DISSERTATION

Reconciling the Theory of Factor Sequences

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter Leitung von Em.O.Univ.Prof. Dipl.-Ing.

Dr. Techn. Manfred Deistler
Inst.Nr.: E105
Institut für Wirtschaftsmathematik

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

MSc with Honors, BSc, BA Philipp Gersing
[redacted]

Wien, am 21.09.2023

Danksagung

Ich bedanke mich herzlichst bei meinem Betreuer Prof. Manfred Deistler: Für die großartige Unterstützung, die vielen interessanten Gespräche und dafür, dass er mich auf die wissenschaftliche Bedeutung der Grundidee dieser Arbeit aufmerksam gemacht hat, sodass ich diese weiter ausbauen konnte.
Mein besonderer Dank gilt auch meinen Eltern, die mich mein Leben lang mit ihren besten Kräften bei meinen Projekten, so auch bei meiner Dissertation unterstützt haben. Ich bedanke mich bei meinem besten Freund und Kollegen Dr. Christoph Rust für die fachliche Zusammenarbeit, und die Hilfe bei der Software-Programmierung und Erstellung der Simulationen in dieser Arbeit.
Ein großer Dank gilt auch Prof. Marco Lippi (Einaudi Institute for Economics and Finance), Prof. Sylvia Frühwirth-Schnatter (WU Wien), Prof. Brian Anderson (ANU Canberra), und Dr. Leopold Sögner (Institut für höhere Studien Wien) für die wertvollen fachlichen Anregungen.
Desweiteren möchte ich mich bei meiner Ex-Frau Zita Barcza-Szabó und bei Bettina Sommer, sowie bei meinem Bruder und meinen Freunden für die Unterstützung bedanken.

Supported by the DOC-Fellowship of the Austrian Academy of Sciences (ÖAW) and the Austrian Central Bank under Anniversary Grant No. 18287.

Philipp Gersing
Sardinien, am 20.09.2023

Deutsche Kurzfassung

Faktorfolgen sind stochastische Prozesse $(y_{it} : i \in \mathbb{N}, t \in \mathbb{Z})$, die in Zeitachse t und Querschnitt i indiziert sind und eine sogenannte Faktorstruktur aufweisen. Dynamische Faktorfolgen wurden von Forni and Lippi (2001) eingeführt. In dieser Arbeit wird die Unterscheidung zwischen dynamischen und statischen Faktorfolgen eingeführt, die wir mit den Modellen in Chamberlain and Rothschild (1983a); Bai and Ng (2002); Stock and Watson (2002a) identifizieren. Der Unterschied besteht darin, was wir die *schwache gemeinsame Komponente* nennen. Diese wird von potenziell unendlich vielen *schwachen Faktoren* aufgespannt. Wie in dieser Arbeit gezeigt wird, kann das Ignorieren dieser schwachen gemeinsamen Komponente erhebliche Folgen für die Anwendung von Faktormodellen haben - sowohl in der Strukturanalyse, als auch bei Prognosen. Wir zeigen auch, dass die dynamische gemeinsame Komponente einer dynamischen Faktorfolge unter recht allgemeinen Bedingungen, den Beobachtungen kausal untergeordnet ist. Folglich sollte nur die dynamische gemeinsame Komponente als Projektion auf die gemeinsamen strukturellen Schocks der Wirtschaft interpretiert werden, während die statische gemeinsame Komponente ausschließlich die *gleichzeitige* gemeinsame Variation erfasst.

Abstract

Factor Sequences are stochastic double sequences $(y_{it} : i \in \mathbb{N}, t \in \mathbb{Z})$ indexed in time and cross-section which have a so called factor structure. The name was coined by Forni and Lippi (2001) who introduced dynamic factor sequences. We introduce the distinction between dynamic- and static factor sequences. The latter are the most common workhorse model of econometric factor analysis building on Chamberlain and Rothschild (1983a); Stock and Watson (2002a); Bai and Ng (2002). The difference consists in what we call the *weak common component* which is spanned by (potentially infinitely many) weak factors. We illustrate that ignoring the weak common component can have substantial consequences for applications of factor models in structural analysis and forecasting. We also show that the dynamic common component of a dynamic factor sequence is causally subordinated to the output under quite general conditions. As a consequence only the dynamic common component can be interpreted as the projection on the infinite past of the common innovations of the economy, i.e. the part which is *dynamically common.* On the other hand the static common component captures the *contemporaneous* co-movement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With the increasing availability of high-dimensional time series data, also the demand for methods to analyse and forecast such data has been growing which has led to the high popularity of factor analysis in macro-econometrics: In factor analysis, we commence from considering such a high-dimensional time series as a double indexed (zero-mean stationary, usually real-valued) stochastic process $(y_{it} : i \in \mathbb{N}, t \in \mathbb{Z}) =: (y_{it})$, where the index i stands for an infinitely growing cross-section and t for time observations. The most common factor model is in a certain sense "static" and of the form

$$y_{it} = C_{it} + e_{it} = \Lambda_i F_t + e_{it} \ , \tag{1.1}$$

where (C_{it}) is called the "common component" and (e_{it}) is called the "idiosyncratic component". The process F_t is a "low"-dimensional $r \times 1$ stochastic vector of factors, the Λ_i's are $1 \times r$ vectors of loadings. Set $y_t^n := (y_{1t}, y_{2t}, ..., y_{nt})'$, we may also write model (1.1) in a corresponding vector representation. The common component accounts for the co-movement in the sense that all r non-zero eigenvalues of $\mathbb{E}\, C_t^n (C_t^n)' =: \Gamma_C^n$ diverge with $n \to \infty$. The idiosyncratic component is allowed to be only weakly correlated, formalised e.g. by assuming that the first eigenvalue of $\mathbb{E}\, e_t^n (e_t^n)' =: \Gamma_e^n$ is bounded in n. This so called *approximate* factor model which has become most common in macro-econometrics, has been introduced by Chamberlain and Rothschild (1983a,b); Stock and Watson (2002a); Bai and Ng (2002).

On the other hand there has been "another kind of factor model" introduced by Forni et al. (2000); Forni and Lippi (2001) commencing from the spectrum rather than the variance matrix. This model has the form

$$y_{it} = \chi_{it} + \xi_{it} = \underline{b}_i(L)u_t + \xi_{it} \ , \tag{1.2}$$

where the common component (χ_{it}) is driven by the orthonormal q-dimensional white noise process (u_t), where $q < \infty$ is usually small, while the $\underline{b}_i(L)$'s are $1 \times q$ square summable filters. In this model the common component has the feature that all of the q non-zero eigenvalues in the spectral densities of (χ_t^n), say f_χ^n, diverge almost everywhere on the frequency band for $n \to \infty$, whereas the first eigenvalue of the spectral densities of the idiosyncratic component (ξ_t^n), say f_ξ^n, is essentially bounded on the frequency band for $n \to \infty$.

It is commonly thought that the main difference between (1.1) and (1.2) is that (1.2) allows for "infinite dimensional factor spaces" (see Forni et al., 2015) in the sense that the space spanned by $(\chi_{1t}, \chi_{2t}, \chi_{3t}, ...)'$, say $\overline{\text{sp}}(\chi_{it} : i \in \mathbb{N})$, is infinite dimensional. Consequently, if we were willing to assume that $\overline{\text{sp}}(\chi_{it} : i \in \mathbb{N})$ is finite dimensional, we could always cast the dynamic factor model (1.2) in static form (see Stock and Watson, 2011, section 2.1) and use static principal components for estimation. The moot point of this dissertation is that this is not true: The essential structural difference is that (1.2) allows for the presence of *weak factors*, while (1.1) does not. Suppose that (y_{it}) has both, a static (associated with variance matrices) and a dynamic (associated with spectral densities) factor structure. In this case the term $e_{it}^\chi := \chi_{it} - C_{it} = e_{it} - \xi_{it}$ is in general non-zero and spanned by (a potentially infinite number of) weak factors. We call (e_{it}^χ) the *weak common component.* It vanishes under static aggregation but is part of the dynamic common component (χ_{it}). By Onatski (2012) it is shown that in a static factor model with weak loadings, the common component cannot be estimated consistently via static principal components. The same applies for e_{it}^χ.

In general we induce an asymptotically non-vanishing bias if we suppose $e_{it}^\chi = 0$

for all $i \in \mathbb{N}$. As a consequence impulse response analysis is asymptotically biased, whenever variables are influenced by weak factors. On the other hand, weak factors can also be essential for forecasting: Though they are "weak" in the sense that their contemporaneous influence is not pervasive, this does not imply that they also load weakly on the subsequent period.

The aim of this thesis is to reconcile what we call the *theory of static factor sequences* or the *American School*, grounded in the work of Chamberlain and Rothschild (1983a,b); Stock and Watson (2002a,b); Bai and Ng (2002) with the theory of *dynamic factor sequences* or the *Italian School* associated with Forni and Reichlin (1996); Forni et al. (2000); Forni and Lippi (2001). We show that both schools are analogous commencing from static versus dynamic aggregation. The two different aggregational schemes entail two different types of common components (static versus dynamic) which differ by a part influenced only by weak static factors - the weak common component. We discuss and evaluate several implications for theory and practice of factor analysis. We also show that the one-sidedness problem of dynamic factor sequences Stock and Watson (2011); Forni et al. (2000, 2005, 2015) is rather a matter of estimation technique than a structural problem. It is an essential feature of the dynamic factor structure, that the innovations of the dynamic common component, i.e. the common structural shocks of (y_{it}), are causally subordinated to the output if the dynamic common component is purely non-deterministic. This justifies the interpretation of the dynamic common component as the projection of the output on the infinite past of the common innovations of the economy.

In chapter 2 we recapture the theory of factor sequences for both schools and demonstrate that the corresponding proofs are analogous. Chapter 3 provides structure theory that reconciles both schools in one model. Section 3.3 is concerned with the solution of the one-sidedness problem of dynamic factor sequences. Chapter 4 is concerned with the structural interpretation of the weak factors, the dynamic and the static common component, and the theoretical consequences of weak factors for forecasting.

Chapter 2

The Theory of Factor Sequences

The aim of this chapter is to give a clear and self-contained account of the theory of factor sequences covering the static and the dynamic case. By the theory of static/ dynamic factor sequences we mainly refer to theorems 2.2.7, 2.3.5 stated below and the lemmas that come with the corresponding proofs. There has been a kind of "cultural clash" in the literature on factor analysis between what we call the American and the Italian school of factor models and their relation has not been understood. The main contribution of this chapter is the insight that both theories can be developed analogously commencing from the notion of static versus dynamic aggregation and that the corresponding structural results (theorems 2.2.7 can 2.3.5) can be stated analogously and proved with analogous techniques. From the distinction between the two aggregational schemes, naturally there arise key definitions that we consider of vital importance for factor analysis: static versus dynamic idiosyncratic component and static versus dynamic common component. In chapter 3 we show that distinguishing between these terms implies a new decomposition that shows that the presence of weak factors is the general - and therefore the "natural" case. This decomposition also clarifies how the American and the Italian school relate to each other.

The second contribution of this chapter is that we provide in section 2.2.5 a new and alternative asymptotic framework to prove consistent estimation of the static common component via sample principal components. The proof is based on transparent and simple assumptions for the idiosyncratic component and aligned to the theory of static factor sequences as presented here. The proof given by Bai and Ng (2002); Stock and Watson (2002a) imposes different and more involved assumptions on the idiosyncratic component. We also discuss heteroskedasticity and non-stationarity. In particular, and opposed to the framework provided by Bai and Ng (2002); Stock and Watson (2002a), it becomes apparent that it is not necessary at all to restrict time dependence in the idiosyncratic component to obtain consistent estimates. The reason for this is that the definition of being statically idiosyncratic, we provide here, is only concerned about contemporaneous covariation - not with covariation over time lags.

2.1 Basic Framework

For a complex matrix $M \in \mathbb{C}^{n\times m}$ we denote by M' the transpose of M and by $M^* = \bar{M}'$ the adjoint of M. Let $\mathcal{P} = (\Omega, \mathcal{A}, P)$ be a probability space and $L_2(\mathcal{P}, \mathbb{C})$ be the Hilbert space of square integrable complex-valued, zero-mean, random-variables defined on Ω equipped with the inner product $\langle u, v\rangle = \mathbb{E}\, u\bar{v}$ for $u, v \in L_2(\mathcal{P}, \mathbb{C})$. We consider complex stochastic double sequences, i.e. a collection of random variables indexed in time and cross-section:

$$(y_{it} : i \in \mathbb{N}, t \in \mathbb{Z}) =: (y_{it}) \qquad \text{where } y_{it} \in L_2(\mathcal{P}, \mathbb{C}) \quad \forall (i,t) \in \mathbb{N}\times\mathbb{Z}\ .$$

Such a process can also be thought of as a nested sequence of multivariate stochastic processes: $(y_t^n : t \in \mathbb{Z}) =: (y_t^n)$, where $y_t^n = (y_{1t}, ..., y_{nt})'$ and $y_t^{n+1} = (y_t^{n\prime}, y_{n+1,t})'$ for $n \in \mathbb{N} \cup \{\infty\}$. In general we will write $(y_t : t \in \mathbb{Z}) =: (y_t)$ for $n = \infty$.
A basic assumption that we will often be employed is stationarity:

Assumption 1 (Stationary Double Sequence)
For all $n \in \mathbb{N}$, the process $(y_t^n :\in \mathbb{Z})$ is weakly stationary with existing (nested) spectrum $f_y^n(\theta)$ for $\theta \in \Theta := [-\pi, \pi]$.

Throughout the thesis, we denote by $f_y^n(\theta)$ the "usual spectrum" times 2π, i.e. $\Gamma_y^n := \mathbb{E}\, y_t^n (y_t^n)^* = (2\pi)^{-1}\int_{-\pi}^{\pi} f_y^n(\theta) d\theta$. This is more convenient from a notational perspective, since the frequently occurring orthonormal white noise process has then spectrum I_q instead of $(2\pi)^{-1}I_q$.

- For a stochastic vector u with coordinates in $L_2(\mathcal{P}, \mathbb{C})$, we will often write $\mathbb{V}\, u := \mathbb{E}\, uu^*$ to denote the variance matrix, which is hermitian and always diagonalisable. If u is real valued this becomes $\mathbb{V}\, u = \mathbb{E}\, uu'$ which is symmetric and diagonalisable.
- Let u be a stochastic vector with coordinates in $L_2(\mathcal{P}, \mathbb{C})$, let $\mathbb{M} \subset L_2(\mathcal{P}, \mathbb{C})$ be a closed subspace. We denote by $\text{proj}(u \mid \mathbb{M})$ the orthogonal projection of u onto $\mathbb{M}$ (see e.g. Deistler and Scherrer, 2022, Theorem 1.2).
- For any finite dimensional multivariate process (z_t) with existing spectrum, we write $\text{spec}\, z_t$ to denote the spectrum of (z_t).
- $\lambda_i(M)$ is the i-th largest eigenvalue of a square matrix M. If M is a spectral density $\lambda_i(M)$ is a measurable function in the frequency $\theta \in [-\pi, \pi]$.

We use the following notation:

$$\begin{aligned}
\mathbb{H}(y) &:= \overline{\text{sp}}\,(y_{it} : i \in \mathbb{N}, t \in \mathbb{Z}) && \text{“time domain”} \\
\mathbb{H}_t(y) &:= \overline{\text{sp}}\,(y_{is} : i \in \mathbb{N}, s \leq t) && \text{“infinite past”} \\
\overline{\text{sp}}(y_t) &:= \overline{\text{sp}}\,(y_{it} : i \in \mathbb{N}) && \text{“contemporaneous span”} \\
\text{sp}(y_t^n) &:= \text{sp}(y_{it} : i = 1, ..., n) && \text{“abbreviation for vector span”}\ ,
\end{aligned}$$

where $\overline{\text{sp}}(\cdot)$ denotes the closure of the linear span. The notation in the fourth line is used to abbreviate the (contemporaneous) span of a stochastic vector.
Given an infinite dimensional row vector $a = (a_1, a_2, \cdots) \in \mathbb{C}^{1\times\infty}$, we denote by $a^{[n]}$ the infinite row vector with zero entries after n, i.e. $(a_1, a_2, \cdots, a_n, 0, \cdots)$, and write $a^{\{n\}} := (a_1, \cdots, a_n)$. As has been shown in (Forni and Lippi, 2001, Lemma 1, 2), also for infinite dimensional stochastic processes (y_t) there exists a corresponding isometric frequency domain to the time domain $\mathbb{H}(y)$:

$$\mathbb{H}(y) \overset{\Phi}{\leftrightarrow} L_2^{\infty}(f_y) \tag{2.1}$$

where $\Phi(\cdot)$ is an isomorphism that preserves the inner product, $L_2^\infty(f_y)$ is the complex linear space of all infinite row vectors of complex valued equivalence classes $c = (c_1(\cdot), c_2(\cdot), \cdots)$, such that for all $i \in \mathbb{N}$, we have

1. $c_i : \Theta \to \mathbb{C}$ is a measurable function,
2. $\lim_n \int_{-\pi}^{\pi} c^{\{n\}}(\theta) f_y^n(\theta) c^{*,\{n\}}(\theta) d\theta < \infty$,
3. the space $L_2^\infty(f_y)$ is endowed with the inner product
$$\langle c, d \rangle_{f_y} = \lim_n (2\pi)^{-1} \int_{-\pi}^{\pi} c^{\{n\}}(\theta) f_y^n(\theta) d^{*,\{n\}}(\theta) d\theta$$
4. and the norm $\|c\|_{L_2^\infty(f_y)} := \sqrt{\langle c, c \rangle_{f_y}}$
5. two vectors c_1, c_2, are equivalent if $\|c_1 - c_2\|_{L_2^\infty(f_y)} = 0$.

For processes in $\mathbb{H}(y)$ that are outputs of filters, we write

$$z_t = \underline{c}(L) y_t := \Phi^{-1}\left(c(\theta) e^{\iota\theta t}\right) \ , \tag{2.2}$$

where $\Phi(\cdot)$ is the isomorphism from (2.1). Accordingly, we write $L_2^\infty(I)$, if f_y^n is the identity matrix I_n for all n.

Next, we consider *sequences* of infinite row vectors (of equivalence classes of functions):

$$(c^{(k)} : k \in \mathbb{N}) = \left((c_1^{(k)}, c_2^{(k)}, \cdots) \mid k \in \mathbb{N}\right) \ .$$

In general, for the limit outputs filtered by such a sequence $(c^{(k)})$, we write:

$$z_t = \underline{\lim}_k \, \underline{c}^{(k)}(L) y_t \ , \tag{2.3}$$

where "$\underline{\lim}$" denotes the limit with respect to mean square convergence. Analogously, we treat the limit of cross-sectional weighted sums:

$$z_t = \underline{\lim}_k \, \hat{c}^{(k)} y_t = \underline{\lim}_k \, \underline{\lim}_n \sum_{i=1}^{n} \hat{c}_i^{(k)} y_{it} \ ,$$

where $\hat{c}_i^{(k)}$ is the i-th entry of the infinite row vector of constant weights $\hat{c}^{(k)}$.
Now drop A1 for the moment, if $(\Gamma_y^n(t) : t \in \mathbb{Z}) = (\mathbb{V}\, y_t^n : t \in \mathbb{Z})$ is the sequence of variance matrices (which may depend on time t) associated with (y_{it}) - not necessarily with existing spectrum - we denote by $\hat{L}_2^\infty(\Gamma_y(t))$ the set of all *constant* vectors $\hat{c} \in \mathbb{C}^{1\times\infty}$ such that $\lim_n \hat{c}^{\{n\}}\Gamma_y^n(t)\left(\hat{c}^{\{n\}}\right)^* < \infty$ and $\hat{L}_2^\infty(I) \subset L_2^\infty(I)$ as the the set of all vectors with $\lim_n \hat{c}^{\{n\}}(\hat{c}^{\{n\}})^* < \infty$. So by infinite row vectors equipped with "$\hat{\cdot}$", we denote vectors of weightings for computing cross-sectional weighted averages.

