ the matrix of corresponding normalized generalized eigenvectors. Then $V^*\Sigma V = I$ and

$$\begin{aligned} \Gamma V &= \Sigma V \Delta, \\ \Gamma V &= \Sigma^{1/2}\Sigma^{1/2} V \Delta, \end{aligned} \tag{A.2.1}$$

and hence

$$\Sigma^{-1/2}\Gamma\Sigma^{-1/2}\Sigma^{1/2}V = \Sigma^{1/2}V\Delta. \tag{A.2.2}$$

Defining $\bar{V} = \Sigma^{1/2}V$, we get that $\bar{V}^*\bar{V} = I_n$ and

$$\Sigma^{-1/2}\Gamma\Sigma^{-1/2}\bar{V} = \bar{V}\Delta. \tag{A.2.3}$$

Hence the generalized eigenvalues of (Γ, Σ) are the eigenvalues of $\bar{\Gamma} := \Sigma^{-1/2}\Gamma\Sigma^{-1/2}$ and the (normalized) generalized eigenvectors of (Γ, Σ) are the (normalized) eigenvectors of $\bar{\Gamma}$ left multiplied by $\Sigma^{-1/2}$. From the canonical decomposition of $\bar{\Gamma}$ we derive the following decomposition

of Γ:

$$\begin{aligned} \Gamma &= \Sigma^{1/2}\bar{V}\Delta\bar{V}^*\Sigma^{1/2} \\ &= \Sigma V\Delta V^*\Sigma. \end{aligned} \tag{A.2.4}$$

As a consequence of the relation between the generalized eigenvalue problem of (Γ, Σ) and the eigenvalue problem of $\Sigma^{-1/2}\Gamma\Sigma^{-1/2}$, Lemma A.1.1 can be extended to the following result.

Lemma A.2.1. *Let Γ and Σ be Hermitian semi-positive definite and positive definite $(n \times n)$-dimensional matrices, and denote the generalized eigenvalues and eigenvectors of (Γ, Σ) by $\delta_1, \ldots, \delta_n$ and $v_1, \ldots, v_n$ respectively. Then the j-th generalized eigenvector v_j, $j = 1, \ldots, n$, is the solution of*

$$\begin{aligned} &\arg\max_{b\in\mathbb{C}^n} b^*\Gamma b \\ &\text{s.t.} \quad b^*\Sigma b = 1 \text{ and for } \; j \geq 2, \; b^*\Sigma b_i = 0, i = 1, \ldots, j-1 \end{aligned} \tag{A.2.5}$$

and the maxima achieved are $v_j^*\Gamma v_j = \delta_j$.

Appendix B

Frequently used notation and acronyms

This section summarizes frequently used notation.

Acronyms

AIC: Akaike information criterion
AR: Autoregressive
ARMA: Autoregressive moving average
ARX: Autoregressive with exogenous variables
BIC: Bayesian information criterion
DAS: Dynamic averaging sequence
FHLR: Generalized principal component analysis as proposed by Forni *et al.* (2005a)
GDFM: Generalized dynamic factor model
GLS: Generalized least squares
GPCA: Generalized principal component analysis
IC: Information criterion
MA: Moving average
MFD: Matrix fraction description
ML: Maximum likelihood
OLS: Ordinary least squares
PC: Principal component
PCA: Principal component analysis
PE: Prediction error

The subsequent list only contains symbols that are regularly used throughout this thesis.

Dimensions

n: cross-sectional dimension of the observable process.
q: dynamic factor dimension, i.e. normal rank of f_y^n.
r: minimal state dimension of an (A, B, C^n) state space representation of the latent variables.
s: static factor dimension, i.e. rank of Γ_y^n.
T: number of observations.

Other integer valued variables

P, Q: orders of the reversed echelon ARMA representation of the static factors.
$\alpha = (r_1, \ldots, r_s)$: Kronecker indices corresponding to $k(z)$.