2.2 Static Factor Sequences: The American School

The theory of static factor sequences, as we propose to call it, originates in the "approximate factor model" by Chamberlain and Rothschild (1983a,b), which is formulated not specifically in a time series context. The authors develop theory for the case that population second moments are known. They do not provide an estimation theory. Stock and Watson (2002a); Bai and Ng (2002) were the first to set the idea of "the approximate factor model" in a time series framework, and to provide estimation theory for it.
In this section, we present the main result of the theory of static factor sequences (see Theorem 2.2.7 below), implicitly contained in Chamberlain and Rothschild (1983a), and prove it by applying and adapting the techniques from Forni and Lippi (2001). So part of the reconciliation of the theory of factor sequences consists in demonstrating that both theories can be proved and understood with analogous mathematical frameworks and proof-techniques: In the static case we work with variance matrices (see section 2.2.3), in the dynamic case we work with spectral densities (see section 2.3.1).

Remark 2.2.1 (Setting the Static Factor Model in a Time Series Context)
The theory developed by Chamberlain and Rothschild (1983a) commences from a nested sequence of variance matrices Γ_y^n corresponding to a stochastic vector $y^n = (y_1, ..., y_n)'$ of growing dimension $n \to \infty$ or alternatively from a "cross-sectional sequence of random variables" $(y_i : i \in \mathbb{N})$. Here, we set the theory in a

time series framework by adding a time index t and supposing that we have such a cross-sectional sequence for every time point $t \in \mathbb{Z}$.
Strictly speaking, the time index is technically irrelevant for the main results of the theory of "static factor sequences"; theorems 2.2.6, 2.2.7. Nonetheless we add the time index to contrast "static factor sequences" which emerge from cross-sectional or contemporaneous aggregations with "dynamic factor sequences" which emerge from aggregations over both - cross-section and time (see section 2.2.1). Thereby, we can reconcile the two schools of factor models in one theoretical framework.

Remark 2.2.2 (Time-Dependent $\Gamma_y^n(t)$)
The corresponding variances $\Gamma_y^n(t) := \mathbb{V}\, y_t^n$ may depend on time or may be constant if (y_{it}) is stationary. For the sake of notational convenience, we present and prove the structural results below for the "stationary case". All results hold also when Γ_y^n or $\Gamma_C^n := \mathbb{V}\, C_t^n$ and $\Gamma_e^n := \mathbb{V}\, e_t^n$ depend on time. In section 2.2.6, we discuss non-stationarity and heteroskedasticity in more detail.

2.2.1 Intuition: Factor Models in Terms of Aggregation

In the literature it is most common to approach factor sequences from a "model-driven point of view": This means that we start by assuming that the data is generated by a model like (1.1), assuming "strong factor loadings" in the sense that $\lambda_r(\Gamma_C^n) \to \infty$ and assuming weakly dependent idiosyncratic terms (e_{it}) (see e.g. Stock and Watson, 2002a; Bai and Ng, 2002). If we formalise the notion of weak dependence by $\lambda_1(\Gamma_e^n) < \infty$, we readily obtain the following about the behavior of eigenvalues of Γ_y^n for $n \to \infty$:

$$\lambda_r(\Gamma_C^n) \leq \lambda_r(\Gamma_y^n) \to \infty \tag{2.4}$$

$$\text{and} \qquad \lambda_{r+1}(\Gamma_y^n) \leq \lambda_{r+1}(\Gamma_C^n) + \lambda_1(\Gamma_e^n) < \infty\ . \tag{2.5}$$

The remarkable result by Chamberlain and Rothschild (1983a) which has been extended by Forni and Lippi (2001) to the "frequency domain" and which we are going to prove in this section, is that the converse is also true: If the eigenvalues of Γ_y^n behave as in equations (2.4), (2.5), i.e. if the following assumption holds

Assumption 2 (r-Static Factor Structure)
There exists $r < \infty$, such that

(i) $\sup_n \lambda_r(\Gamma_y^n) = \infty$

(ii) $\sup_n \lambda_{r+1}(\Gamma_y^n) < \infty$,

it can be shown (theorem 2.2.7) that y_{it} has a factor model representation (1.1). For this reason, we will call a stochastic double sequence (y_{it}) for which A2 holds a static "factor sequence" (see definition 2.2.8). In contrast, as discussed below, there are also double sequences (y_{it}) which do not have a factor structure, e.g. if no or all eigenvalues of Γ_y^n diverge. Summing up, we distinguish the "structural approach" from the "model driven approach" in the follwoing sense: The existence of (1.1) does not hold by a "model-assumption" but is implied by the specific eigenvalue structure, i.e. A2, of the observed output process.
In my point of view, one of the key contributions of Forni and Lippi (2001) is to provide the clear link of factor models/sequences to aggregation. To give an intuition for this, consider taking the cross-sectional average over the first k variables. The variance of that average is

$$\mathbb{V}\,\bar{y}_t^k = \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} y_{it}\right)^2 = \frac{1}{k^2}\left\{\sum_{i=1}^{k}\mathbb{V}\,y_{it} + \sum_{i\neq j, i,j\leq k}\mathbb{E}\,y_{it}y_{jt}\right\} .$$

The term on the RHS is the sum of the variances plus the sum of all cross-covariances multiplied by $1/k^2$. Now let $k \to \infty$: If the y_{it}'s were cross-sectionally independent and identically distributed, we know that by the law of large numbers $\mathbb{V}\,\bar{y}_t^k \to 0$. On the other hand if the cross-sectional covariances among the variables are "strong enough" such that they are not killed by $1/k^2$, the variance of $\bar{y}_t^k$ does not vanish with increasing k. The operation of computing the limit of the cross-sectional average is a special case of *static aggregation* which is performed with the *static averaging sequence* (see definition 2.2.3 below)

$$(\hat{c}^{(k)}) = (\hat{c}_1^{(k)}, \hat{c}_2^{(k)}, \ldots) = (\underbrace{1/k, 1/k, \cdots, 1/k}_{k \text{ times}}, 0, 0, \cdots) \in \mathbb{C}^{1\times\infty} ,$$

with the special property that

$$\lim_k \hat{c}^{(k)}(\hat{c}^{(k)})^* = \lim_k \sum_{i=1}^{k} \frac{1}{k^2} = \frac{k}{k^2} = \lim_k \frac{1}{k} = 0 \ .$$

As it turns out the factor space, i.e. the Hilbert space spanned by the static factors $\text{sp}(F_t)$, is the space that is generated by the set of all static aggregates, namely all possible random variables z_t that can be represented as the limit of a static aggregation $z_t = \underline{\lim}_k \hat{c}^{(k)} y_t^k$ where $\hat{c}^{(k)}(\hat{c}^{(k)})^* \to 0$ for $k \to \infty$. So if (y_{it}) has a non-trivial static factor space, it means that the cross-sectional covariances are strong enough such that they are not killed under cross-sectional aggregation for suitably chosen averaging sequences $(\hat{c}^{(k)})$. In this sense "factor modelling is about capturing covariances not variances" (see Barigozzi, 2018).

2.2.2 The Structure Theory of Static Factor Sequences

In the following, we formalise the theory static factor sequences adapting the concepts of Forni and Lippi (2001) to the static case. As pointed out in remark 2.2.1, all of the results in this section hold also if we allow $\Gamma_y^n(t)$ to change over time. To keep notation simple, we omit the time index on the variance matrices involved and refer to the discussion in section 2.2.6.

Definition 2.2.3 (Static Averaging Sequence (SAS))
Let $\hat{c}^{(k)} \in \hat{L}_2^{\infty}(I) \cap \hat{L}_2^{\infty}(\Gamma_y)$ for all $k \in \mathbb{N}$. The sequence of cross-sectional aggregations $\left(\hat{c}^{(k)} : k \in \mathbb{N}\right)$ is called Static Averaging Sequence (SAS) if

$$\lim_k \left\|\hat{c}^{(k)}\right\|_{\hat{L}_2^{\infty}(I)} := \lim_k \hat{c}^{(k)} \left(\hat{c}^{(k)}\right)^* = 0 \ .$$

We denote the set of all static averaging sequences corresponding to (y_{it}) as

$$\mathcal{S}(\Gamma_y) := \left\{ \left(\hat{c}^{(k)}\right) : \hat{c}^{(k)} \in \hat{L}_2^{\infty}(I) \cap \hat{L}_2^{\infty}(\Gamma_y) \cap \mathbb{C}^{1\times\infty} \ \forall k \in \mathbb{N} \text{ and } \lim_k \left\|\hat{c}^{(k)}\right\|_{\hat{L}_2^{\infty}(I)} = 0 \right\} \ .$$

The set of all random variables that can be written as the limit of a SAS defines a closed subspace of $\overline{\text{sp}}(y_t)$ (see Forni and Lippi, 2001, Lemma 6).

Definition 2.2.4 (Static Aggregation Space)
The space $\mathbb{S}_t(y) := \left\{ z_t : z_t = \underline{\lim}_k \hat{c}^{(k)} y_t, \text{ where } \left(\hat{c}^{(k)}\right) \in \mathcal{S}(\Gamma_y) \right\} \subset \overline{\mathrm{sp}}(y_t)$ *is called Static Aggregation Space at time* t.

Henceforth, we will write $\mathbb{S}_t$ instead of $\mathbb{S}_t(y)$ when it is clear from the context. Note that the Static Aggregation Space changes over $t \in \mathbb{Z}$ as it emerges from aggregation of the cross-section of y_{it} holding t fixed.

Definition 2.2.5 (Statically Idiosyncratic)
A stochatic double sequence (z_{it}) *is called statically idiosyncratic, if* $\underline{\lim}_k \hat{c}^{(k)} z_t = 0$ *for all* $(\hat{c}^{(k)}) \in \mathcal{S}(\Gamma_z)$ *for all* $t \in \mathbb{Z}$.

The concept of a *statically* idiosyncratic double sequence is implicitly contained in Chamberlain and Rothschild (1983b). In particular Forni and Lippi (2001) introduce a notion of idiosyncraticness that we identify with being *dynamically* idiosyncratic. On the other hand Bai and Ng (2002) define idiosyncratic in a different way that involves also a limitation of time dependence (see also sections 2.2.6, 2.2.5). One of the main aspects of this thesis is, that it is fundamentally important to distinguish between the two ideas of static and dynamic aggregation (see below) since two different types of "common-ness" and "idiosyncraticness" are associated with that. As we will illustrate in section 3, a double sequence that vanishes under static aggregation does not need to do so under dynamic aggregation.
Fortunately, there is an easy and very useful characterisation of a statically idiosyncratic double sequence:

Theorem 2.2.6 (Characterisation of Statically Idiosyncratic)
The following statements are equivalent:

(i) A stochastic double sequence (z_{it}) *is statically idiosyncratic.*

(ii) The first eigenvalue of the variance matrix is bounded, i.e.

$$\sup_n \lambda_1(\Gamma_z^n) < \infty, \text{ for all } t \in \mathbb{Z} .$$

The proof of theorem 2.2.6 works analogously to the proof of theorem 2.3.4 (see Forni and Lippi, 2001, theorem 1) and is given in the Appendix.
As is well known, we can compute *static low rank approximations* (SLRA) of rank r of y_t^n by "static" principal components. For this consider the eigen-decomposition of the variance:

$$\Gamma_y^n = P_{(n)}^* \Lambda_{(n)} P_{(n)} \ , \tag{2.6}$$

where $P_{(n)}$ is a unitary matrix of row eigenvectors and $\Lambda_{(n)}$ is a diagonal matrix of eigenvalues sorted from the largest to the smallest. Denote by p_{nj} the j-th row of $P_{(n)}$ and by P_{nr} the sub-unitary matrix consisting of the first r rows of $P_{(n)}$. Recall that we associate r with the number of divergent eigenvalues of Γ_y^n (see A2). As we will frequently use P_{nr}, we make notation easier by writing $P_n := P_{nr}$ which is $r \times n$ in contrast to $P_{(n)}$ which is $n \times n$, analogously we write Λ_n to denote the $r \times r$ diagonal matrix of the largest r eigenvalues of Γ_y^n. Set

$$\mathcal{K}_{ni} := \mathcal{K}_{ni}(\Gamma_y^n) := p_{ni}^* P_n \quad \text{the } i\text{-th row of } P_n^* P_n \tag{2.7}$$

$$C_t^{[n]} := P_{nr}^* P_{nr} y_t^n = P_n^* P_n y_t^n \tag{2.8}$$

$$C_{it,n} := \mathcal{K}_{ni} y_t^n \quad \text{the } i\text{-th row of } C_t^{[n]} \ . \tag{2.9}$$

Recall that $C_t^{[n]}$ is the best (with respect to squared error) possible approximation of y_t^n by an r dimensional vector of linear combinations of $y_{1t}, ..., y_{nt}$. The $r \times 1$ vector $P_n y_t^n$ are the first r principal components of y_t^n and provide such a vector of linear combinations, though not uniquely. We call $C_t^{[n]}$ the *static rank r approximation* of y_t^n which is unique. For details on principal component analysis/ low rank approximations see Jolliffe (2002).
Here comes the main result of the theory of static factor sequences. This theorem is implicitly contained in Chamberlain and Rothschild (1983a). Here we present a different formulation aligned to Forni and Lippi (2001):

Theorem 2.2.7 (Chamberlain and Rothschild (1983a): r-Static Factor Sequence)
Consider a stochastic double sequence (y_{it}) *in* $L_2(\mathcal{P}, \mathbb{C})$

1. *A2 holds if and only if we can decompose*

$$y_{it} = C_{it} + e_{it} = \Lambda_i F_t + e_{it} \qquad \textit{where } \mathbb{E}\, C_{jt} e_{it} = 0 \quad \forall i, j, t, \tag{2.10}$$

where F_t is such that $\mathbb{E}\, F_t F_t^* = I_r$ and (C_{it}) and (e_{it}) are stochastic double sequences such that for $\Gamma_C^n := \mathbb{V}\, C_t^n$, $\Gamma_e^n := \mathbb{V}\, e_t^n$ it holds that

   (i) $\sup_n \lambda_r(\Gamma_C^n) = \infty$

   (ii) $\sup_n \lambda_1(\Gamma_e^n) < \infty$

2. $C_{it} = \underline{\lim}_n C_{it,n}$,
3. r, C_{it}, e_{it} *are uniquely determined from the output sequence,*
4. $C_{it} = \operatorname{proj}(y_{it} \mid \mathbb{S}_t(y))$.

The proof is given in section 2.2.3. Some comments in order. The "if" direction of the first statement has already been proved in section 2.2.1. The "only if" part of the first statement is a *representation result*: It allows us to conclude from the characteristic behaviour of the eigenvalues of the variance matrices Γ_y^n, i.e. A2, to the existence of unique representation as a sum of (C_{it}) which we call the *static common component* (static CC) and (e_{it}) which we call *static idiosyncratic component*. The uniqueness of that decomposition holds by the third statement. By the fourth statement, the static common component is the projection of the output variables on the static aggregation space. So it represents the part that does not vanish under static, i.e. cross-sectional aggregation. As a consequence, in general, *it is not orthogonal to the static idiosyncratic component at all leads and lags* but *only contemporaneously orthogonal*. The second statement provides the link to principal components analysis: The i-th static common component is the mean square limit of the i-th element of the static rank r approximation of the output process letting $n \to \infty$. This is very useful for proving that the sample low rank approximations converge in probability to the true common component (see section 2.2.5). It also provides another way of interpreting an approximate factor model: The number of static factors, r, is a "good choice" for the number of principal components in a low rank approximation. Since asymptotically $(n \to \infty)$

the relative gain in terms of explained variance from choosing $r+1$ instead of r principal components is negligible compared to the relative gain from choosing r instead of $r-1$ principal components. The first method for determining r has been given by Bai and Ng (2002). In Ahn and Horenstein (2013) an eigenvalue ratio test for r is provided.
By theorem 2.2.7.1 the eigenvalue structure of Γ_y^n in A2 is equivalent to a representation as a factor model. This justifies the name "static factor sequence":

Definition 2.2.8 (r-Static Factor Sequence (r-SFS))
A stochastic double sequence (y_{it}) in $L_2(\mathcal{P}, \mathbb{C})$ that satisfies A2 is called r-Static Factor Sequence, r-SFS.

If we assume in addition that (y_{it}) is weakly stationary, e.g. A1, it readily follows that, factors, common component and idiosyncratic component are stationary as well. Note that we do not need the existence of the spectrum (A1) for weak stationarity. The following corollary holds whenever (y_{it}), which is always supposed to be zero mean, satisfies $\mathbb{E}\, y_{it}^2 < \infty, \mathbb{E}\, y_{it}y_{is}^* = \mathbb{E}\, y_{i,t-s}y_0^*$.

Corollary 2.2.9 (Stationarity of the Static Factors)
If (y_{it}) is a stationary r-SFS then $(F_t), (C_{it})$ and (e_{it}) are stationary processes.

Proof. If (y_{it}) is stationary, the auto-covariance function of $(y_t^n : t \in \mathbb{Z})$ does not depend on time, so $\mathbb{E}\, y_s^n (y_t^n)^* = \mathbb{E}\, y_{t-s}^n (y_0^n)^* =: \Gamma_y^n(t-s)$, say and therefore $\langle C_{it,n}, C_{j,t-k,n}\rangle = \mathcal{K}_{ni}\Gamma_y^n(k)\mathcal{K}_{nj}^*$ for all $t \in \mathbb{Z}$ and $k \in \mathbb{N}$. By continuity of the inner product (see also lemma A.0.3), for every $t \in \mathbb{Z}$, we have

$$\langle \underline{\lim}_n C_{it,n}, \underline{\lim}_n C_{j,t-k,n}\rangle = \lim_n \langle C_{it,n}, C_{j,t-k,n}\rangle = \gamma_{C,ij}(k) \quad \text{say}$$

which is independent of t. Consequently, also $e_{it} = y_{it} - C_{it}$ is stationary. Since the factors $F_t = (F_{1t}, ..., F_{rt})'$ can be also defined as the mean square limit of a cross-sectional weighted average (see lemma 2.2.13, propositions 2.2.15 and 2.2.20) for the same arguments (F_t) is stationary. ■

2.2.3 Proof of the Main Result: Static Case

In this section we prove Theorem 2.2.7. The proof for the static case is very analogous to the proof for the dynamic case. In the static case we deal with eigenvectors and eigenvalues of the nested sequence of variance matrices $\mathbb{V}\, y_t^n = \Gamma_y^n$, whereas in the dynamic case we deal with the eigenvalues and eigenvectors of the nested sequence of spectral densities, i.e. $\operatorname{spec} y_t^n = f_y^n$. We will use the following notation throughout the proofs in this section

- **Canonical projection static case**: $y_{it} = \operatorname{proj}(y_{it} \mid \mathbb{S}_t) + \delta_{it} = \gamma_{it} + \delta_{it}$ where γ_{it} is the orthogonal Hilbert space projection onto the static aggregation space
- $\Gamma_y^n = P_n^* \Lambda_n P_n + Q_n^* \Phi_n Q_n$, where Λ_n, Φ_n are diagonal matrices containing the first r largest and the remaining $n-r$ eigenvalues in order and P_n, Q_n are the associated unitary matrices of row-eigenvectors.
- Normalised static principal components: $\psi_t^n := \Lambda_n^{-1/2} P_n y_t^n$
- Rotation matrix $C \in \mathbb{C}^{r\times r}$ matrix: $C^*C = I_r = CC^*$
- Rotated version of the normalised principal components: $\psi_t^{m,C} := C\psi_t^m$
- We write $P_m y_t^n$ short for $(P_m, 0 \cdots 0) y_t^n = P_m y_t^m$, i.e. we add $n-m$ zero columns to fit the dimension
- $\lambda_{n,j} = \lambda_j(\Gamma_y^n)$
- $\hat{\psi}_t^{m,n,C} := \operatorname{proj}(C\psi_t^m \mid \overline{\operatorname{sp}}(\psi_t^n))$.
- $D(m,n,C) =: D = C\Lambda^{-1/2} P_m P_n^* \Lambda_n^{-1/2}$ is the projection matrix that transforms ψ_t^n to $\hat{\psi}_t^{m,n,C}$, so $\hat{\psi}_t^{m,n,C} = D\psi_t^n$ and $\mathbb{V}\, \hat{\psi}_t^{m,n,C} = D\psi_t^n = DD^*$
- (α_n) is a scalar sequence, such that $\lim_n \alpha_n = \infty$ and $\lambda_{n,r} := \lambda_r(\Gamma_y^n) \geq \alpha_n$

In this section, we prove the "only if" direction of theorem 2.2.7.1, i.e. we show that a factor model representation exists if A2 holds. We also prove the other

statements of theorem 2.2.7. We follow the steps in Forni and Lippi (2001) adapting them to the static case.
The proof consists of the following main steps: **a)** We construct a r-dimensional stochastic vector V_t in $\mathbb{S}_t(y)$ having unit variance using suitably rotated normalised principal components (proposition 2.2.15). **b)** We show that this vector is a basis for $\mathbb{S}_t(y)$ (proposition 2.2.15). **c)** We show that the normalised principal components (ψ_t^n) generate a Cauchy sequence of static spaces $\mathrm{sp}(\psi_t^n)$ (see definition 2.2.16). Projections $\mathrm{proj}(y_{it} \mid \mathrm{sp}(\psi_t^n))$ must converge in mean square and they converge to $\mathrm{proj}(y_{it} \mid \mathbb{S}_t)$. **d)** We show that the residual of this projection is idiosyncratic (lemma 2.2.21). **e)** All this implies the uniqueness of the number of factors, the common and the idiosyncratic component (proposition 2.2.22) and that $C_{it} = \mathrm{proj}(y_{it} \mid \mathbb{S}_t)$.
We start with constructing an r-dimensional basis V_t for $\mathbb{S}_t$ with the following lemmas:

Lemma 2.2.10
Suppose $m \leq n$ then

$$\begin{aligned} C\psi_t^m &= \mathrm{proj}(\psi_t^{m,C} \mid \mathrm{sp}(\psi_t^n)) + R_t(m,n) \\ &= \underbrace{C\Lambda^{-1/2}P_m P_n^* \Lambda_n^{1/2}}_{D(m,n,C)} \underbrace{\Lambda_n^{-1/2}P_n y_t^n}_{\psi_t^n} + \underbrace{C\Lambda_m^{-1/2}P_m Q_n^* Q_n y_t^n}_{R_t(m,n,C)} . \end{aligned}$$

Proof.

$$\begin{aligned} C\psi_t^m &= C\Lambda_m^{-1/2}P_m y_t^n = C\Lambda^{-1/2}P_m \underbrace{(P_n^* P_n + Q_n^* Q_n)}_{I_n} y_t^n \\ &= C\Lambda_m^{-1/2}P_m P_n^* \Lambda_n^{1/2}\Lambda_n^{-1/2}P_n y_t^n + \underbrace{C\Lambda_m^{-1/2}P_m Q_n^* Q_n}_{R(m,n,C)} y_t^n \end{aligned}$$

which is by the projection theorem already the decomposition in projection plus residual, as the first term is in $\mathrm{sp}(\psi_t^n)$ and the second term is orthogonal to $\mathrm{sp}(\psi_t^n)$. ■

Note that $D(m,n,C) =: D$ is the population projection matrix, when pro-

jecting $\psi_t^{m,C}$ on $\operatorname{sp}\psi_t^n$. On the other hand $R(m,n,C) =: R$ is the matrix that transforms y_t^n to the residual of that projection.
The main idea is that even though the normalised principal components do not converge, the spaces which they span generate a "Cauchy sequence". The notion that the spaces $\operatorname{sp}(\psi_t^m)$, $\operatorname{sp}(\psi_t^n)$ become close for $m, n \to \infty$ is formalised by looking at the projections $\hat{\psi}_t^{m,n,C} := \operatorname{proj}(C\psi_t^m \mid \overline{\operatorname{sp}}(\psi_t^n))$ for $m, n \in \mathbb{N}$ and showing that the residual term of this projection tends to zero for $m, n \to \infty$. The idea of a Cauchy sequence of spaces will be formalised in more detail below.

Lemma 2.2.11
Suppose A2 holds, then $\lambda_1(\mathbb{V}\, R_t(m,n,C)) \leq \lambda_{n,r+1}/\lambda_{m,r}$, *where* $\lambda_{n,j} := \lambda_j(\Gamma_y^n)$.

Proof. We know that

$$
\begin{aligned}
&I_n - Q_n^* Q_n = P_n^* P_n \geq 0 \\
&\lambda_{n,r+1} I_{n-r} - \Phi_n \geq 0 \quad \Rightarrow \quad Q_n^*(\lambda_{n,r+1} - \Phi_n) Q_n \geq 0\,. \\
&\lambda_{n,r+1}(I_n - Q_n^* Q_n) + \lambda_{n,r+1} Q_n^* Q_n - Q_n^* \Phi_n Q_n \quad \text{(using the two equations above)} \\
&\ = \lambda_{n,r+1} I_n - Q_n^* \Phi_n Q_n \geq 0 \text{ by Fact A.0.1.}
\end{aligned}
$$

Next we pre- and postmultiply $C\Lambda_m^{-1/2} P_m$, which yields:

$$
\begin{aligned}
&\lambda_{n,r+1} C \Lambda_m^{-1} C^* - C\Lambda_m^{-1/2} P_m Q_n^* \underbrace{Q_n \Gamma_y^n Q_n^*}_{\Phi_n} Q_n P_m^* \Lambda_m^{-1/2} C^* \geq 0 \\
&\lambda_{n,r+1} C \Lambda_m^{-1} C^* - \mathbb{V}\, R_t(m,n,C) \geq 0 \\
&\lambda_1(\lambda_{n,r+1} C \Lambda_m^{-1} C^*) = \frac{\lambda_{n,r+1}}{\lambda_{m,r}} \geq \lambda_1(\mathbb{V}\, R_t(m,n,C)) \quad \text{by Fact A.0.1.}
\end{aligned}
$$

■

Clearly if $m, n \to \infty$, the first eigenvalue of $\mathbb{V}\, R_t(m,n,C)$ tends to zero by A2. We can regard this as a crucial feature of an r-SFS (y_{it}) (see A2), that for increasing cross-sectional dimension the spaces spanned by the first r principal components become closer.
Next, we look at the projection $\hat{\psi}_t^{m,n,C} := \operatorname{proj}(C\psi_t^m \mid \overline{\operatorname{sp}}(\psi_t^n))$. We show that the

eigenvalues of $DD^* = \mathbb{V}\,\hat{\psi}_t^{m,n,C}$ tend to one from below. For this let $\alpha_n \to \infty$ be an increasing sequence such that $\lambda_{n,r} \geq \alpha_n$ and let $\lambda_{n,r+1} < W$ for all n.

Lemma 2.2.12 (Squeeze Lemma for eigenvalues of $\mathbb{V}\,\psi_t^n$)
There exists m^*, such that $1 \geq \lambda_j(DD^*) \geq 1 - W/\alpha_m > 0$ for all $m \geq m^*$ and $j = 1, ..., r$.

Proof.

$$\begin{aligned}
I_r = \mathbb{V}\,\psi_t^{m,C} &= DD^* + R\Gamma_y^n R^* \quad \text{from Lemma 2.2.10} \\
DD^* = I_r - R\Gamma_y^n R^* &\geq I_r - I_r\lambda_1(R\Gamma_y^n R^*) \\
&\geq I_r - I_r\frac{\lambda_{n,r+1}}{\lambda_m} \quad \text{Lemma 2.2.11} \\
&\geq I_r - I_r\frac{W}{\alpha_m} \ .
\end{aligned}$$

So choosing m^* such that $W/\alpha_m < 1$ for all $m \geq m^*$, we have $1 \geq \lambda_r(DD^*) \geq \lambda_r(I_r - I_r(W/\alpha_m)) = 1 - W/\alpha_m$. ∎

Note that even though the normalised principal components $C\psi_t^m$ are orthonormal, the projection $\hat{\psi}_t^{m,n,C}$ is not. In the next step we orthonormalise $\hat{\psi}_t^{m,n,C}$: Let $H\Delta H^* = DD^*$ be the eigen-decomposition of DD^* (in the usual manner with orthonormal eigenvectors in H), we define an orthonormalised version of $\hat{\psi}_t^{m,n,C}$ as

$$\bar{\psi}_t^{m,C} = H\Delta^{-1/2}H^* \underbrace{\operatorname{proj}(C\psi_t^m \mid \operatorname{sp}(\psi_t^n))}_{\hat{\psi}_t^{m,n,C}} = \underbrace{H\Delta^{-1/2}H^*D}_{=:F(D)=:F}\,\psi_t^n = F\psi_t^n \ . \tag{2.11}$$

In Lemma 2.2.10, we showed that for large m, n the projection $\hat{\psi}_t^{m,n,C}$ gets close to $\psi_t^{m,C}$. Now we show that this holds also for the orthonormalised version of the projection, i.e. $\bar{\psi}_t^{m,C}$ of $\psi_t^{m,C}$. The idea behind the construction of a basis for $\mathbb{S}_t$ is to construct a Cauchy sequence from the orthonormalised versions of the projections by iteration:

Lemma 2.2.13 (The orthonormalised projection)
For all $\tau \in (0, 2)$, there exists m_τ, such that

- $W/\alpha_{m_\tau} < 1$
- *For all* $n \geq m \geq m_\tau$ *we have* $\lambda_1\left(\mathbb{V}\left[\psi_t^{m,C} - \bar{\psi}_t^{m,n,C}\right]\right) < \tau$.

Proof. We have

$$\psi_t^{m,C} - \bar{\psi}_t^{m,n,C} = C\psi_t^m - F\psi_t^n = D\psi_t^n + Ry_t^n - F\psi_t^n = Ry_t^n + (D-F)\psi_t^n \ ,$$

hence $\mathbb{V}(C\psi_t^m - F\psi_t^n) = I_r - DD^* + (D-F)(D-F)^* = I_r - DD^* + DD^* - DF^* + FD^* + \underbrace{FF^*}_{I_r}$

$$= 2I_r - D(H\Delta^{-1/2}H^*D)^* - (H\Delta^{-1/2}H^*D)D^* = 2I_r - DD^*H\Delta^{-1/2}H^* - H\Delta^{-1/2}H^*DD^*$$
$$= 2I_r - H\Delta H^*H\Delta^{-1/2}H^* - H\Delta^{-1/2}H^*H\Delta H^* = H(2I_r - 2\Delta^{1/2})H^*$$

hence $\lambda_1(\mathbb{V}(C\psi_t^m - F\psi_t^n)) = \lambda_1(2(I_r - \Delta^{1/2})) = 2 - 2\underbrace{\sqrt{\lambda_r(DD^*)}}_{\leq 1}$

$$\leq 2 - 2(1 - W/\alpha_{m_\tau}) = 2W/\alpha_{m_\tau} < \tau \quad \text{by Lemma 2.2.12} \ .$$

■

Next we use Lemma 2.2.13 to show by construction that there exists an orthonormal process V_t in the static aggregation space $\mathbb{S}_t$ which is also a basis for $\mathbb{S}_t$.

Proposition 2.2.14 (Existence of an orthonormal stochastic vector in the static aggregation space)
Suppose A2 holds, then there exists a zero-mean process $V_t = (V_{1t}, ..., V_{rt})'$, *such that*
1. $V_{jt} \in \mathbb{S}_t$ *for all* $t \in \mathbb{Z}, j = 1, ..., r$
2. $\mathbb{V}\, V_t = I_r$.

Proof. We construct a Cauchy sequence V_t^k in $\mathbb{S}_t$.
$\underline{(k=1)}$: We start with $\tau_1 = (\frac{1}{2})^{2k} = (\frac{1}{2})^2$. Let $C \in \mathbb{C}^{r\times r}$ be unitary. By Lemma 2.2.13, we know that there exists m_{τ_1}, such that

$$\lambda_1\left(\mathbb{V}\left[\psi_t^{m,C} - \bar{\psi}_t^{m,n,C}\right]\right) < \tau_1 = \left(\frac{1}{2}\right)^2 \quad \forall n \geq m \geq s_1 \ . \tag{2.12}$$

Set $s_1 = m_{\tau_1}$, which is a number from which all orthonormalised projections are closer to $\psi_t^{m,C}$ $(m \geq s_1)$ than τ_1. Set $V_t^1 = \psi_t^{s_1,C} = C\psi_t^{s_1} = C\Lambda_{s_1}^{-1/2} P_{s_1} y_{s_1,t}$. So for all $n \geq s_1$, we have $\lambda_1 \left(\mathbb{V} \left[\psi_t^{s_1,C} - \bar{\psi}_t^{s_1,n,C} \right] \right) < \tau_1$. In other words, for $n \geq s_1$ the orthonormalised projection $\bar{\psi}_t^{s_1,n,C}$ is "closer" than τ_1 to $C\psi_t^{s_1}$ - measuring closeness in terms of the first eigenvalue of the variance of the difference.

$\underline{(k=2)}$: $\tau_2 = (\frac{1}{2})^{2k} = (\frac{1}{2})^4$. Now consider $F_2 := F(C)$ as in equation (2.11) and set $s_2 := m_{\tau_2}$:
Now by Lemma 2.2.13, we know that for $s_2 = m_{\tau_2}$, we have

$$\lambda_1 \left(\mathbb{V} \left[\psi_t^{m,F_2} - \bar{\psi}_t^{m,n,F_2} \right] \right) < \tau_2 = \left(\frac{1}{2} \right)^4 \qquad \forall n \geq m \geq s_2.$$

We set $V_t^2 := \psi_t^{s_2,F_2} = F_2 \psi_t^{s_2} = \underbrace{H\Delta^{-1/2}H^*}_{F(C)} C\psi_t^{s_2} = F_2 \Lambda_{s_2}^{-1/2} P_{s_2} y_t^{s_2} = \bar{\psi}_t^{s_1,s_2,C}$. Here H, Δ is the corresponding eigen-decomposition as in equation (2.11) for $D = F_1$. Thus

$$\lambda_1 \left(\mathbb{V} \left[V_t^1 - V_t^2 \right] \right) = \lambda_1 \left(\mathbb{V} \left[\psi_t^{s_1,C} - \bar{\psi}_t^{s_1,s_2,C} \right] \right) < \left(\frac{1}{2} \right)^2 ,$$

by equation (2.12) for $m = s_1$, $n = s_2$ and therefore also for the coordinates $V_{jt}^{\cdot}$ of $V_t^{\cdot}$ with $\cdot = 1, 2$, $j := 1, ..., r$ we have

$$\left\| V_{jt}^1 - V_{jt}^2 \right\| = \mathbb{V} \left[V_{jt}^1 - V_{jt}^2 \right] \leq \lambda_1 \left(\mathbb{V} \left[V_t^1 - V_t^2 \right] \right) \leq \left(\frac{1}{2} \right)^2 .$$

$\underline{(k=3)}$: $\tau_3 = (\frac{1}{2})^{2k} = (\frac{1}{2})^6$. Set $F_3 := F(F_2)$ and $s_3 = m_{\tau_3}$.
Again by Lemma 2.2.13, we know that for $s_3 = m_{\tau_3}$

$$\lambda_1 \left(\mathbb{V} \left[\psi_t^{m,F_3} - \bar{\psi}_t^{m,n,F_3} \right] \right) < \tau_3 = \left(\frac{1}{2} \right)^6 \qquad \forall n \geq m \geq s_3.$$

We set $V_t^3 := \psi_t^{s_3,F_3} = F_3 \psi_t^{s_3} = F(F_2) F_2 \psi_t^{s_3} = F_3 \Lambda_{s_3}^{-1/2} P_{s_3} y_t^{s_3} = \bar{\psi}_t^{s_2,s_3,F_2}$ and have therefore that:

$$\left\|V_{jt}^2 - V_{jt}^3\right\| \leq \lambda_1\left(\mathbb{V}\left[V_t^2 - V_t^3\right]\right) \leq \left(\frac{1}{2}\right)^6 .$$

We continue like this for $k = 4, 5, 6, ...$ and obtain sequence $(V_t^k : k \in \mathbb{N})$ which is a Cauchy sequence since:

$$\left\|V_{jt}^k - V_{jt}^{k+h}\right\| \leq \underbrace{\left\|V_{jt}^k - V_{jt}^{k+1}\right\|}_{<(1/2)^k} + \cdots + \underbrace{\left\|V_{jt}^{k+h-1} - V_{jt}^{k+h}\right\|}_{<(1/2)^{k+h}}$$
$$\leq \sum_{l=0}^{h-1}\left(\frac{1}{2}\right)^{k+l} = \left(\frac{1}{2}\right)^k \times \underbrace{\sum_{l=0}^{h-1}\left(\frac{1}{2}\right)^l}_{<2} \leq \left(\frac{1}{2}\right)^{k-1} \to 0 \text{ for } k \to \infty .$$

Since V_t^k is a Cauchy sequence for every t, it converges in $\mathbb{H}(y)$.
We are left with showing that the limit V_t is in the static aggregation space, i.e. $V_t \in \mathbb{S}_t$ for all $t \in \mathbb{Z}$. For this consider $G_k := F_k \Lambda_{s_k}^{-1} P_{s_k}$. Obviously:

$$G_k G_k^* = F_k \Lambda_{s_k}^{-1} F_k^* \text{ so } \quad \lambda_1(G_k G_k^*) \leq \frac{1}{\lambda_r(\Gamma_y^{s_k})} \to 0 \text{ for } k \to \infty ,$$

consequently considering the row vectors, say g_{jk} with $j = 1, ..., r$ from $G_k = \begin{pmatrix} g_{1k} \\ \vdots \\ g_{rk} \end{pmatrix}$, we know that $g_{jk} g_{jk}^* \leq |1/\lambda_r(\Gamma_y^{s_k})| \to 0$ and (g_{jk}) is a SAS. By lemma A.0.3, we know that $\mathbb{V}\, V_t = \mathbb{V} \underline{\lim}_k V_t^k = \lim \mathbb{V}\, V_t^k = I_r$ which completes the proof. ■

The process (V_t) which we define by the limit of the Cauchy sequence V_t^k for $t \in \mathbb{Z}$, is as the following proposition shows a basis for the static aggregation space. However the uniqueness of (V_t) has not been shown, since we start with an arbitrary rotation C of normalised principal components ψ_t^m which are also not unique.

Proposition 2.2.15 (V_t is a basis for $\mathbb{S}_t$)
For all $t \in \mathbb{Z}$, $\overline{\text{sp}}(V_t) = \mathbb{S}_t$.