Random variables

ε_t: fundamental shocks, innovations of the latent variables, (see (3.3.2)).
e_t^n: residuals of the projection of y_t^n onto the set of all aggregates, (see (3.1.2)).
$e_{h,t}$: residuals of the regression of f_t onto $f_{t-1}, \ldots, f_{t-h}$, (see (6.1.4)).
f_t: minimal static factor.
γ_t^n: projection of y_t^n onto the set of all aggregates (see (3.1.2)).
ψ_t, ψ_t^n: vector containing the principal components of $y_t^{(n)}$.
u_t, u_t^n: noise in the factor model representation.
$\tilde{u}_t$, $\tilde{u}_t^n$: noise in the PC model.
x_t: minimal state vector (see (3.3.13).
ξ_t: dynamic factor.
χ_t, χ_t^n: latent variable.
$\tilde{\chi}_t$, $\tilde{\chi}_t^n$: latent variable in the PC model.
y_t, y_t^n: observations.

Factor loadings and transfer functions

$k(z) = \sum_{j=0}^{\infty} K_j z^j$: transfer function from ε_t to f_t (see 3.3.32).
$\tilde{k}(z)$: strictly proper transfer function corresponding to $k(z)$ (see 3.3.36).
$\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j, \Lambda^n(z) = \sum_{j=-\infty}^{\infty} \Lambda_j^n z^j$: factor loadings corresponding to ξ_t.
$\bar{\Lambda}$: static factor loading matrix in the quasi-static representation (3.3.63).
$w^n(z) = \sum_{j=0}^{\infty} W_j^n z^j$: factor loadings corresponding to ε_t (see 3.3.2).

State space, ARMA and AR-realizations

(A, B, C^n): system matrices of a state space realization of $w^n(z)$ (see Section 3.3.2).
(a, b): ARMA realization of $k(z)$ with $a(z) = A_0 - \sum_{j=1}^{P} A_j z^j$, $b(z) = \sum_{j=0}^{Q} B_j z^j$.

$(\tilde{a}, \tilde{b})$: ARMA realization of $\tilde{k}(z)$ (see Section 3.3.4).
$\mathscr{C}_r = (B, AB, \ldots, A^{r-1}B)$: controllability matrix.
$\mathscr{H}_\infty^n, \mathscr{H}_\infty(k)$: Hankel matrices of w^n and k, respectively.
$h^n(i,j), h(i,j)$: j-th row in the i-th block row of $\mathscr{H}_\infty^n$ and $\mathscr{H}_\infty(k)$, respectively.
$\mathscr{O}_r = (C^{n\prime}, A'C^{n\prime}, \ldots, A^{r-1\prime}C^{n\prime})'$: observability matrix.
$\phi(z) = I_s - \sum_{k=1}^{\infty} \Phi_k z^k$: AR representation of f_t (see (3.3.69)).
$\tau = \mathrm{vec}(I_s - A_0, \ldots, A_P, B_1, \ldots, B_Q)$: vector of stacked ARMA coefficients.
τ_α: vector containing the free parameters in τ corresponding to the Kronecker indices α.

Vector spaces

$\mathscr{A}(y)$: set of all aggregates corresponding to (y_t^n), $n \in \mathbb{N}$ (see Definition 3.1.1).
$\mathbb{H}_\xi$: Hilbert space spanned by ξ_{it}, $i = 1, \ldots, q$, $t \in \mathbb{Z}$.
$\mathbb{H}_y, \mathbb{H}_\chi$: Hilbert space spanned by y_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$ and χ_{it}, $i \in \mathbb{N}$, $t \in \mathbb{Z}$, respectively.
$\mathbb{H}_\varepsilon(t^-)$: Hilbert space spanned by $\varepsilon_{i\tau}$, $i = 1, \ldots, q$, $\tau \leq t$.
$\mathbb{H}_y^n(t^-), \mathbb{H}_\chi^n(t^-)$: Hilbert space spanned by $y_{i\tau}$, $i \in \mathbb{N}$, $\tau \leq t$ and $\chi_{i\tau}$, $i \in \mathbb{N}$, $\tau \leq t$, respectively.
$\mathbb{K}_f(t)$: Hilbert space spanned by $f_{i\tau}$, $i = 1, \ldots, s$, $\tau = 1, \ldots, t$.
$L_2^n([-\pi, \pi], \mathbb{C}, f_y^n)$: frequency domain of (y_t^n) (see Section 2.1).