Proof. Suppose that $w_t \in \mathbb{S}_t$. We have to show that in decomposition of w_t into projection onto $\text{sp}(V_t)$ and residual, i.e.

$$w_t = \underbrace{\text{proj}\left(w_t \mid \text{sp}\left(V_t\right)\right)}_{\in \mathbb{S}_t} + \underbrace{r_t}_{\in \mathbb{S}_t}$$

the residual r_t is equal to zero. For this consider the vector

$$\bar{V}_t := \begin{pmatrix} V_{1t} \\ \vdots \\ V_{rt} \\ r_t \end{pmatrix} = \begin{pmatrix} \bar{V}_{1t} \\ \vdots \\ \bar{V}_{r+1,t} \end{pmatrix}$$

for each of these elements there exist static averaging sequences $\left(\hat{a}_j^{(k)}\right)$ for $j = 1, ..., r+1$, such that $\bar{V}_{jt} = \underline{\lim}_k \hat{a}_j^{(k)} y_t = \underline{\lim}\, \bar{V}_{jt}^k$ and

$$\bar{V}_{jt}^k := \hat{a}_j^{(k)} y_t = \underbrace{\text{proj}\left(\bar{V}_{jt}^k \mid \text{sp}(\psi_t^k)\right)}_{h_{jt}^k} + \rho_{jt}^k = \underbrace{\hat{a}_j^{(k)} P_k^* P_k}_{\hat{f}_j^{(k)}} y_t^k + \underbrace{\hat{a}_j^{(k)} (I_k - P_k^* P_k)}_{\hat{g}_j^{(k)}} y_t^k \tag{2.13}$$

Clearly,

$$\left(\hat{a}_j^{(k)}\right)^* \hat{a}_j^{(k)} = \left(\hat{a}_j^{(k)}\right)^* \left(P_k^* P_k + Q_k^* Q_k\right) \hat{a}_j^{(k)} = \left(\hat{f}_j^{(k)}\right)^* \hat{f}_j^{(k)} + \left(\hat{g}_j^{(k)}\right)^* \hat{g}_j^{(k)} \to 0 \,,$$

which implies that also $\left(\hat{g}_j^{(k)}\right)^* \hat{g}_j^{(k)} \to 0$ and consequently

$$\mathbb{V}\, \rho_{jt}^k = \left(\hat{g}_j^{(k)}\right)^* \Gamma_y^k \hat{g}_j^{(k)} \leq \left(\sup_n \lambda_{r+1}\left(\Gamma_y^n\right)\right) \left(\hat{g}_j^{(k)}\right)^* \hat{g}_j^{(k)} \to 0 \quad \text{for } k \to \infty \,.$$

The main argument of the proof is the following: Denoting by $h_t^k = (h_{1t}^k \cdots h_{r+1,t}^k)^*$ and ρ_t^k likewise, we obtain

$$\begin{aligned} \det \mathbb{V}\, \bar{V}_t &= \det \mathbb{V}\left(\underline{\lim}_k V_t^k\right) = \det \mathbb{V}\, \underline{\lim}_k (h_t^k + \rho_t^k) = \det \mathbb{V}\, \underline{\lim}_k h_t^k \\ &= \det \underline{\lim}_k \mathbb{V}\, h_t^k = \underline{\lim}_k \det \mathbb{V}\, h_t^k \;= 0, \end{aligned}$$

where by Lemma A.0.3 (or by continuity of the inner product), we know that $\det \mathbb{V} \underline{\lim}_k h_t^k = \det \lim_k \mathbb{V} h_t^k$ and the last equality follows since $\det \mathbb{V} h_t^k$ is a zero-sequence ($\mathbb{V} h_t^k$ is singular for every $k \in \mathbb{N}$). ∎

A key insight of this proof is (2.13): In fact we can represent every element in the aggregation space as a limit of a sequence of projections on the first r principal components corresponding to cross-sectional dimension k. If we let $k \to \infty$ the corresponding projection will converge. In the following, we will show that this holds not only for elements in the aggregation space but in general, since the normalised principal components (ψ_t^n) generate a Cauchy sequence of static spaces.
Even though the normalised principal components themselves do not converge (we can always change the direction by a change of sign), the spaces which they span converge in the sense that projections of random variables on the spaces spanned by the principal components (irrespective of directions) converge.

Definition 2.2.16 (Cauchy sequence of static spaces generating process)
Suppose W_t^n is stationary orthonormal, i.e. $\mathbb{V} W_t^n = I_r$ lives in $\mathbb{H}(y)$ and is co-stationary with (y_{it}) for every $n \in \mathbb{N}$. We say that (W_t^n) generates a Cauchy sequence of static spaces if

$$\operatorname{tr} \mathbb{V}\left[W_t^m - \operatorname{proj}(W_t^m \mid \operatorname{sp}(W_t^n))\right] \to 0 \quad \textit{for } m, n \to \infty \ .$$

This feature ensures that the sequence of projections on these spaces converges:

Lemma 2.2.17 (The projection onto a Cauchy sequence of static spaces is convergent)
Let (W_t^n) be a Cauchy sequence of static spaces generating process. Suppose (z_t) is scalar stationary and lives in $\mathbb{H}(y)$ - co-stationary with (y_{it}), then $\operatorname{proj}(z_t \mid \operatorname{sp}(W_t^n))$ *converges in $\mathbb{H}(y)$.*

Proof. For this proof set $\hat{z}_t^n := \operatorname{proj}(z_t \mid \operatorname{sp}(W_t))$, and let r_{nt} be the residual from the projection of z_t on $\mathrm{sp}(W_t)$. We have that

$$\begin{aligned} y_t &= \hat{z}_t^n + r_{nt} = \hat{z}_t^m + r_{mt} \ , \quad \text{so } \hat{z}_t^n - \hat{z}_t^m = r_{mt} - r_{nt} \\ \mathbb{V}(\hat{z}_t^n - \hat{z}_t^m) &= \mathbb{E}(\hat{z}_t^n - \hat{z}_t^m)(\hat{r}_{mt} - \hat{r}_{nt}) = \underbrace{\mathbb{E}\, \hat{z}_t^n r_{mt}}_{\mathcal{S}_1} + \underbrace{\mathbb{E}\, \hat{z}_t^m r_{nt}}_{\mathcal{S}_2} \ , \end{aligned}$$

where the last equal sign follows by the orthogonality relations corresponding to the orthogonal projection. Next, we show that both terms $\mathcal{S}_1, \mathcal{S}_2$ go to zero for $n, m \to \infty$. For this consider writing

$$\begin{aligned} z_t^n &= \hat{z}_t^n + r_{nt} = B^n W_t^n + r_{nt} \\ W_t^m &= \mathrm{proj}(W_t^m \mid \mathrm{sp}(W_t^n)) + \rho_t^{mn} = A^{mn} W_t^n + \rho_t^{mn} \\ \text{so} \quad \mathcal{S}_1 &= \mathbb{E}\, B^m W_t^m r_{nt} = \mathbb{E}\, (B^m A^{mn} W_t^n + B^m \rho_t^{mn})\, r_{nt} = \mathbb{E}\, B^m \rho_t^{mn} r_{nt} \ . \end{aligned}$$

Now by Cauchy-Schwarz inequality, we have

$$|\,\mathbb{E}(B^m \rho_t^{mn} r_{nt})| \leq (\mathbb{V}\, B^m \rho_t^{mn})(\mathbb{V}\, r_{nt}) \to 0 \text{ for } m, n \to \infty \ ,$$

since by assumption $\rho_t^{mn} \to 0$ in mean square. Analogously, we can proceed with $\mathcal{S}_2$. ∎

Lemma 2.2.18
The sequence (ψ_t^n) generates a Cauchy sequence of static spaces.

Proof. We distinguish between two cases:
<u>1. Case:</u> Projection with $n > m$:

$$\begin{aligned} \psi_t^m &= \mathrm{proj}(\psi_t^m \mid \mathrm{sp}(\psi_t^n)) + \rho_t^{mn} = D\psi_t^n + \rho_t^{mn} \\ \text{so} \quad \mathrm{tr}(\mathbb{V}\, \rho_t^{mn}) &\leq q\lambda_1(\mathbb{V}\, \rho_t^{mn}) \leq r\frac{\lambda_{n,r+1}}{\lambda_{m,r}} \end{aligned}$$

where $\lambda_{n,j} = \lambda_j(\Gamma_y^n)$.
<u>2. Case:</u> Projection of ψ_t^n on $\mathrm{sp}(\psi_t^m)$ with $n > m$: Since $D = \Lambda_m^{1/2} P_n P_m^* \Lambda_n^{-1/2}$ we

have $D^{-1} = \Lambda_n^{1/2} P_m P_n^* \Lambda_m^{-1/2} = D^*$. We now show that

$$D^* \psi_t^m = \operatorname{proj}(\psi_t^n \mid \operatorname{sp}(\psi_t^m)) \qquad \text{i.e. } \rho_t^{nm} = \psi_t^n - D^* \psi_t^m \perp D^* \psi_t^m$$

where the equivalence of these statements follows by the projection theorem.

$$\begin{aligned}
&\mathbb{E}\left(\Lambda_n^{-1/2} P_n - D^* \Lambda_m^{1/2} P_m y_t^n\right)\left(D^* \Lambda_m^{-1/2} P_m y_t^n\right)^* \\
=&\Lambda_n^{-1/2} P_n \underbrace{\Gamma_n^y}_{P_n^* \Lambda_n P_n + Q_n^* \Lambda_n Q_n} P_m^* \Lambda_m^{-1/2} D - D^* \Lambda_m^{-1/2} P_m \underbrace{\Gamma_n^y}_{P_m^* \Lambda_m P_m + Q_m^* \Lambda_m Q_m} P_m^* \Lambda_m^{-1/2} D \\
=& \Lambda_n^{-1/2} \underbrace{P_n P_n^*}_{I_r} \Lambda_n P_n P_m^* \Lambda_m^{-1/2} D - D^* D = D^* D - D^* D = 0 \ .
\end{aligned}$$

It follows that

$$\begin{aligned}
&I_r = \mathbb{V}\,\psi_t^m = DD^* + \mathbb{V}\,\rho_t^{nm} = \mathbb{V}\,\psi_t^n = D^* D + \mathbb{V}\,\rho_t^{mn} \\
\text{so} \quad &\operatorname{tr} \mathbb{V}\,\rho_t^{nm} = \operatorname{tr} \mathbb{V}\,\rho_t^{mn} \to 0 \text{ for } n, m \to \infty \ .
\end{aligned}$$

■

Lemma 2.2.19
The sequence $(\mathcal{K}_{ni} : n \in \mathbb{N})$, is a static averaging sequence.

Proof. Consider

$$\operatorname{proj}(y_{it} \mid \operatorname{sp}(\psi_t^n)) = \pi_{ni} \Lambda_n^{1/2} \underbrace{\Lambda_n^{-1/2} P_n y_t^n}_{\psi_t^n}$$

$$\begin{aligned}
&\pi_{ni} \pi_{ni}^* \lambda_{nq} \leq \pi_{ni} \Lambda_n \pi_{ni}^* = \mathbb{V}\,\gamma_{it}^n \leq \mathbb{V}\,y_{it}^n < \infty \ , \\
\text{so} \quad &\mathcal{K}_{ni} \mathcal{K}_{ni}^* = \pi_{ni} \pi_{ni}^* \leq \frac{\mathbb{V}\,y_{it}}{\lambda_{nq}} \to 0 \text{ for } m \to \infty \ ,
\end{aligned}$$

so $(\mathcal{K}_{ni})$ is a static averaging sequence for every $i \in \mathbb{N}$. ■

Proposition 2.2.20
If (y_{it}) is a r-SFS, then the i-th element in the SLRA of rank r converges in mean

square to the projection of y_{it} on the static aggregation space, i.e.

$$\underline{\lim}_n \operatorname{proj}(y_{it} \mid \operatorname{sp}(\psi_t^n)) = \underline{\lim}_n \mathcal{K}_{ni} y_t^n = \underline{\lim}_n \gamma_{it}^n = \operatorname{proj}(y_{it} \mid \mathbb{S}_t) \ .$$

Proof. Since (ψ_t^n) generates a Cauchy sequence of static spaces (Lemma 2.2.18) it follows by Lemma 2.2.17, that $\operatorname{proj}(y_{it} \mid \operatorname{sp}(\psi_t^n)) = \mathcal{K}_{ni} y_t^n =: \gamma_{it}^n$ is convergent in L^2 to the limit say $\gamma_{it}^\times$. Call the L^2 limit of the residual $\delta_{it}^\times = y_{it} - \gamma_{it}^\times = \underline{\lim}_n \delta_{it}^n$, say. By Lemma 2.2.19, we know that $(\mathcal{K}_{ni})$ is a SAS, therefore $\gamma_{it}^\times \in \mathbb{S}_t$.
We show that $\delta_{it}^\times \perp \mathbb{S}_t$ which implies that $\gamma_{it}^\times = \operatorname{proj}(y_{it} \mid \mathbb{S}_t)$. For this suppose $z_t \in \mathbb{S}_t$, so we can represent $z_t = \underline{\lim}_n \hat{a}^{(n)} \psi_t^n$ (see proof of proposition 2.2.15), thus

$$\langle z_t, \delta_{it}^\times \rangle = \langle \underline{\lim}_n \hat{a}^{(n)} \psi_t^n, \underline{\lim}_n \delta_{it}^n \rangle = \lim_n \langle \hat{a}^{(n)} \psi_t^n, \delta_{it}^n \rangle = 0 \ ,$$

by continuity of the inner product. ∎

Lemma 2.2.21
The double sequence of residuals (δ_{it}) from the orthogonal projection of the output on the static aggregation space, i.e. $\gamma_{it} = \operatorname{proj}(y_{it} \mid \mathbb{S}_t)$ is statically idiosyncratic.

Proof. Let $\delta_t^n = (\delta_{1t}, ..., \delta_{nt})'$ be the n-dimensional L^2 limit of $\delta_t^{m,n} = (\delta_{1t}^n, ..., \delta_{nt}^n)'$. Set $\Gamma_\delta^n := \mathbb{V}\, \delta_t^n$ and $\Gamma_\delta^{m,n} = \mathbb{V}\, \delta_t^{m,n}$. To show that (δ_{it}) is statically idiosyncratic, we use the characterisation theorem and show that $\sup_n \lambda_1(\Gamma_\delta^n) \leq \sup_n \lambda_{r+1}(\Gamma_y^n) < \infty$.
Let m be fix. Since $\underline{\lim}_n \delta_{it}^n = \delta_{it}$ by proposition 2.3.18, it follows that also starting with $n > m$, the sequence of variance matrices converges, i.e. $\Gamma_\delta^{n,m} \to \Gamma_\delta^m$ (note that the roles of n, m are exchanged here). By continuity of eigenvalues we have

$$\lim_n \lambda_1 \left(\underbrace{\Gamma_\delta^{n,m}}_{m \times m} \right) = \lambda_1 \left(\lim_n \Gamma_\delta^{n,m} \right) = \lambda_1 \left(\Gamma_\delta^m \right) \ ,$$

but also $\lambda_1(\Gamma_\delta^{n,m}) \leq \lambda_1(\Gamma_\delta^{n,n}) = \lambda_{r+1}(\Gamma_y^n)$ for all $n \geq m$ therefore $\lambda_1(\Gamma_\delta^n) \leq \sup_n \lambda_{r+1})(\Gamma_y^n)$ and since this holds for all m, also $\sup_n \lambda_1(\Gamma_\delta^n) \leq \sup_n \lambda_{r+1}(\Gamma_y^n)$ so (δ_{it}) is idiosyncratic by theorem 2.2.6. ∎

Proposition 2.2.22
Suppose that (y_{it}) has a factor model representation as in (2.10) satisfying the assumptions from theorem 2.2.7.1, then
1. $\overline{\mathrm{sp}}(C_t) = \mathrm{sp}(F_t) = \mathbb{S}_t$
2. $C_{it} = \mathrm{proj}(y_{it} \mid \mathbb{S}_t)$.

Proof. Suppose (y_{it}) has a representation as a factor model as in equation (2.10), so

$$y_{it} = C_{it} + e_{it} = \Lambda_i F_t + e_{it}$$

while $\sup_n \lambda_r(\Gamma_C^n) = \infty$ and $\sup_n \lambda_1(\Gamma_e^n) < \infty$.
Furthermore, we know that

$$\begin{aligned} y_{it} &= \mathrm{proj}(y_{it} \mid \mathbb{S}_t) + \delta_{it} = \gamma_{it} + \delta_{it} \\ &= \underbrace{c_i}_{1\times r}\underbrace{V_t}_{r\times 1} + \delta_{it} \text{ from proposition 2.2.15 .} \end{aligned}$$

Since (e_{it}) is idiosyncratic, we know that $\mathbb{S}_t \subset \mathbb{H}(C_t)$, because for any $z_t = \underline{\lim}_k\, \hat{c}^{(k)} y_t \in \mathbb{S}_t$, where $(\hat{c}^{(k)})$ is the SAS corresponding to z_t, we have

$$z_t = \underline{\lim}_k\, \hat{c}^{(k)} y_t = \underline{\lim}_k\, \hat{c}^{(k)}(C_t + e_t) = \underline{\lim}_k\, \hat{c}^{(k)} C_t \ .$$

But also $\overline{\mathrm{sp}}(C_t) \subset \mathrm{sp}(F_t)$ since $C_{it} = \Lambda_i F_t$. It follows that

$$\mathrm{sp}(V_t) = \mathbb{S}_t \subset \overline{\mathrm{sp}}(C_t) \subset \mathrm{sp}(F_t).$$

On the other hand, we know that $\mathbb{V}\, F_t = I_r = \mathbb{E}\, V_t V_t'$, so noting that there exists an $r \times r$ matrix b, we can write $V_t = \underbrace{b}_{r\times r} F_t$ while $bb' = I_r$, which implies that b is invertible and consequently $F_t = b^{-1} V_t$, so from

$$\mathrm{sp}(F_t) \subset \mathrm{sp}(V_t) = \mathbb{S}_t$$
$$\text{it follows that } \mathbb{S}_t = \mathrm{sp}(F_t) = \overline{\mathrm{sp}}(C_t)$$

and $C_{it} \in \mathbb{S}_t$ and $e_{it} \perp \mathbb{S}_t = (C_t)$ which implies that $C_{it} = \text{proj}(y_{it} \mid \mathbb{S}_t)$ and $\delta_{it} = e_{it}$ by the projection theorem. ■

2.2.4 Real Valued Data

Observed data is usually real valued. In this section we show that we can formulate the structural results presented above also for real valued (y_{it}): We show that if (y_{it}) is real valued, the static common component and the static idiosyncratic component are as well and we can choose (F_t) to be real valued.
First, recall that if (y_{it}) is real valued, in terms of orthogonal projections, we can without loss of generality only consider real valued spans:

Remark 2.2.23 (see Deistler and Scherrer (2022), excercise on p.55)
Let (y_{it}) be a real valued stochastic double sequence and

- *let $\mathbb{H}^{\mathbb{R}}(y)$ be the real expansion of (y_{it}), i.e. we use only real numbers as scalars,*
- *denote by $\mathbb{H}^{\mathbb{C}}(y)$ the corresponding complex span, with complex numbers as scalars,*
- *let $\mathbb{M} \subset \mathbb{H}^{\mathbb{R}}(y)$ be a subspace of $\mathbb{H}^{\mathbb{R}}(y)$,*
- *let $\mathbb{M}^{\mathbb{C}} := \overline{\text{sp}}(z \mid z \in \mathbb{M}^{\mathbb{R}}) \subset \mathbb{H}^{\mathbb{C}}(y)$ be the subspace generated by $\mathbb{M}$ in $\mathbb{H}^{\mathbb{C}}(y)$,*
- *let $z \in \mathbb{H}^{\mathbb{R}}(y)$ be real valued,*

then the projection of z on $\mathbb{M}$ in $\mathbb{H}^{\mathbb{R}}(y)$ is the same as the projection of z on $\mathbb{M}^{\mathbb{C}}$ in $\mathbb{H}^{\mathbb{C}}(y)$.

If $\Gamma_y^n \in \mathbb{R}^{n \times n}$ is real, we can choose real orthonormal eigenvectors for $P_{(n)}$ and all eigenvalues of Γ_y^n are real, so equation (2.6) becomes

$$\Gamma_y^n = P'_{(n)} \Lambda_{(n)} P_{(n)} \ . \tag{2.14}$$

Accordingly, for real valued (y_{it}) instead of equation (2.8) and (2.9) we can write

$$\begin{aligned} \mathcal{K}_{ni} &:= \mathcal{K}_{ni}(\Gamma_y^n) = p'_{ni}P_n \quad \text{the } i\text{-th row of } P'_nP_n \\ C_t^{[n]} &:= P'_nP_ny_t^n \\ C_{it,n} &:= \mathcal{K}_{ni}y_t^n \quad \text{the } i\text{-th row of } C_t^{[n]} \ . \end{aligned} \tag{2.15}$$

Now define $\mathbb{S}_t^{\mathbb{C}}(y)$ as the "complex aggregation space", which we obtain from building the limits with complex SAS, i.e. $\hat{c}^{(k)} \in \mathbb{C}^{1\times\infty}$. Set $\mathbb{S}_t^{\mathbb{R}}(y)$ as the "real aggregation space", obtained from limits with real SAS, i.e. $\hat{c}^{(k)} \in \mathbb{R}^{1\times\infty}$. Define

$$\begin{aligned} C_{it}^{\mathbb{C}} &:= \operatorname{proj}(y_{it} \mid \mathbb{S}_t^{\mathbb{C}}(y)) \quad \text{in} \quad \mathbb{H}^{\mathbb{C}}(y) \\ C_{it}^{\mathbb{R}} &:= \operatorname{proj}(y_{it} \mid \mathbb{S}_t^{\mathbb{R}}(y)) \quad \text{in} \quad \mathbb{H}^{\mathbb{R}}(y) \ , \end{aligned}$$

where $\mathbb{H}^{\mathbb{R}}(y), \mathbb{H}^{\mathbb{C}}(y)$ were defined in remark 2.2.23. Note that $C_{it}^{\mathbb{C}}$ corresponds to C_{it} in theorem 2.2.7.

Corollary 2.2.24
Suppose (y_{it}) is a real valued static factor sequence. We have $C_{it}^{\mathbb{R}} = C_{it}^{\mathbb{C}} =: C_{it}$. Furthermore, the factors (F_t) and factor loadings Λ_i can be chosen to be real valued. Also static common and static idiosyncratic component are real valued.

Proof. By remark 2.2.23, we know that

$$\begin{aligned} C_{it}^{\mathbb{R}} &= \operatorname{proj}(y_{it} \mid \mathbb{S}_t^{\mathbb{R}}(y)) \quad \text{in } \mathbb{H}^{\mathbb{R}}(y) \\ &= \operatorname{proj}\left(y_{it} \mid \left(\mathbb{S}_t^{\mathbb{R}}(y)\right)^{\mathbb{C}}\right) \ , \end{aligned}$$

where $\left(\mathbb{S}_t^{\mathbb{R}}(y)\right)^{\mathbb{C}}$ is the subspace generated by $\mathbb{S}_t^{\mathbb{R}}(y)$ in $\mathbb{H}^{\mathbb{C}}(y)$. Now, working with the real eigen-decomposition of equation (2.14), we can reproduce the proofs from section 2.2.3 by using only real eigenvectors and eigenvalues. This produces a real valued $r \times 1$ basis, say $(V_t^{\mathbb{R}})$, in proposition 2.2.15 of $\mathbb{S}_t^{\mathbb{R}}(y)$. However, $V_t^{\mathbb{R}}$ is also a basis for $\mathbb{S}_t^{\mathbb{C}}(y)$, replacing equation (2.13) by

$$\bar{V}_{jt}^k := \hat{a}_j^{(k)} y_t = \underbrace{\operatorname{proj}\left(\bar{V}_{jt}^k \mid \operatorname{sp}(\psi_t^k)\right)}_{h_{jt}^k} + \rho_{jt}^k = \underbrace{\hat{a}_j^{(k)} P_k' P_k}_{\hat{f}_j^{(k)}} y_t^k + \underbrace{\hat{a}_j^k (I_k - P_k' P_k)}_{\hat{g}_j^{(k)}} y_t^k \ ,$$

where $\hat{a}_j^{(k)} \in \mathbb{C}^{1\times\infty}$ and P_k are the real eigenvectors from (2.14) and proceeding analogously with the proof. It follows that

$$\left(\mathbb{S}_t^{\mathbb{R}}(y)\right)^{\mathbb{C}} = \operatorname{sp}\left(V_t^{\mathbb{R}}\right)^{\mathbb{C}} = \mathbb{S}_t^{\mathbb{C}}(y) \ .$$

Therefore $C_{it}^{\mathbb{R}} = C_{it}^{\mathbb{C}} = C_{it}$, where C_{it} is the static common component from theorem 2.2.7, which completes the proof. ■

2.2.5 Estimation

We will present the section in terms of real valued observations (y_{it}). In Chamberlain and Rothschild (1983a) only population results are provided and estimation is not investigated. Proofs for the consistent estimation of the approximate factor model via principal components have first been given in the seminal work of Stock and Watson (2002a); Bai and Ng (2002, 2020, 2021). In these papers idiosyncraticness is formalised and assumed by limiting cross-sectional and time dependence in (e_{it}). This involves quite technical conditions compared to assuming that $\sup_n \lambda_1(\Gamma_e^n) < \infty$ (see A2). To provide the link to the theory presented above, we prove consistency of the static principal components estimator under A2. A key element of the proof is to use that we know already that the population SLRA of rank-r converges in mean square to the true common component (theorem 2.2.7.3).

Assumption 3 (Consistent Estimation of the Covariance)
Let $\gamma_{ij}^T(0)$ be an estimator for the covariance $\gamma_{ij}(0) = \mathbb{E}\, y_{it} y_{jt}$ using T time observations. For every $\delta > 0$, we have

$$\lim_T \mathbb{P}(|\gamma_{ij}^T(0) - \gamma_{ij}(0)| > \delta) = 0.$$

For example Brockwell and Davis (2009), ch. 7, discuss consistent estimation of the auto-covariance in the stationary case. Accordingly, we define the coordinates of the sample rank-r approximation as follows:

$$\hat{\Gamma}_y^n := \left(\gamma_{ij}^T(0)\right)_{i,j=1,\dots,n} \tag{2.16}$$

$$\hat{C}_{it} := \mathcal{K}_{ni}^T y_t^n \ , \tag{2.17}$$

which depend both on (n, T) and where $\mathcal{K}_{ni}^T$ is obtained from equation (2.7), replacing the population variances by their sample counterparts $\gamma_{ij}^T(0)$.

Theorem 2.2.25 (Consistency of the Static Rank-r Approximation)
If (y_{it}) is a r-SFS and A3 holds, then $\hat{C}_{it,n} \xrightarrow{P} C_{it}$ in probability for $\min(n, T) \to \infty$ for all $i \in \mathbb{N}$.

The corresponding result for the dynamic rank-q approximation has been shown in Forni et al. (2000). In the following proof, we employ similar techniques as in Forni et al. (2000).

Proof of theorem 2.2.25. Since the suitably normalised eigenvectors are continuous functions of the matrix entries, we know that for all $\delta, \eta > 0$ there exists $T_1(n, \delta, \eta)$ such that for all $T \geq T_1(n, \delta, \eta)$

$$\mathbb{P}(\left\|\mathcal{K}_{ni}^T - \mathcal{K}_{ni}\right\| > \delta) \leq \eta \ .$$

Therefore we can write:

$$\begin{aligned}\mathbb{P}\left(|\mathcal{K}_{ni}^T y_t^n - C_{it}| > \delta\right) &\leq \mathbb{P}\left(|(\mathcal{K}_{ni}^T - \mathcal{K}_{ni}) y_t^n| > \delta/2\right) + \mathbb{P}\left(|\mathcal{K}_{ni} y_t^n - C_{it}| > \delta/2\right) \\ &= R_{n1}^T + R_{n2}^T \text{ , say.}\end{aligned} \tag{2.18}$$

Let $\varepsilon > 0$, we show there exists n, T such that $R_{n1}^T + R_{n2}^T < \varepsilon$. By theorem 2.2.7.4, we know that R_{n2}^T converges to 0 for $n \to \infty$, so choose n large enough such that $R_{n2}^T < \varepsilon/2$.
Next, we show that we can choose T large enough, such that $R_{n1}^T < \varepsilon/2$. Set

$B_{ni}^T := (\mathcal{K}_{ni}^T - \mathcal{K}_{ni})$ and $A_\delta := \left\{\left\|B_{ni}^T\right\| \leq \delta\right\}$ and $A_\delta^c = \Omega \setminus A_\delta$, then for some $\eta > 0$, we have

$$
\begin{aligned}
R_{n1}^T &= \mathbb{P}\left(\left\{|B_{ni}^T y_t^n| > \delta/2\right\} \cap \left(\left\{\left\|B_{ni}^T\right\| \leq \delta\right\} \cup \left\{\left\|B_{ni}^T\right\| > \delta\right\}\right)\right) \\
&\leq \mathbb{P}\left(\left\{|B_{ni}^T y_t^n| > \delta/2\right\} \cap A_\delta\right) + \mathbb{P}\left(A_\delta^c\right) \\
&\leq \mathbb{P}\left(|B_{ni}^T y_t^n| > \delta/2 \mid A_\delta\right) + \eta(\delta) \quad \text{for } \; T \geq T_1(n, \delta, \eta) \; .
\end{aligned}
$$

Next, by Chebyshev's inequality and the properties of conditional expectation, we have

$$
\mathbb{P}\left(|B_{ni}^T y_t^n| > \delta/2 \mid A_\delta\right) \leq \frac{\mathbb{E}\left(|B_{ni}^T y_t^n|^2 \mid A_\delta\right)}{(\delta^2/4)} \; .
$$

$$
\begin{aligned}
\mathbb{E}\left(|B_{ni}^T y_t^n|^2 \mid A_\delta\right) &= \frac{\int_{A_\delta} B_{ni}^T y_t^n y_t^{n\prime} B_{ni}^{T\prime}}{\mathbb{P}(A_\delta)} d\,\mathbb{P} \leq 2 \int_{A_\delta} \left\|B_{ni}^T\right\|^2 \underbrace{\frac{B_{ni}^T}{\|B_{ni}^T\|}}_{l_{ni}^T} y_t^n y_t^{n\prime} \frac{B_{ni}^{T\prime}}{\|B_{ni}^T\|} d\,\mathbb{P} \\
&\leq 2\delta^2 \int_{A_\delta} l_{ni}^T y_t^n y_t^{n\prime} l_{ni}^{T\prime} d\,\mathbb{P} \leq 2\delta^2 \int_{A_\delta} \lambda_1(y_t^n y_t^{n\prime}) d\,\mathbb{P} \\
&\leq 2\delta^2\, \mathbb{E}\, \lambda_1(y_t^n y_t^{n\prime}) \leq 2\delta^2 \lambda_1(\mathbb{V}\, y_t^n) < \infty \; .
\end{aligned}
$$

where the last inequality, we used Jensen's inequality since the first eigenvalue $\lambda_1(\cdot)$ is a continuous and convex function in the set of non-negative definite matrices.

■

In Stock and Watson (2002a); Bai and Ng (2002, 2020, 2021) the authors provide consistency rates and prove consistent estimation of the factors and the loadings-matrix. Here, we do not provide rates and prove consistency for the common component (C_{it}) instead. Furthermore note also that Stock and Watson (2002a); Bai and Ng (2002, 2020, 2021) start from assumptions about an underlying true factor model rather than imposing assumptions on the structure of the output process (A2).

2.2.6 Time Dependent Variance

In this section, we discuss heteroscedasticity and non-stationarity. Firstly, note that the idiosyncratic part in the variance matrix Γ_y^n is irrelevant for the computation of the common component:

Corollary 2.2.26
Let (y_{it}) be a r-SFS with $y_{it} = C_{it} + e_{it}$, then for all Γ^n with $\sup_n \lambda_1(\Gamma^n) < \infty$, we have

$$C_{it} = \underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) y_t^n \ ,$$

where $\mathcal{K}_{ni}(\cdot)$ is defined in (2.15).

Proof. From lemma 2.2.19, we know that $\mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n)$ is a SAS since $\sup_n \lambda_r(\Gamma_C^n + \Gamma^n) = \infty$ and $\sup_n \lambda_{r+1}(\Gamma_C^n + \Gamma^n) < \infty$ by the same arguments as in (2.4), (2.5). Suppose (ν_{it}) is a statically idiosyncratic double sequence with $\mathbb{V}\, \nu_t^n = \Gamma^n$. It follows that

$$\begin{aligned}
\underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) y_t^n &= \underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) C_t^n + \underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) e_t^n \\
&= \underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) C_t^n \ , \quad \text{since } \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) \text{ is a SAS} \\
&= \underline{\lim}_n \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n)(C_t^n + \nu_t^n) \ , \quad \text{since } \mathcal{K}_{ni}(\Gamma_C^n + \Gamma^n) \text{ is a SAS} \\
&= C_{it} \ , \quad \text{by theorem 2.2.7.}
\end{aligned}$$

■

Assumption 4 (Heteroscedasticity)
Suppose (y_{it}) is an r-SFS with $y_{it} = \Lambda_i F_t + e_{it} = C_{it} + e_{it}$, where C_{it} is weakly stationary and (e_{it}) is such that

(i) Heteroscedasticity: $\lambda_1(\Gamma_e^n(t)) < \infty$ for all $t \in \mathbb{Z}$

(ii) Average variance: $\lim_T T^{-1} \sum_{t=1}^T \Gamma_e^n(t) = \bar{\Gamma}_e^n$, and $\sup_n \lambda_1(\bar{\Gamma}_e^n) < \infty$

(ii) $\hat{\Gamma}_y^n \xrightarrow{P} \Gamma_C^n + \bar{\Gamma}_e^n$ in probability for $t \to \infty$.

We can estimate the static common component consistently also under heteroscedasticity:

Corollary 2.2.27
Suppose A4 holds for (y_{it}), then $\hat{C}_{it,n} \xrightarrow{P} C_{it}$ in probability for $\min(n,T) \to \infty$ for all $i \in \mathbb{N}$.

Proof. We replace $\mathcal{K}_{ni}(\Gamma_y^n)$ by $\mathcal{K}_{ni}(\Gamma_C^n + \bar{\Gamma}_e^n)$ and $\mathcal{K}_{ni}^T(\hat{\Gamma}_y^n(t))$ by $\mathcal{K}_{ni}^T(\hat{\Gamma}_y^n)$ in the proof of theorem 2.2.25. By corollary 2.2.26, we know that $R_{n2}^T \to 0$ in equation (2.18). Complete the proof by proceeding in the same way as in the proof of theorem 2.2.25. ■

The setup of Stock and Watson (2002a); Bai and Ng (2002, 2020, 2021) restricts serial correlation in the idiosyncratic component. However this restriction is kind of unnecessary, since by definition a statically idiosyncratic double sequence only limits the contemporaneous covariation and not the covariation over time, as 2.2.28 illustrates.

Example 2.2.28 (Random Walks as Static Idiosyncratic Component)
Let (u_t) be a scalar white noise process with unit variance and let $e_{it} = e_{i,t-1} + \varepsilon_{it}^e$ be a random work where (ε_{it}^e) is zero-mean i.i.d. with variance σ^2 - orthogonal to u_t. Say $e_{i0} = 0$ for all $i \in \mathbb{N}$.
Consider the following factor model

$$y_{it} = u_t + e_{it} = u_t + e_{i,t-1} + \varepsilon_{it}^e = C_{it} + e_{it} \ .$$

The variance matrix of y_t^n depends on t:

$$\mathbb{V}\, y_t^n = \Gamma_y^n(t) = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} + t \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} = \Gamma_C^n + \Gamma_e^n(t) \ .$$

Here for all t we have $\lambda_1(\Gamma_C^n(t)) = n \to \infty$ for $n \to \infty$ and $\lambda_1(\Gamma_e^n(t)) = t\sigma^2 < \infty$ for all n, which satisfies the conditions of theorem 2.2.7.

In general, if we have a procedure to consistently estimate $\Gamma_y^n(t)$, we can consistently estimate common and idiosyncratic component also in the non-stationary case as long as A2 holds for every $t \in \mathbb{Z}$.

2.3 Dynamic Factor Sequences: Italian School

All results and definitions stated in sections 2.3, 2.3.1 are from Forni and Lippi (2001). The common component in the *Generalized Dynamic Factor Model* introduced in Forni et al. (2000); Forni and Lippi (2001) emerges from dynamic aggregation rather than from static aggregation: This means that we consider limits of averaging sequences, which are weighted averages over time and cross-section:

$$z_t = \underline{\lim}_k \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} c_{ij}^{(k)} y_{i,t-j} = \underline{\lim}_k c^{(k)}(L) y_t \ . \tag{2.19}$$

Forni and Lippi (2001) allow very general transfer functions from the "dynamic shocks" to what we call the *dynamic* common component (see also Forni et al., 2005).

Definition 2.3.1 (Dynamic Averaging Sequence (DAS))
Let $c^{(k)} \in L_2^\infty(I) \cap L_2^\infty(f_y)$ for $k \in \mathbb{N}$. The sequence of filters $(c^{(k)} : k \in \mathbb{N})$ is called Dynamic Averaging Sequence (DAS) if

$$\lim_k \left\| c^{(k)} \right\|_{L_2^\infty(I)} = \int_\Theta c^{(k)}(\theta) \left(c^{(k)}(\theta) \right)^* d\theta = 0 \ .$$

If $\left(c^{(k)}\right)$ in equation (2.19) is a DAS, the scalar valued output process (z_t) is called *dynamic aggregate*. It is useful to introduce a notation for the set of all DAS corresponding to (y_{it}):

$$\mathcal{D}(f_y) := \left\{ \left(c^{(k)}\right) : c^{(k)} \in L_2^\infty(I) \cap L_2^\infty(f_y) \ \forall k \in \mathbb{N} \text{ and } \lim_k \left\| c^{(k)} \right\|_{L_2^\infty(I)} = 0 \right\} \ .$$

Definition 2.3.2 (Dynamic Aggregation Space)
The set $\mathbb{G}(y) := \left\{ z_t : z_t = \underline{\lim}_k \underline{c}^{(k)}(L) y_t \text{ and } \left(c^{(k)}\right) \in \mathcal{D}(f_y) \right\}$ is called Dynamic

Aggregation Space.

Henceforth we might often write $\mathbb{G}$ to denote $\mathbb{G}(y)$ when it is clear from the context. For a stationary double sequence (y_{it}) (see A1), the dynamic aggregation space $\mathbb{G}(y)$ is a closed subspace of the time domain $\mathbb{H}(y)$ (see Forni and Lippi, 2001, Lemma 6). Therefore orthogonal projections onto $\mathbb{G}$ are again in $\mathbb{G}$.

Definition 2.3.3 (Dynamically Idiosyncratic)
We call a stationary (A1) double sequence (z_{it}) dynamically idiosyncratic, if

$$\underline{\lim}_k \underline{c}^{(k)}(L) z_t = 0$$

for all $(c^{(k)}) \in \mathcal{D}(f_z)$.

In other words, a double sequence which is dynamically idiosyncratic vanishes under all possible dynamic aggregations. Analogously to theorem 2.2.6, we can characterise dynamic idiosyncraticness by the boundedness of the first eigenvalue of the spectrum (see Forni and Lippi, 2001, Thm 1).

Theorem 2.3.4 (Forni and Lippi (2001), Dynamically Idiosyncratic)
The following statements are equivalent:

(i) A stochastic double sequence (z_{it}) is dynamically idiosyncratic.

(ii) The first eigenvalue of spectrum is essentially bounded, i.e.

$$\operatorname{ess\,sup}_\theta \sup_n \lambda_1(f_z^n)(\theta) < \infty \ .$$

From what follows in the remainder of this section, we may say that a stationary stochastic double sequence (y_{it}) with the properties specified in A5 below has a dynamic factor structure:

Assumption 5 (q-Dynamic Factor Structure)
There exists $q < \infty$, such that

(i) $\sup_n \lambda_q(f_y^n) = \infty$ *a.e. on* Θ.