Spectral densities and covariances

f_u, f_u^n: spectral density of (u_t), (u_t^n).
f_ξ: spectral density of (ξ_t).
f_χ, f_χ^n: spectral density of (χ_t), (χ_t^n).
f_y, f_y^n: spectral density of (y_t), (y_t^n).
$\tilde{f}_u, \tilde{f}_u^n$: spectral density of the PC noise $(\tilde{u}_t)$, $(\tilde{u}_t^n)$.
$\tilde{f}_\chi, \tilde{f}_\chi^n$: spectral density of the PC latent variable $(\tilde{\chi}_t)$, $(\tilde{\chi}_t^n)$.
Γ_f: covariance matrix of (f_t). Γ_u, Γ_u^n: covariance matrix of (u_t), (u_t^n).
Γ_ξ: covariance matrix of (ξ_t).
$\Gamma_\chi, \Gamma_\chi^n$: covariance matrix of (χ_t), (χ_t^n).
Γ_y, Γ_y^n: covariance matrix of (y_t), (y_t^n).
$\tilde{\Gamma}_u, \tilde{\Gamma}_u^n$: covariance matrix of the PC noise $(\tilde{u}_t)$, $(\tilde{u}_t^n)$.
$\tilde{\Gamma}_\chi, \tilde{\Gamma}_\chi^n$: covariance matrix of the PC latent variable $(\tilde{\chi}_t)$, $(\tilde{\chi}_t^n)$.

Eigenvalues and eigenvectors for (dynamic) PCA estimation

$\mu_i^n, \mu_{\chi,i}^n, \mu_{u,i}^n$: eigenvalues of Γ_y^n, Γ_χ^n and Γ_u^n.
$\omega_i^n, \omega_{\chi,i}^n, \omega_{u,i}^n$: eigenvalues of f_y^n, f_χ^n and f_u^n.
Ω_1^n, Ω_2^n: diagonal matrices containing the s largest and $n-s$ smallest eigenvalues of Γ_y^n in their diagonal, arranged in descending order of magnitude.
$\Omega_1^n(\lambda), \Omega_2^n(\lambda)$: diagonal matrices containing the q largest and $n-q$ smallest eigenvalues of $f_y^n(\lambda)$

in their diagonal, arranged in descending order of magnitude.
O_1^n, O_2^n: matrices of eigenvectors corresponding to Ω_1^n and Ω_2^n, respectively.
$O_1^n(e^{-i\lambda}), O_2^n(e^{-i\lambda})$: matrices of eigenvectors corresponding to $\Omega_1^n(\lambda)$ and $\Omega_2^n(\lambda)$, respectively.

Generalized eigenvalues and eigenvectors for GPCA estimation

Δ_1^n, Δ_2^n: diagonal matrices containing the s largest and $n-s$ smallest generalized eigenvalues of (Γ_y^n, Γ_u^n) in their diagonal, arranged in descending order of magnitude.
ν_i^n: generalized eigenvalues of (Γ_y^n, Γ_u^n).
V_1^n, V_2^n: matrices of eigenvectors corresponding to Δ_1^n and Δ_2^n, respectively.

Generalized eigenvalues and eigenvectors for FHLR estimation

D_1^n, D_2^n: diagonal matrices containing the s largest and $n-s$ smallest generalized eigenvalues of $(\tilde{\Gamma}_\chi^n, \tilde{\Gamma}_u^n)$ in their diagonal, arranged in descending order of magnitude.
$\tilde{\nu}_i^n$: generalized eigenvalues of $(\tilde{\Gamma}_\chi^n, \tilde{\Gamma}_u^n)$.
W_1^n, W_2^n: matrices of eigenvectors corresponding to D_1^n and D_2^n, respectively.