(ii) $\operatorname{ess\,sup}_\theta \sup_n \lambda_{q+1}(f_y^n) < \infty$,

where "ess sup" denotes the essential supremum of a measurable function. Analogously to the static case, we can also compute dynamic low rank approximations with dynamic principal components (see Brillinger, 2001, ch. 9). For this, consider the eigen-decomposition of the spectrum

$$f_y^n(\theta) = P_{(n)}^*(\theta)\Lambda_{(n)}(\theta)P_{(n)}(\theta) \text{ for } \theta \in \Theta \ ,$$

where $P_{(n)}(\theta)$ is a unitary matrix of row eigenvectors, $\Lambda_{(n)}(\theta)$ is a diagonal matrix of eigenvalues sorted from the largest to the smallest both having equivalence classes of (measurable) functions as entries. Denote by $p_{nj}(\theta)$ the j-th row of $P_{(n)}(\theta)$ and by $P_{nq}(\theta) = P_n(\theta)$ the sub unitary matrix consisting of the first q rows of $P_{(n)}(\theta)$. We set

$$\begin{aligned} \mathcal{K}_{ni}(\theta) &:= p_{ni}^*(\theta)P_n(\theta) \quad \text{the } i\text{-th row of } P_n^*(\theta)P_n(\theta) \\ \chi_t^{[n]} &:= \underline{P}_n^*(L)\underline{P}_n(L)y_t^n \qquad (2.20) \\ \chi_{it,n} &:= \underline{\mathcal{K}}_{ni}(L)y_t^n \quad \text{the } i\text{-th row of } \chi_t^{[n]} \ . \qquad (2.21) \end{aligned}$$

where (2.20) is the *dynamic low rank approximation* (DLRA) of rank q. The process $\chi_t^{[n]}$ emerges from a filter of rank q a.e. applied to y_t^n in order to best approximate y_t^n with respect to mean squares.
Here comes the main result for the dynamic case:

Theorem 2.3.5 (Forni and Lippi (2001): q-Dynamic Factor Sequence or q-DFS)
Suppose A1 holds, then
1. A5 holds if and only if we can decompose

$$y_{it} = \chi_{it} + \xi_{it} = \underline{b}_i(L)u_t + \xi_{it} \quad \textit{where } \mathbb{E}\,\chi_{it}\xi_{js} = 0 \quad \forall i,j,t,s, \qquad (2.22)$$

such that u_t is an orthonormal white noise $q \times 1$ process, $b_i(\theta)$ is a square summable (in general two-sided) filter and (χ_{it}), (ξ_{it}) are stationary double sequences (fullfilling A1) with *(i)* $\sup_n \lambda_q \left(f_\chi^n\right) = \infty$ *a.e. on* Θ

(ii) $\operatorname{ess\,sup}_\theta \sup_n \lambda_1 \left(f_\xi^n\right) < \infty$.

Furthermore in this case, it holds that

2. $\chi_{it} = \underline{\lim}_n \chi_{it,n}$,

3. q, χ_{it} *and* ξ_{it} *are uniquely determined from the output* (y_{it}),

4. $\chi_{it} = \text{proj}(y_{it} \mid \mathbb{G})$ *for all* $i \in \mathbb{N}, t \in \mathbb{Z}$.

The proof is given in section 2.3.1. The double sequence (χ_{it}) is called *dynamic common component* (dynamic CC) and (ξ_{it}) is called *dynamic idiosyncratic component* since it is dynamically idiosyncratic by theorem 2.3.4. The equivalence statement 1. justifies the wording *dynamic factor sequence* as the "structure" described in A5 corresponds to an underlying factor model. Therefore we also call the "only if" part in theorem 2.3.5.1 *representation result.* By 4. we obtain the dynamic common component through projecting the output on the dynamic aggregation space. Statement 2. provides the link to dynamic principal component analysis: the dynamic common component is the mean square limit of the DLRA of rank q (see equation 2.21).

Definition 2.3.6 (Forni and Lippi (2001): q-Dynamic Factor Sequence (q-DFS)) *A stationary stochastic double sequence (assumption A1) that satisfies A5 is called q-Dynamic Factor Sequence, q-DFS.*

Methods to estimate q have been proposed by Hallin and Liška (2007); Bai and Ng (2007). In Onatski (2009) a test for determining q is provided.

2.3.1 Proof of the Main Result: Dynamic Case

In this section we prove theorem 2.3.5. In the proof of theorem 2.2.7, we dealt with the static principal components arising from the eigen-decomposition of the variance. The proof of theorem 5 for dynamic factor sequences is mainly analogous to the static case, but here we operate with dynamic rather than static principal components. We replace the eigenvectors/eigenvalues of Γ_y^n with the eigenvectors/eigenvalues of the spectrum spec $y_t^n = f_y^n(\theta)$. As such, they are filters, but we will suppress that in the notation for simplicity. We will use the following notation throughout the proof in this section:

- **canonical projection dynamic case**: $y_{it} = \text{proj}(y_{it} \mid \mathbb{G}) + \delta_{it} = \gamma_{it} + \delta_{it}$ where γ_{it} is the orthogonal Hilbert space projection onto the dynamic aggregation space

- For a multivariate stationary zero-mean stochastic vector process (z_t) with existing spectrum, we denote by $\text{spec}(v_t)$ the spectrum of (z_t)

- $\lambda_i(A)$ is the i-th largest eigenvalue of a square matrix A. If A is a spectral density $\lambda_i(A)$ is an equivalence class of (measurable) functions in the frequency $\theta \in [-\pi, \pi]$.

- $f_y^n(\theta) = P_n^*(\theta)\Lambda_n(\theta)P_n(\theta) + Q_n^*(\theta)\Phi_n(\theta)Q_n(\theta) =: P_n^*\Lambda_n P_n + Q_n^*\Phi_n Q_n$ (compare notation in section 2.2.3)

- normalized Dynamic Principal Components: $\psi_t^n := \underline{\Lambda}_n^{-1/2}(L)\underline{P}_n(L)y_t^n =: \Lambda_n^{-1/2}P_n y_t^n$

- In the following $M \subset [-\pi, \pi]$ is a subset of the frequency band of non-zero measure that will be given in the context

- unitary bandpass filter: Let $M \subset \Theta$, and C be $q \times q$ filter

$$\mathcal{B}_M := \{C : C(\theta) = 0 \ \theta \notin M, C(\theta)C^*(\theta) = I_q\} \tag{2.23}$$

- Let $\Pi \subset \Theta$ be the set of θ's for which $\lambda_q(f_y^n)(\theta) \to \infty$. Recall that by A5, the set $\Theta \setminus \Pi$ is a Lebesgue zero set.

- Rotated version of a normalised principal components: $\psi_t^{m,C} := \underline{C}(L)\psi_t^m$

- If we write $P_m y_t^n$ we mean that $(\underline{P}_m(L), 0 \cdots 0)y_t^n$ filled up with $n - m$ zeros to match the dimension

- We use $\lambda_{n,j} = \lambda_j(f_y^n)$ as a shorthand notation for the j-th spectral eigenvalue.

- $\hat{\psi}_t^{m,n,C} := \text{proj}(\underset{\sim}{C}(L)\psi_t^m \mid \mathbb{H}(\psi_t^n))$. Note that we project here on the whole time domain of (ψ_t^n) rather than on the contemporaneous space.

- $D(m,n,C) =: D = C(\theta)\Lambda^{-1/2}(\theta)P_m(\theta)P_n^*(\theta)\Lambda_n^{-1/2}(\theta)$ is the projection matrix that transforms ψ_t^n to $\hat{\psi}_t^{m,n,C} := \text{proj}(\psi_t^{m,C} \mid \mathbb{H}(\psi^n))$, so $\hat{\psi}_t^{m,n,C} = \underset{\sim}{D}(L)\psi_t^n$ and $\text{spec}\,\hat{\psi}_t^{m,n,C} = \text{spec}\,\underset{\sim}{D}(L)\psi_t^n = D(\theta)D^*(\theta) =: DD^*$. We suppress dependence on n in the notation.

- (α_n) an increasing scalar sequence with $\alpha_n \to \infty$ for $n \to \infty$ such that $\lambda_{n,q}(\theta) := \lambda_q(f_y^n(\theta)) \geq \alpha_n$ for all $\theta \in M$.

From a birds' eye view, the main steps of the proof are the same as in the static case (see steps a) - d) in the beginning of section 2.2.3). Now, in steps **a)-b)**, we construct a basis that is orthonormal and white noise, (v_t), that spans $\mathbb{G}$ in time domain, i.e. $\mathbb{G} = \mathbb{H}(v)$. The main additional complication of this basis construction compared to the static case, is that the first q-eigenvalues might not diverge uniformly. As pointed out by Forni and Lippi (2001), this is for example the case for elementary common components like $\chi_{it} = (1-L)u_t$. Therefore the basis construction for the dynamic aggregation space is conducted "piecewise" on the frequency band. The remaining steps **c)-e)** are essentially analogous to the static case.

Lemma 2.3.7
Suppose $m \leq n$ and $C \in \mathcal{B}_M$ defined in (2.23), then

$$\begin{aligned}
\underset{\sim}{C}(L)\psi_t^m =: C\psi_t^m &= \text{proj}(\psi_t^{m,C} \mid \mathbb{H}(\psi_t^n)) + R_t(m,n) \\
&= \underbrace{C\Lambda^{-1/2}P_m P_n' \Lambda_n^{1/2}}_{D(m,n,C)} \underbrace{\Lambda_n^{-1/2}P_n y_t^n}_{\psi_t^n} + \underbrace{C\Lambda_m^{-1/2}P_m Q_n^* Q_n y_t^n}_{R_t(m,n,C)} .
\end{aligned}$$

Proof. The proof is analogous to lemma 2.2.10, replacing matrices by the corresponding filters. ∎

Note that $D(m,n,C) =: D(\theta) =: D$ is the population projection matrix, when projecting $\psi_t^{m,C}$ onto the time domain of ψ_t^n. On the other hand $R(m,n,C)(\theta) := C\Lambda_m^{-1/2}P_m Q_n^* Q_n =: R(\theta)$ is the filter matrix that transforms y_t^n to the residual of that projection, so $R_t(m,n,C) =: \underline{R}(L)y_t^n =: Ry_t^n$.

Lemma 2.3.8
Suppose A5 holds, then $\lambda_1\left(\text{spec}\left[R_t\left(m,n,C\right)\right]\left(\theta\right)\right) \leq \lambda_{n,q+1}(\theta)/\lambda_{m,q}(\theta)$, *where* $\lambda_{n,j} := \lambda_j(f_y^n(\theta))$ *for all* $\theta \in M$.

Proof. The proof is analogous to the proof of lemma 2.2.11, replacing matrices by the analogous filters and replacing the variance operator "$\mathbb{V}$" by "spec". ■

Lemma 2.3.9 (Squeeze Lemma for eigenvalues of spec ψ_t^n)
For given Lebesgue non-zero $M \subset [-\pi,\pi]$, *there exists* m^*, *such that* $1 \geq \lambda_j(DD^*) \geq 1 - W/\alpha_m > 0$ *for all* $\theta \in M$, $m \geq m^*$ *and* $j = 1, ..., q$.

Proof. The proof is analogous to the proof of lemma 2.2.12, replacing matrices by the analogous filters, operating on the subset M of the frequency band. ■

Note that even though the normalised dynamic principal components $\underline{C}(L)\psi_t^m =: C\psi_t^m$ are orthonormal, the projection $\hat{\psi}_t^{m,n,C}$ is not. In the next step we orthonormalise $\hat{\psi}_t^{m,n,C}$: Let $H(\theta)\Delta(\theta)H^*(\theta) =: H\Delta H^* = DD^*$ be the eigen-decomposition of DD^* (in the usual manner with orthonormal eigenvectors in H), we define an orthonormalised version of $\hat{\psi}_t^{m,n,C}$ as

$$\begin{aligned}\bar{\psi}_t^{m,C} &= \underline{H}(L)\underline{\Delta}^{-1/2}(L)\underline{H}^*(L)\underbrace{\text{proj}(\underline{C}(L)\psi_t^m \mid \mathbb{H}(\psi_t^n))}_{\hat{\psi}_t^{m,n,C}} \\ &= \underbrace{H\Delta^{-1/2}H'D}_{=:F(D)=:F}\psi_t^n = \underline{F}(L)\psi_t^n =: F\psi_t^n\ . \end{aligned} \tag{2.24}$$

Lemma 2.3.10 (The orthonormalised projection)
For all $\tau \in (0,2)$, *there exists* m_τ, *such that*

- $W/\alpha_{m_\tau} < 1$
- *For all* $n \geq m \geq m_\tau$ *we have* $\lambda_1\left(\text{spec}\left[\psi_t^{m,C} - \bar{\psi}_t^{m,n,C}\right]\right) < \tau$ *for all* $\theta \in M$.

Proof. The proof is analogous to the proof of lemma 2.2.13, replacing matrices by filters and replacing "$\mathbb{V}$" by "spec". ■

Next we use the previous lemmas to construct a q-dimensional orthonormal white noise "dynamic" basis for $\mathbb{G}$, i.e. $\mathbb{H}(v) = \mathbb{G}$. We begin with constructing a process (v_t^M) that is orthonormal on M and lives in $\mathbb{G}$

Lemma 2.3.11 (Existence of a band-white noise process in the Aggregation Space)
Suppose A1 and A5 hold and $M \subset [-\pi, \pi]$ as above, then there exists a q-dimensional process (v_t^M), such that

1. $v_{jt}^M \in \mathbb{G}$ *for* $j = 1, ..., q$
2. $\operatorname{spec} v_t^M = \begin{cases} I_q \ \textit{for} \ \ \theta \in M \ , \\ 0 \ \textit{else} \ . \end{cases}$

Proof. Analogous to the proof of proposition 2.2.15, replacing matrices by the corresponding filters and "$\mathbb{V}$" by "spec", we obtain a Cauchy sequence $(v_t^{M,k})$, with limit (v_t^M) having the required spectrum and living in $\mathbb{H}(y)$.
We are left with showing that the limit process (v_t^M) lives also in $\mathbb{G}$. For this we show that the sequence of filters $\underline{G}_k(\theta) := \underline{F}_k(\theta)\underline{\Lambda}_{s_k}^{-1/2}(\theta)\underline{P}_{s_k}(\theta) =: F_k\Lambda_{s_k}^{-1/2}P_{s_k}$ defines a DAS. For this note that

$$G_kG_k^* = F_k\lambda_{s_k}^{-1}F_k^* \text{ so } \quad \lambda_1(G_kG_k^*) \leq \frac{1}{\lambda_q(f_y^{s_k})} \to 0 \text{ a.e. on } \Theta \ ,$$

consequently considering the row vectors, say $g_{j,k}$ with $j = 1, ..., q$ from $G_k = \begin{pmatrix} g_{1k} \\ \vdots \\ g_{qk} \end{pmatrix}$ we know that $g_{jk}g_{jk}^* < |1/\lambda_q(f_y^{s_k})|$ which is integrable and tends to zero a.e. in $M \subset \Theta$. By the theorem of dominated convergence, it follows that $\lim_k \int_\Theta g_{jk}g_{jk}^* = \int_\Theta \lim_k g_{jk}g_{jk}^* = 0$. Since $\operatorname{spec} v_t^{M,k} = I_q$ for all $\theta \in M$ and $\operatorname{spec} v_t^{M,k} = 0$ else, we have that $\operatorname{spec} v_t^M = \operatorname{spec}(\underline{\lim}_k v_t^{M,k}) = \lim \operatorname{spec}(v_t^{M,k}) = I_q$ for $\theta \in M$ and $\operatorname{spec} v_t^M = 0$ else by lemma A.0.3 which completes the proof. ■

Proposition 2.3.12 (There exists an orthonormal white noise process in $\mathbb{G}$) *Suppose A1, A5 hold, then there exists an orthonormal white noise process* $(v_t) \sim WN(I_q)$ *in* $\mathbb{G}$.

Proof. As usual set $\lambda_{j,q}(\theta) =: \lambda_{j,q} = \lambda_q(f_y^j(\theta))$
We construct a sequence of sets $(N_b : b \in \mathbb{N})$ as follows. At first set $b = 1$.

Algorithm $a)$: We construct a sequence of sets $M_a^1 \subset \Pi$ together with a sequence of integers ν_a^1 while we set $M_0^1 = \Pi$, $\pi_1 = \pi$
($a = 1$, determine ν_1^1): Let ν_1^1 be the smallest natural number, such that $\mathcal{L}(\{\theta \in M_0 : \lambda_{\nu_1^1,q} > 1\}) > \pi_1$. Such ν_1^1 must exist since $\lambda_{n,q}$ diverges at every $\theta \in \Pi$, set

$$M_1^1 := \{\theta \in M_0 : \lambda_{v_1^1,1} > 1\} \text{ and } \mathcal{L}(M_1^1) > \pi_1 = \pi \ .$$

($a = 2$, determine ν_2^1): Let ν_2^1 be the smallest natural number, such that $\mathcal{L}(\{\theta \in M_1^1 : \lambda_{\nu_2^1,q} > 2\}) > \pi_1$, set

$$M_2^1 := \{\theta \in M_1^1 : \lambda_{\nu_2^1,q} > 2\} \text{ and } \mathcal{L}(M_2^1) > \pi_1 \ .$$

(For any a, determine ν_a^1): Let ν_a^1 be the smallest natural number, such that $\mathcal{L}(\{\theta \in M_{a-1}^1 : \lambda_{\nu_a^1,q} > a\})$, set

$$M_a^1 := \{\theta \in M_{a-1}^1 : \lambda_{\nu_a^1,q} > a\} \text{ and } \mathcal{L}(M_a^1) > \pi_1 \ .$$

It follows that for

$$N_1 := M_1^1 \cap M_2^1 \cap \cdots = \bigcap_{a=2}^{\infty} M_a^1 \text{ we have } \mathcal{L}(N_1) \geq \pi_1 = \pi \ .$$

Algorithm $b)$: We construct the sequence (N_b) of subsets of Π using (M_a^b) and ν_a^b:
($b = 1$): Set $N_1 := \bigcap_{a=1}^{\infty} M_a^1$ using algorithm $a)$ with $\pi_1 := \pi$ we know that $\mathcal{L}(N_1) \geq \pi_1 = \pi$.
($b = 2$): Set $M_0^2 := \Pi \setminus N_1$, $\pi_2 := \mathcal{L}(\Pi \setminus N_1)/2$ and apply algorithm $a)$ starting with M_0^2 to obtain the sequence of sets $(M_a^2 : a \in \mathbb{N})$ together with the index

sequence $(\nu_{a^2} : a \in \mathbb{N})$ while

$$M_a^2 := \{\theta \in M_{a-1}^2 : \lambda_{\nu_a^2, q} > a\} \text{ and } \mathcal{L}(M_a^2) > \pi_2 \ .$$

setting $N_2 := \bigcap_{a=1}^{\infty} M_a^2$, we know that $\mathcal{L}(N_2) \geq \pi_2$.
$(b = 3)$: Set $M_0^2 := \Pi \setminus (N_1 \cap N_2)$, $\pi_3 := \mathcal{L}(\Pi \setminus (N_1 \cap N_2))$, algorithm $a)$ yields N_3.
$(b = 4)$: ...
Consequently $(N_b : b \in \mathbb{N})$ is a sequence of disjoint subsets of Π such that

$$\sum_{b=1}^{\infty} \pi_b \leq \mathcal{L}\left(\bigcup_{b=1}^{\infty} N_b\right) = \sum_{b=1}^{\infty} \mathcal{L}(N_b) = 2\pi \ .$$

We apply lemma 2.3.11 with $M = N_b$ to obtain processes $(v_t^{N_1}), (v_t^{N_2}), \ldots$ which oscillate on disjoint sets of the frequency band, but $v_t := \sum_{b=1}^{\infty} v_t^{N_b}$ is white noise and lives in $\mathbb{G}$, which completes the proof. ■

So far we have shown the existence of a white noise process (v_t) in $\mathbb{G}$. Next we prove that such a (v_t) is a "dynamic basis" for $\mathbb{G}$:

Proposition 2.3.13 ((v_t) is a dynamic basis for $\mathbb{G}$)
Given (v_t) from proposition 2.3.12, it holds that $\mathbb{H}(v) = \mathbb{G}$.