Bibliography

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *J. of Business and Economic Statistics*, **25**, 91–96.

Anderson, B. and Deistler, M. (2008). Properties of zero-free transfer function matrices. *SICE Journal of Control, Measurement, and System Integration*. To appear.

Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics*, **12**, 1–45.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **V**, 111–150.

Apostol, T. M. (1974). *Mathematical Analysis*. Addison-Wesley.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, **71(1)**, 135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70(1)**, 191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *J. of Business and Economic Statistics*, **25**, 52–60.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British J. of Psychology*, **28**, 97–104.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *J. of Econometrics*, **127(1)**, 169–194.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Holden Day, San Francisco.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

Burt, C. (1909). Experimental tests of general intelligence. *British J. of Psychology*, **3**, 94–177.

Chamberlain, G. (1983). Funds, factors and diversification in arbitrage pricing models. *Econometrica*, **51(5)**, 1305–1324.

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

Deistler, M. (2001a). Multivariate linear systems. In D. Pena, G. C. Tiao, and R. S. Tsay, editors, *A Course in Time Series Analysis*. Wiley series in probability and statistics.

Deistler, M. (2001b). System identification - General aspects and structure. In G. Goodwin, editor, *System Identification and Adaptive Control (Festschrift for B.D.O. Anderson)*. Springer, London.

Deistler, M. and Hamann, E. (2005). Identification of factor models for forecasting returns. *J. of Financial Econometrics*, **3**, 256–281.

Donald, S. G. (1997). Inference concerning the number of factors in a multivariate nonparametric relationship. *Econometrica*, **65**(1), 103–132.

Doz, C., Giannone, D., and Reichlin, L. (2007). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. CEPR Discussion Papers 6043.

Engle, R. F. and Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rates. *J. of the American Statistical Association*, **76**, 774–781.

Findley, D. F., Pötscher, B. M., and Wei, C. Z. (2004). Modeling of time series arrays by multistep prediction or likelihood methods. *J. of Econometrics*, **118**, 151–187.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric Theory*, **17**, 1113–1141.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economic Studies*, **65**, 453–473.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005a). The generalized dynamic factor model: one-sided estimation and forecasting. *J. of the American Statistical Association*, **100**, 830–840.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2005b). Opening the black box: structural factor models versus structural VARs. *Manuscript, Universite Libre de Bruxelles*.

Geweke, J. (1977). The dynamic factor analysis of economic time series models. In D. J. A. et al., editor, *Latent Variables in Socio-Economic Models*, pages 365–383. North-Holland Amsterdam.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *J. of the American Statistical Association*, **102**, 603–617.

Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

Hannan, E. J. and Kavalieris, L. (1986). Regression, autoregression models. *J. of Time Series Analysis*, **7**, 27–49.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of ARMA order. *Biometrika*, **69**, 81–94.

Heij, C., Scherrer, W., and Deistler, M. (1997). System identification by dynamic factor models. *SIAM J. on Control and Optim.*, **35**(6), 1924–1951.

Hillmer, S. C. and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *J. of the American Statistical Association*, **74**, 652–660.

Hotelling, H. (1933). Analysis of a complex of statistical variables with principal components. *J. of Educational Psychology*, **24**, 417–441.

Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *SIAM J. on Control and Optim.*, **1**, 152–192.

Koreisha, S. and Pukkila, T. (1989). Fast linear estimation methods for autoregressive moving-average models. *J. of Time Series Analysis*, **10**, 325–339.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*. Butterworth & Co., 2nd ed.

Lewbel, A. (1991). The rank of demand systems: theory and nonparametric estimation. *Econometrica*, **59**(3), 711–730.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice Hall, New Jersey.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

Lütkepohl, H. and Poskitt, D. (1996). Specification of echelon-form VARMA models. *J. of Business and Economic Statistics*, **14**, 69–79.