Proof. The proof works analogous to the proof of proposition 2.2.15 replacing matrices by the corresponding filters, replacing "$\mathbb{V}$" by "spec" and working with the projection onto $\mathbb{H}(v)$ instead of projecting onto $\text{sp}(V_t)$. ■

Definition 2.3.14 (Cauchy sequence of dynamic spaces generating process)
Suppose $(w_t^n) \sim WN(I_q)$ is orthonormal white noise, lives in $\mathbb{H}(y)$ and is co-stationary with (y_{it}) for every $n \in \mathbb{N}$. We say that (w_t^n) generates a Cauchy sequence of dynamic spaces if

$$\text{tr}\,\text{spec}\left[w_t^m - \text{proj}(w_t^m \mid \text{sp}(w_t^n))\right] \to 0 \quad \textit{for } m, n \to \infty \ \textit{a.e. on } \Theta \ .$$

Lemma 2.3.15 (The projection onto a Cauchy sequence of dynamic spaces is convergent)

Let (w_t^n) be a Cauchy sequence of dynamic spaces generating processes as in definition 2.3.14. Suppose (z_t) is scalar stationary and lives in $\mathbb{H}(y)$ - co-stationary with (y_{it}), then $\operatorname{proj}(z_t \mid \mathbb{H}(w_t^n))$ converges in $\mathbb{H}(y)$.

Proof. The proof works analogous to the proof of proposition 2.2.17 replacing matrices by the corresponding filters, replacing "$\mathbb{V}$" by "spec", covariances by cross-spectrum and working pointwise on the frequency band with the projection onto $\mathbb{H}(w^n)$ instead of projecting onto $\operatorname{sp}(W_t^n)$.
For the last part of the proof, consider $\hat{z}_t^n := \operatorname{proj}(z_t \mid \mathbb{H}(w^n))$, we have that

$$\operatorname{spec}(\hat{z}_t^m - \hat{z}_t^n) \leq \operatorname{spec}(z_t) < \infty$$

by the theorem of dominated convergence, we have

$$0 = \int_\Theta \lim \operatorname{spec}(\hat{z}_t^m - \hat{z}_t^n) = \lim \int_\Theta \operatorname{spec}(\hat{z}_t^m - \hat{z}_t^n) \text{ where } m, n \to \infty \ ,$$

which implies that $\hat{z}_t^n$ is a Cauchy sequence and therefore convergent. ■

Lemma 2.3.16

The sequence (ψ_t^n) generates a Cauchy sequence of dynamic spaces.

Proof. The proof works analogous to the proof of proposition 2.2.17 replacing matrices by the corresponding filters, replacing "$\mathbb{V}$" by "spec", covariances by cross-spectrum and working pointwise on the frequency band. ■

Lemma 2.3.17

The sequence $\underline{\mathcal{K}}_{ni}(\theta)$, is a dynamic averaging sequence.

Proof. Consider

$$\operatorname{proj}(y_{it} \mid \mathbb{H}(\psi_t^n)) = \pi_{ni}\Lambda_n^{1/2} \underbrace{\Lambda_n^{-1/2} P_n y_t^n}_{\psi_t^n}$$

$$\pi_{ni}(\theta)\pi_{ni}^*(\theta)\lambda_{nq}(\theta) =: \pi_{ni}\pi_{ni}^*\lambda_{nq} \leq \operatorname{spec} y_{it} < \infty \ ,$$

$$\text{so} \quad \mathcal{K}_{ni}\mathcal{K}_{ni}^* = \pi_{ni}\pi_{ni}^* \leq \frac{\operatorname{spec} y_{it}}{\lambda_{nq}} \to 0 \text{ a.e. on } \Theta$$

$$\text{so} \quad 0 = \int_\Theta \lim_n \pi_{ni}\pi_{ni}^* = \lim_n \int_\Theta \pi_{ni}\pi_{ni}^* \ ,$$

where the last equality follows by the theorem of dominated convergence. ■

Proposition 2.3.18
If (y_{it}) is a q-DFS, then the i-th element in the DLRA of rank r converges in L^2 to the projection of y_{it} on the dynamic aggregation space, i.e.

$$\underline{\lim}_n \operatorname{proj}(y_{it} \mid \mathbb{H}(\psi_t^n)) = \underline{\lim}_n \underline{\mathcal{K}}_{ni}(L) y_t^n = \underline{\lim}_n \gamma_{it}^n = \operatorname{proj}(y_{it} \mid \mathbb{G}) \ .$$

Proof. The proof is analogous to the proof of proposition 2.2.20. ■

Lemma 2.3.19
The double sequence of residuals (δ_{it}) from the orthogonal projection of the output on the dynamic aggregation space, i.e. $\gamma_{it} = \operatorname{proj}(y_{it} \mid \mathbb{G})$ is dynamically idiosyncratic.

Proof. The proof is analogous to the proof of proposition 2.2.21. ■

Proposition 2.3.20
Suppose that (y_{it}) has a factor model representation as in (2.22) satisfying the assumptions from theorem 2.3.5.1, then
1. $\mathbb{H}(\chi) = \mathbb{H}(u) = \mathbb{G}$
2. $\chi_{it} = \operatorname{proj}(\chi_{it} \mid \mathbb{G})$.

Proof. Suppose (y_{it}) has a representation as a factor model as in equation (2.22), so

$$y_{it} = \chi_{it} + \xi_{it} = \underline{b}_i(L) u_t + \xi_{it}$$

while $\sup_n \lambda_q(f_\chi^n) = \infty$ a.e. on Θ and $\sup_n \lambda_1(f_\xi^n) < \infty$.
Furthermore, we know that

$$\begin{aligned} y_{it} &= \operatorname{proj}(y_{it} \mid \mathbb{G}) + \delta_{it} = \gamma_{it} + \delta_{it} \\ &= \underbrace{\underline{c}_i(L)}_{1\times q} \underbrace{v_t}_{q\times 1} + \delta_{it} \text{ from proposition 2.3.13} \ . \end{aligned}$$

Since (ξ_{it}) is idiosyncratic, we know that $\mathbb{G} \subset \mathbb{H}(\chi)$, because for any $z_t = \underline{\lim}_k\, \underline{c}^{(k)}(L)y_t \in \mathbb{G}$, where $(c^{(k)})$ is the DAS corresponding to z_t, we have

$$z_t = \underline{\lim}_k\, \underline{c}^{(k)}(L)y_t = \underline{\lim}_k\, \underline{c}^{(k)}(L)(\chi_t + \xi_t) = \underline{\lim}_k\, \underline{c}^{(k)}(L)\chi_t \ .$$

But also $\mathbb{H}(\chi) \subset (u_t)$ since $\chi_{it} = \underline{b}_i(L)u_t$. It follows that

$$\mathbb{H}(v) = \mathbb{G} \subset \mathbb{H}(\chi) \subset \mathbb{H}(u).$$

On the other hand, we know that $\operatorname{spec} u_t = I_q = \operatorname{spec} v_t$ a.e. on Θ, so noting that there exists an $q \times q$ filter $\underline{d}(L)$, we can write $v_t = \underbrace{d}_{q\times q} u_t$ while $dd^* = I_q$ a.e. on Θ, which implies that d is invertible and consequently $u_t = \underline{d}^{-1}(L)v_t$, so from

$$\operatorname{sp}(u_t) \subset \operatorname{sp}(v_t) = \mathbb{G}$$
$$\text{it follows that } \mathbb{G} = \mathbb{H}(u) = \mathbb{H}(\chi)$$

and $\chi_{it} \in \mathbb{G}$ and $\xi_{it} \perp \mathbb{G} = \mathbb{H}(\chi)$ for all i, t which implies that $\chi_{it} = \operatorname{proj}(y_{it} \mid \mathbb{G})$ and $\delta_{it} = \delta_{it}$ by the projection theorem. ■

Chapter 3

Reconciling the Schools in One Model

Given the discussion of the previous chapter, naturally the question arises whether every static factor sequence is a dynamic factor sequence. However, the "Generalised Dynamic Factor Model" (A5), is actually not more general than the class of static factor sequences (A2): For example, we might have divergence only on a subset of the frequency band:

Example 3.0.1 (1-SFS but not a q-DFS)
Let (u_t) *be scalar white noise and* (ε_{it}^{ξ}) *be i.i.d. across all leads and lags and independent of* (u_t)*. Consider a double sequence with common component given by*

$$\chi_{it} = F_t = 2\underline{I}_{[-\pi/2,\pi/2]}(L)u_t \ ,$$

where $I_{[-\pi/2,\pi/2]}$ *is the index function being equal to one for* $\theta \in [-\pi/2, \pi/2]$ *and zero otherwise. The spectrum of* (χ_{it}) *is*

$$\begin{pmatrix} 1 & 1 & \cdots \\ 1 & 1 & \\ \vdots & & \ddots \end{pmatrix}$$

on $[-\pi/2, \pi/2]$ *and zero on* $[-\pi, \pi] \setminus [-\pi/2, \pi/2]$*, so* $\sup_n \lambda_1(\Gamma_\chi^n) = \infty$*. This is a* 1*-SFS but not a DFS since there divergence does not happen a.e. in* $[-\pi, \pi]$*.*

On the other hand A5 is also violated if all eigenvalues of f_y^n diverge:

Example 3.0.2 (Infinite Number of Blocks: ∞-DFS)
Let (u_t^j) *be a collection of mutually orthogonal scalar white noise processes: So for each* j*,* $(u_t^j : t \in \mathbb{Z})$ *is white noise and for all* $j \neq l$ *we have* $u_t^j \perp u_s^l$ *for all* $t, s \in \mathbb{Z}$*. Define filters* $\underline{b}_i^{(j)}(L)$ *for* $i, j \in \mathbb{N}$ *such that* $\chi_t^{(j)} = \underline{b}_i^{(j)}(L) u_t^j$ *is the common component of a* 1*-DFS for* $i \in \mathbb{N}$*. Hence the collection* $\{(\chi_{it}^{(j)}) : j \in \mathbb{N}\}$ *is an infinite set of mutually orthogonal dynamic factor sequences.*
Next we use these factor sequences and blend them to a new double sequence in the following manner:

$$(\tilde{\chi}_{lt} : l \in \mathbb{N}, t \in \mathbb{Z}) := (\chi_{1t}^{(1)} \mid \chi_{2t}^{(1)}, \chi_{1t}^{(2)} \mid \chi_{3t}^{(1)}, \chi_{2t}^{(2)}, \chi_{1t}^{(3)} \mid \cdots) \,. \tag{3.1}$$

In Hallin and Liška (2011), it is shown that if two dynamic factor sequences with dynamic dimensions q_1, q_2 *are combined to one factor sequence, the resulting double sequence is a again a factor sequence of dimension* q *with* $\max(q_1, q_2) \leq q \leq q_1 + q_2$*. In particular the dimension* q *of the blended factor sequence is* $q_1 + q_2 - q_\cap$*, where* $q_\cap$ *is the dimension of the intersection of the two aggregation spaces.*
Since by construction of the model above, the dimension of the interesction of the aggregation spaces of two factor sequences $(\chi_{it}^{(j)})$ *and* $(\chi_{it}^{(l)})$ *is zero, as the dynamic shocks are mutually orthogonal. Set* $\operatorname{spec} \tilde{\chi}_t^n =: f_{\tilde{\chi}}^n$*. By iteratively applying this result, it follows that there exists no finite* q*, such that* $\sup_n \lambda_{q+1}(f_{\tilde{\chi}}^n)$ *is essentially bounded, where* $f_{\tilde{\chi}}^n$ *is the* n*-dimensional spectrum corresponding to* $(\tilde{\chi}_{lt})$ *from equation (3.1).*

In this chapter, we reconcile the two schools in one model: We suppose that (y_{it}) is weakly stationary (A1), and has both, a static (A2) and a dynamic factor structure (A5). In section 3.1, we provide a new decomposition of y_{it} which has been overlooked in the literature and which is the moot point of this thesis: We can decompose y_{it} into the static common component, the weak common component and the dynamic idiosyncratic component. The weak common component lives in the dynamic aggregation space, vanishes under static aggregation and is spanned by a potentially infinite number of weak factors. We interpret this result in detail from a structural point of view. In section 3.2, we provide a canonical representation for double sequences which have a dynamic and a static factor structure and model the dynamic common component as a state space system. Finally, in section 3.3, we show that under quite general and common conditions, there exists a representation of the dynamic common component that is causally subordinated to the output.

3.1 A New Decomposition Theorem

The main statement of this thesis emerges from a combination of theorem 2.3.5 by Forni and Lippi (2001) and the reformulation of the results from Chamberlain and Rothschild (1983a) in theorem 2.2.7.

Theorem 3.1.1

1. For every $t \in \mathbb{Z}$ the static aggregation space is contained in the dynamic aggregation space, i.e. $\mathbb{S}_t \subset \mathbb{G}$.

2. If A1, A5 and A2 hold, then $C_{it} = \text{proj}(\chi_{it} \mid \mathbb{S}_t)$. In particular we can decompose (y_{it}) into three parts:

$$y_{it} = C_{it} + e^{\chi}_{it} + \xi_{it}, \tag{3.2}$$

where $\chi_{it} = C_{it} + e^{\chi}_{it}$ in equation (2.22) and $e_{it} = e^{\chi}_{it} + \xi_{it}$ in equation (2.10), with $e^{\chi}_{it}, C_{it} \perp \xi_{js}$ for all i, j, s, t and $C_{it} \perp e^{\chi}_{jt}$ for all j, i. Furthermore (e^{χ}_{it}) is statically idiosyncratic while $(\chi_{it}), (\xi_{it}), (C_{it})$ and (e_{it}) satisfy the conditions of theorems 2.3.5 and 2.2.7.

Proof. Clearly, under the given assumptions $\hat{L}^{\infty}_2(\Gamma_y) \subset L^{\infty}_2(f_y)$ and $\hat{L}^{\infty}_2(I) \subset L^{\infty}_2(I)$, so every static averaging sequence is a dynamic averaging sequence, i.e. $\mathcal{S}(\Gamma_y) \subset \mathcal{D}(f_y)$. It follows that every static aggregate is a dynamic aggregate and therefore $\mathbb{S}_t \subset \mathbb{G}$ for all $t \in \mathbb{Z}$.

For the second statement note that we have

$$
\begin{aligned}
C_{it} &= \text{proj}\,(y_{it} \mid \mathbb{S}_t) && \text{by theorem 2.2.7.4} \\
&= \text{proj}\,(\chi_{it} + \xi_{it} \mid \mathbb{S}_t) && \text{by theorem 2.3.5.1} \\
&= \text{proj}\,(\chi_{it} \mid \mathbb{S}_t) + \text{proj}\,(\xi_{it} \mid \mathbb{S}_t) \\
&= \text{proj}\,(\chi_{it} \mid \mathbb{S}_t) && \text{since } \mathbb{S}_t \subset \mathbb{G}.
\end{aligned}
$$

Since $e^{\chi}_{it} = \chi_{it} - C_{it} \in \mathbb{G}$ it follows that $e^{\chi}_{it} \perp \xi_{js}$ for all i, j, s, t. Furthermore $e^{\chi}_{it} = e_{it} - \xi_{it}$, and both terms on the right hand side vanish under static aggregation, so does e^{χ}_{it}. ∎

Definition 3.1.2 (Weak Common Component)
In (3.2), we call (e^{χ}_{it}) *the weak common component.*

Reconciling the American and the Italian school, we can state that the approaches can be regarded as mathematically analogous (see chapter 2) while employing two different types of aggregation being *static-* (obtained via SLRA) versus *dynamic* aggregation (obtained via DLRA). The two schools have *structurally* two different types of common components - a dynamic and a static one. The dynamic common component arises from a projection onto a (much) larger Hilbert space $\mathbb{G}$. Note that $\mathbb{G}$ contains *all* static aggregation spaces, i.e. the union $\overline{\text{sp}}\left(\bigcup_{t\in\mathbb{Z}} \mathbb{S}_t\right) = \mathbb{H}(C)$ which is the whole time domain of the static common component. So naturally, the dynamic common component, in general, explains a larger part of the variation of the outputs (y_{it}). This however does not imply that static aggregation is in any sense "worse" than dynamic aggregation (as we will examine below). The moot point of this thesis, and the discussion that follows, is that a careful distinction between the two concepts has theoretical and empirical relevance and implies a number of interesting research questions.
The weak common component is the residual term from the projection of the dynamic common component on the static aggregation space. As such, it is the static idiosyncratic component of the dynamic common component and lives in the dynamic aggregation space. On the other hand it is also the projection of the static idiosyncratic component on the dynamic aggregation space, i.e. $e^{\chi}_{it} = \text{proj}(e_{it} \mid \mathbb{G})$

or that part of the static idiosyncratic component which is dynamically common to the output sequence. It vanishes under static aggregation and is spanned by (a potentially infinite number) of *weak static factors* (see definition 3.2.1 below). The weak common component - though always being statically idiosyncratic - can be dynamically idiosyncratic or not: In an extreme case, the static common component can be even zero while (e_{it}^{χ}) is not equal to zero for every cross-sectional unit: (see a similar example in a different narrative also in Hallin and Lippi, 2013):

Example 3.1.3 (1-DFS but 0-SFS)
Let (u_t) be a scalar white noise process with unit variance. Consider a dynamic common component of the form

$$\chi_{it} = u_{t-i+1} \ .$$

he spectrum of (χ_t) is

$$f_{\chi}(\theta) = \begin{pmatrix} 1 & e^{\iota\theta} & e^{2\iota\theta} & \cdots \\ e^{-\iota\theta} & 1 & e^{\iota\theta} & \cdots \\ e^{-2\iota\theta} & & 1 & \\ \vdots & & & \ddots \end{pmatrix} .$$

Firstly note that $\Gamma_{\chi}^{n} = I_n$ for all n, so the first eigenvalue of Γ_{χ}^{n} is bounded and $C_{it} = 0$ by theorem 2.2.6 and $\mathbb{S}_t = \{0\}$.
The first row of f_{χ}^{n} equals the k-th row of f_{χ}^{n} times $e^{\iota k\theta}$. Thus f_{χ}^{n} has rank one a.e. on Θ and therefore $\lambda_1(f_{\chi}^{n}(\theta)) = \operatorname{tr} f_{\chi}^{n}(\theta) = n \to \infty$. It follows that (χ_{it}) is a 1-DFS by theorem 2.3.5.
Relating to equation (3.2), we have $e_{it}^{\chi} = \chi_{it}$ and by the special construction of this double sequence, we have that $\chi_{2,t+1} = u_{t+1-2+1} = u_t = \chi_{1t}$ and $\chi_{3,t+1} = u_{t+1-3+1} = u_{t-1} = \chi_{2t}$ and so on. Here we can perfectly predict $\chi_{i,t+1}$ for $i \geq 2$ through χ_{it}, that means that all the predictive power is due to the term (e_{it}^{χ}) which would be lost under static aggregation.

Admitted, this is example is really pathological, though illuminating, as it demonstrates the range of possibilities when distinguishing between dynamic and static aggregation. Note also that in this example $1 = q > r = 0$, so in general it *does not* hold that $q \leq r$ as is commonly assumed. In other words the number of dynamic shocks is in general not less or equal than the number of strong static factors.

We know that $(e^{\chi}_{it} : t \in \mathbb{Z})$ lives in $\mathbb{G}$, but this does not imply that the double sequence as a whole (e^{χ}_{it}) has a non-trivial aggregation space, i.e. is *not* dynamically idiosyncratic.