Mayne, D. Q., Aström, K. J., and Clark, J. M. C. (1984). A new algorithm for recursive estimation of parameters in controlled ARMA processes. *Automatica*, **20**(6), 751–760.

Molenaar, P. C. M. (1987). Dynamic factor analysis in the frequency domain: Causal modeling of multivariate psychophysiological time series. *Multivariate Behavioral Research*, **22**, 329–353.

Nsiri, S. and Roy, R. (1992). On the identification of ARMA echelon-form models. *Canadian Journal of Statistics*, **20**(4), 369–386.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, **2**, 559–572.

Poskitt, D. (1992). Identification of echelon canonical forms for vector linear processes using least squares. *The Annals of Statistics*, **20**, 195215.

Reinsel, G. C. (1993). *Elements of Multiple Time Series Analysis*. Springer, New York.

Royden, H. L. (1988). *Real Analysis*. Macmillan, New York.

Rozanov, Y. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modelling without pretending to have too much a priori economic theory. In C. A. Sims, editor, *New Methods in Business Cycle Research*. Minneapolis: Federal Reserve Bank of Minneapolis.

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

Spearman, C. (1904). General intelligence, objectively determined and measured. *American J. of Psychology*, **15**, 201–203.

Stock, J. H. and Watson, M. (1998). Diffusion indexes. *NBER Working Paper 6702*.

Stock, J. H. and Watson, M. (2002a). Forecasting using principal components from a large number of predictors. *J. of the American Statistical Association*, **97**, 1167–1179.

Stock, J. H. and Watson, M. (2002b). Macroeconomic forecasting using diffusion indexes. *J. of Business and Economic Statistics*, **20**, 147–162.

Stock, J. H. and Watson, M. (2005). An empirical comparison of methods for forecasting using many predictors. *Mimeo, Princeton University*.

Thomson, G. H. (1951). *The Factorial Analysis of Human Ability*. London University Press.

Thurstone, L. L. (1932). The vectors of the mind. *Psychological Review*, **41**, 1–32.

Tiao, G. (2001). Vector ARMA models. In D. Pena, G. C. Tiao, and R. S. Tsay, editors, *A Course in Time Series Analysis*. Wiley series in probability and statistics.

Tiao, G. and Tsay, R. (1989). Model specification in multivariate time series (with discussions). *J. of the Royal Statistical Society B*, **51**, 157213.

Tyrtyshnikov, E. E. (1997). *A Brief Introduction to Numerical Analysis*. Birkhaeuser Boston.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *J. of Econometrics*, **23**, 385–400.

White, H. (2000). A reality check for data snooping. *Econometrica*, **68(5)**, 1097–1126.

Curriculum Vitae

Christiane Zinner

Date of Birth:	November 12, 1981
Place of Birth:	Ried, Austria
Citizenship:	Austrian
Marital Status:	Single

Education

05/2000	Matura (corresponding to A-levels) at Bundesgymnasium Stockerau.
10/2000 - 03/2005	M.Sc. in Technical Mathematics with emphasis on Mathematical Methods in Economics, at Vienna University of Technology. First and second diploma passed with distinction, Master thesis: *Wavelet-basierte Trendschätzung einer endlichen Zeitreihe* (Wavelet based trend estimation of a finite time series).
09/2003- 06/2004	Studies of Applied Mathematics at ENSTA (Ecole Nationale Supérieure de Techniques Avancées), Paris.
Since 03/2005	Ph.D. studies in Technical Mathematics at Vienna University of Technology, Subject matter: Modeling of high-dimensional time series by generalized dynamic factor models.

Professional career

03/2004 - 06/2004	BNP Paribas Asset Management, Paris. Internship
11/2004 - 03/2007	Vienna University of Technology. Project assistant
11/2004 - 08/2007	Siemens AG Österreich Fin4Cast. Project manager in the areas of quantitative analysis and financial forecasting.
Since 02/2008	University of Vienna. Project assistant