Example 3.1.4

Consider a double sequence where the dynamic common component is given by

$$\chi_{1t} = u_t$$
$$\chi_{it} = u_{t-1} \ \mathit{for} \ \ i > 1 \ ,$$

where (u_t) *is as in example 3.1.3. Here* $C_{it} = u_{t-1}$ *for all* $i > 1$ *and* $e^{\chi}_{it} = u_t$ *for* $i = 1$ *and* $e^{\chi}_{it} = 0$ *for* $i > 1$*. So* (e^{χ}_{it}) *itself is dynamically and statically idiosyncratic.*

Also here, we can perfectly predict $\chi_{i,t+1}$ *for* $i > 1$ *from* χ_{1t}*.*

3.2 A Canonical Representation

Next we would like to construct a "canonical representation" of y^n_t for finite n in terms of strong and weak factors. Again suppose A1, A5 and A2 hold. Denote now by (F^s_t) the $r \times 1$ dimensional stochastic vector of "strong" factors obtained from static aggregation (corresponding formerly to (F_t) in theorem 2.2.7). By theorem 3.1.1, we know that $F^s_t \in \overline{\mathrm{sp}}(\chi_t)$ and $\mathrm{sp}(F^s_t) = \mathbb{S}_t(y)$. We use the Gram-Schmidt-orthogonalisation procedure to iteratively add weak factor basis dimensions to $\overline{\mathrm{sp}}(\chi_t)$:

Choose the first i in order for which $\chi_{it} - \mathrm{proj}(\chi_{it} \mid \mathrm{sp}(F^s_t)) \neq 0$, set this to i_1. Set $v_{1t} = \chi_{i_1,t} - \mathrm{proj}(\chi_{i_1,t} \mid \mathrm{sp}(F^s_t))$ and set $F^w_{1t} = \|v_{1t}\|^{-1} v_{1t}$. Let $i_2 > i_1$ be the next i in order such that $\chi_{it} - \mathrm{proj}(\chi_{it} \mid \mathrm{sp}(F^s_t, F^w_{1t})) \neq 0$ and set $v_{2t} = \chi_{i_2,t} - \mathrm{proj}(\chi_{i_2,t} \mid \mathrm{sp}(F^s_t, F^w_{1t}))$ and $F^w_{2t} = \|v_{2t}\|^{-1} v_{2t}$. This way we obtain indices $i_1, i_2, ..., i_{r^+_\chi(n)}$ with

$r_\chi(n)^+ \leq n$ along with $F_t^{w,n} = (F_{1t}^w, \dots, F_{r_\chi^+(n),t}^w)'$ having orthonormal variance matrix and being contemporaneously orthogonal to F_t^s. Set $r_\chi(n) := r + r_\chi^+(n)$. The static factors $F_t^n = (F_t^s, F_t^{w,n})'$ are $r_\chi(n) \times 1$. For every finite n, we can write the decomposition of theorem 3.1.1 in vector form as

$$
\begin{aligned}
y_t^n &= C_t^n + e_t^{\chi,n} + \xi_t^n \\
&= \underbrace{\Lambda_s^n F_t^s + \Lambda_w^n F_t^{w,n}}_{\chi_t^n} + \xi_t^n = \begin{bmatrix} \Lambda_s^n & \Lambda_w^n \end{bmatrix} \begin{bmatrix} F_t^s \\ F_t^{w,n} \end{bmatrix} + \xi_t^n \ ,
\end{aligned} \tag{3.3}
$$

with $\mathbb{E}\, F_t^n (F_t^n)' = I_{r_\chi(n)}$ by construction. Furthermore theorems 2.2.7 and 3.1.1 imply that $\lambda_r \left((\Lambda_s^n)' \Lambda_s^n \right) \to \infty$ and $\sup_n \lambda_1 \left((\Lambda_w^n)' \Lambda_w^n \right) < \infty$. In general, it is clear that the dimension $r_\chi^+(n)$ of $F_t^{w,n}$ may increase, when we add new variables in equation (3.3).

Definition 3.2.1 (Strong and Weak Static Factors)
We shall use the term static factor for any basis coordinate of $\overline{\mathrm{sp}}(\chi_t)$ *- distinguishing as in equation (3.3) between strong static factors,* (F_t^s) *associated with strong loadings, and weak static factors,* $(F_t^{w,n})$ *associated with weak loadings.*

The term “weak factor” has first been used by Onatski (2012) which gives a notion of weak factors that is consistent with what we defined above. In particular Onatski (2012) considers the model $y_t^n = \Lambda_w^n F_t^w + e_t^n$ with weak loadings $\sup_n \lambda_1 \left((\Lambda_w^n)' \Lambda_w^n \right) < \infty$ and shows that the principal components estimator is not consistent.

Weak factors may load e.g. only on a finite number of cross-sectional units - though their influence might be large for those units - or their loadings are “thinly” distributed in the cross-section with vanishing influence or both. On the other hand their influence on the subsequent period might be large and consequently, they can be important for forecasting (see section 4.3). We shall use the term *dynamic factor* for a shock that is a dynamic basis coordinate for $\mathbb{G}$ as in theorem 2.3.5. Naturally in this framework, there is no such a thing as a weak dynamic factor.

In light of the discussion above, the term “factor space” is to be used with care as we have to distinguish between $\mathbb{S}_t = \text{sp}(F_t^s)$, $\overline{\text{sp}}(\chi_t)$ and $\mathbb{G}$. For example Forni et al. (2015, 2017) provide theory and methods for "infinite dimensional factor spaces" by which they actually mean that $\overline{\text{sp}}(\chi_t)$ is infinite dimensional. However, $\overline{\text{sp}}(\chi_t)$ might be infinite dimensional even though $\mathbb{S}_t$ is finite dimensional (see example 4.4.2).

We may suppose that $\dim \overline{\text{sp}}(\chi_t)$ is finite dimensional:

Assumption 6

The contemporaneous space of the dynamic common component is finite dimensional, i.e. $\sup_n r_\chi(n) =: r_\chi = \dim \overline{\text{sp}}(\chi_t) < \infty$.

This guarantees that we have a SFS. The static and dynamic common component coincide if and only if all r_χ non-zero eigenvalues of the variance of the dynamic common component diverge. Some structural results in order:

Theorem 3.2.2

Suppose (y_{it}) *is a q-DFS (A1, A5) .*

1. If $\sup_n r_\chi(n) = r_\chi < \infty$, *then* (y_{it}) *is a* r*-SFS (A2) with* $r \leq r_\chi$.

2. If (y_{it}) *is also a* r*-SFS (A2), then* $C_{it} = \chi_{it}$, *for all* $i \in \mathbb{N}, t \in \mathbb{Z}$ *if and only if there exists an* r*-dimensional process* z_t *with non-singular variance matrix* $\Gamma_z = \mathbb{E}\, z_t z_t'$ *together with a nested sequence of* $n \times r$ *loadings-matrices* L^n *such that* $\chi_t^n = L^n z_t$ *and* $\lambda_r(L^{n\prime} L^n) \to \infty$.

3. Suppose (y_{it}) *is also a* r*-SFS (A2) and consider a representation of the dynamic common component of the form*

$$\chi_t^n = \begin{bmatrix} L_1^n & L_2^n \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^{2,n} \end{bmatrix} ,$$

where x_t^1 *is* $(r \times 1)$ *and* $\sup_n \lambda_r \left((L_1^n)' L_1^n \right) = \infty$ *and* $\sup_n \lambda_1 \left((L_2^n)' L_2^n \right) < \infty$, $\Gamma_{x^1} = \mathbb{V}\, x_t^1$ *and* $\Gamma_{x^2}^n = \mathbb{V}\, x_t^{2,n}$ *are non-singular and* $\mathbb{E}\, x_t^1 (x_t^{2,n})' = 0$ *for all* $n \in \mathbb{N}$. *Then* $C_t^n = L_1^n x_t^1$ *and* $e_t^{\chi,n} = L_2^n x_t^{2,n}$ *for all* $n \in \mathbb{N}$.

Proof. If $\dim \mathbb{H}(\chi_t) = r_\chi$, then there exists some r_χ dimensional process, together with a loadings L^n such that $\chi_t^n = L^n z_t$ for all $t \in \mathbb{Z}$. By theorem 3.1.1 we know that $\mathbb{S}_t(y) \subset \mathbb{H}(\chi_t)$, therefore there exists F_t from theorem 2.2.7 as a linear transformation of z_t of dimension $r \leq r_\chi$.

For the second statement, note that if $C_{it} = \chi_{it}$ for all $i \in \mathbb{N}, t \in \mathbb{Z}$, we know that χ_{it} is finite dimensional with $r_\chi = r$ being the number of divergent eigenvalues in Γ_C^n by theorem 2.2.7.3. Setting $z_t = F_t$ (theorem 2.2.7.1), we obtain the desired result.

On the other hand, let now $\chi_t^n = L^n z_t$, with $\lambda_r(L^{n\prime} L^n) \to \infty$ with $\Gamma_z = \mathbb{V}\, z_t = PDP'$ having full rank and D is a diagonal matrix of eigenvalues and P is an orthonormal matrix of eigenvectors. Since (y_{it}) is a SFS by assumption A2, there exists some $\tilde{r} \leq r$ such that $\lambda_{\tilde{r}}(\Gamma_\chi^n) \to \infty$ and $\lambda_{\tilde{r}+1}(\Gamma_\chi^n) < \infty$ by theorem 3.1.1 and theorem 2.2.7.1. Suppose $\tilde{r} < r$, so $\lambda_r(\Gamma_\chi) = \lambda_r(L^n \Gamma_z L^{n\prime}) < \infty$. But

$$\lambda_r(L^n \Gamma_z L^{n\prime}) = \lambda_r(L^n PDP' L^{n\prime}) = \lambda_r(D^{1/2} P' L^{n\prime} L^n P D^{1/2}) \tag{3.4}$$

$$\geq \lambda_1(D^{1/2} P') \lambda_r(L^{n\prime} L^n) \lambda_1(P D^{1/2}) \; , \tag{3.5}$$

which is a contradiction as $\lambda_1(PD^{1/2}) > 0$ since Γ_z is of full rank and $\lambda_r(L^{n\prime} L^n) \to \infty$.

Set $\chi_{it}^1 = L_{i,1} x_t^1$ and $\chi_{it}^2 = \chi_{it} - \chi_{it}^1$. Recall that $\mathbb{S}_t(y) = \mathbb{S}_t(\chi)$ by theorem 3.1.1. By theorem 2.2.7.3 the decomposition into static common and static idiosyncratic component of (χ_{it}) is unique - given the number r of divergent eigenvalues and the contemporaneous orthogonality between χ_{it}^1 and χ_{it}^2 is satisfied. This completes the proof. ∎

3.3 Causal Subordination

As we have seen above, the dynamic common component retains a larger part of the variation of the output process by accounting for the presence of weak factors which is a desirable feature. However, a theoretical downside at first glance is that dynamic principal components is in general a two-sided operation on the observed data. As a consequence neither the dynamic common component nor the dynamic shocks (u_t) as they arise from computation via DLRA are causally subordinated to (y_{it}), and therefore cannot be used for forecasting. In this section, we show that under quite general conditions there exists a representation of (χ_{it}) and of the innovations of (χ_{it}) causally subordinated to the output (y_{it}).

Suppose (z_t) is a multivariate $d \times 1$ zero-mean weakly stationary process. Let $\varepsilon_t := z_t - \text{proj}(z_t \mid \mathbb{H}_{t-1}(z))$ be the innovation of z_t. Recall that by the Wold representation theorem (see e.g. Deistler and Scherrer, 2022, theorem 2.3) we may say that (z_t) is purely non-deterministic iff we can represent $z_t = \sum_{j=0}^{\infty} K(j)\varepsilon_{t-j}$, where $K(0) = I_d$ and $\sum_{j=0}^{\infty} \|K(j)\|^2 < \infty$. We call $k(\theta) = \sum_{j=0}^{\infty} K(j)e^{-i\theta j}$ transfer function of (z_t) in the frequency domain. Corresponding to equation (2.2) we can also write $z_t = \underline{k}(L)\varepsilon_t$. Consider $\underline{k}(z) = \sum_{j=0}^{\infty} K(j)z^j$ where z is a complex number. Recall that the transfer function $\underline{k}(z)$ of a purely non-deterministic process is analytic on the open unit disc.

We extend the definition of pure non-deterministicness to stationary double sequences in the following way:

Definition 3.3.1 (Innovation for a stochastic double sequence)
Suppose that (z_{it}) is a stationary stochastic double sequence of dynamic dimension q, i.e. there exists a $q\times 1$ orthonormal white noise process (u_t) together with square summable filters $(b_i : i \in \mathbb{N})$ (not necessarily causal), s.t. $z_{it} = \underline{b}_i(L)u_t$ for every $i \in \mathbb{N}$.

(i) We call an orthonormal white noise $q \times 1$ process (ε_t) innovation of (z_{it}) if there exists an index set $(i_1, ..., i_q)$ such that

$$\varepsilon_t \in \text{sp}\left(\begin{pmatrix} z_{i_1,t} \\ \vdots \\ z_{i_q,t} \end{pmatrix} - \text{proj}\left(\begin{pmatrix} z_{i_1,t} \\ \vdots \\ z_{i_q,t} \end{pmatrix} \middle| \mathbb{H}_{t-1}(z_{i_1}, ... z_{i_q})\right)\right)$$

(ii) If (ε_t) is an innovation of (z_{it}), we say (z_{it}) is purely non-deterministic if $z_{it} \in \mathbb{H}_t(\varepsilon)$ for all $i \in \mathbb{N}$ with $z_{it} = \sum_{j=0}^{\infty} K_i(j)\varepsilon_{t-j}$ and $\sum_{j=0}^{\infty} \|K_i(j)\|^2 < \infty$.

We impose the following assumption on the common component:

Assumption 7 (Purely Non-Deterministic Dynamic Common Component)
The common component (χ_{it}) *is purely non-deterministic with orthonormal white noise innovation* (ε_t) *and innovation-form*

$$\chi_{it} = \sum_{j=0}^{\infty} K_i(j)\varepsilon_{t-j} \ .$$

Note that we can also write $\chi_{it} = \underline{k}_i(L)\varepsilon_t$ as in equation (2.2). In factor analysis (American and Italian school) the common component or the factors respectively are modelled by a linear system (Stock and Watson, 2011; Forni et al., 2005, 2009; Anderson and Deistler, 2008) which implies that the common component is purely non-deterministic. Assumption A7 is much more general than supposing a rational spectrum for (χ_{it}).
Consider the sequence of $1 \times q$ row transfer functions $(k_i : i \in \mathbb{N})$. We look at partitions of consecutive $q \times q$ blocks $k^{(j)} = \left(k'_{(j-1)q+1}, ..., k'_{jq}\right)'$. Intuitively, if (χ_{it}) is the dynamic common component of a q-DFS, we would expect to "find" all q coordinates of the innovation process (ε_t) infinitely often: In other words, looking at Definition 3.3.1 there exists an infinite number of selections of the form $\tilde{\chi}_t = (\chi_{i_1,t}, ..., \chi_{i_q,t})'$, such that $(\tilde{\chi}_t)$ has innovation (ε_t). This is confirmed by the following lemma:

Lemma 3.3.2
If (χ_{it}) *is the dynamic common component of a* q*-DFS and is purely non-deterministic (A5, A7), there exists a reordering* $(k_{i_l} : l \in \mathbb{N})$ *of the sequence* $(k_i : i \in \mathbb{N})$ *such that all* $q \times q$ *blocks* $(k^{(j)})$ *of* $(k_{i_l} : l \in \mathbb{N})$ *have full rank* q *a.e. on* Θ.

Proof. By A5 and theorem 2.3.5.1, we know that

$$\lambda_q\left(f_\chi^n\right) = \lambda_q\Big(\left(k^n\right)^* k^n\Big) \to \infty \quad \text{a.e. on } \Theta \ . \tag{3.6}$$

We proof the statement by constructing such a reordering using induction.
Clearly, we can build the first $q \times q$ block, having full rank a.e. on Θ by selecting the first linearly independent rows $i_1, ..., i_q$ that we find in the sequence of row transfer functions $(k_i : n \in \mathbb{N})$, i.e. set $k^{(1)} = (k'_{i_1}, ...k'_{i_q})'$.

Now look at the block $j+1$: We start by using the next k_i available in order, as the first row of $k^{(j+1)}$, i.e. $k_{i_{jq+1}}$. Suppose we cannot find k_i with $i \in \mathbb{N} \backslash \{i_l : l \leq jq+1\}$ linearly independent of $k_{i_{jq+1}}$. Consequently, having built already j blocks of rank q, all subsequent blocks that we can built from any reordering are of rank 1. In general, for $\bar{q} < q$, suppose we cannot find rows $k_{i_{jq+\bar{q}+1}}, \ldots, k_{i_{jq+q}}$ linearly independent of $k_{i_{jq+1}}, \ldots, k_{i_{jq+\bar{q}}}$, then all consecutive blocks that we can built from any reordering have at most rank $\bar{q}$.
For all $m = j+1, j+2, \ldots$ by the RQ-decomposition we can factorise $k^{(m)} = R^{(m)}(\theta)Q^{(m)}$, where $Q^{(m)} \in \mathbb{C}^{q\times q}$ is orthonormal and $R^{(m)}(\theta)$ is lower triangular $q \times q$ filter which is analytic in the open unit disc.
Now suppose $n \geq i_j$ and say without loss of generality that n is a multiple of q:

$$
\begin{aligned}
(k^n)^* k^n &= \sum_{i=1}^{n} k_i^* k_i \\
&= \left[\left(k^{(1)}\right)^* \quad \cdots \quad \left(k^{(j)}\right)^*\right] \begin{bmatrix} k^{(1)} \\ \vdots \\ k^{(j)} \end{bmatrix} + \left[\left(R^{(j+1)}\right)^* \quad \cdots \quad \left(R^{(n/q)}\right)^*\right] \begin{bmatrix} R^{(j+1)} \\ \vdots \\ R^{(n/q)} \end{bmatrix} \\
&= \left[\left(k^{(1)}\right)^* \quad \cdots \quad \left(k^{(j)}\right)^*\right] \begin{bmatrix} k^{(1)} \\ \vdots \\ k^{(j)} \end{bmatrix} + \begin{pmatrix} \times & 0 \\ 0 & 0 \end{pmatrix} = A + B^n, \text{ say,}
\end{aligned}
$$

where $\times$ is a placeholder. By the structure of the reordering, there are $q - \bar{q}$ zero end columns/rows in B^n for all $n \geq jq$ where A remains unchanged.
Now by Lancaster and Tismenetsky (1985, theorem 1, p.301), we have

$$
\begin{aligned}
\lambda_q\Big((k^n)^* k^n \Big) &= \lambda_q(A + B^n) \\
&\leq \lambda_1(A) + \lambda_q(B^n) \\
&= \lambda_1(A) < \infty \text{ for all } n \in \mathbb{N} \text{ a.e. on } \Theta .
\end{aligned}
$$

However this is a contradiction to (3.6) which completes the induction step and the proof. ∎

We impose the following regularity condition that will be relaxed below:

Assumption 8 (Strictly Miniphase)
There exists a "strictly miniphase" sub-sequence $k^{(j_l)}$ of $k^{(j)}$ of transfer functions in the sense that $\lambda_q\left(k^{j_l}(k^{j_l})^*\right) > \delta > 0$ a.e. on Θ for all j_l.

Theorem 3.3.3
Suppose A1, A5, A7, A8 hold for the q-DFS (y_{it}) with dynamic common component (χ_{it}), then the innovations (ε_t) of (χ_{it}) are causally subordinated to (y_{it}), i.e. $\varepsilon_t \in \mathbb{H}_t(y)$.

Proof. Suppose $(k_i : i \in \mathbb{N})$ is re-ordered such that all $q \times q$ blocks $k^{(j)}$ for $j = 1, 2, ...$ are of full rank a.e. on Θ (such as in lemma 3.3.2). Next, redefine $(k_i : i \in \mathbb{N})$ as a suitable sub-sequence of itself such that A8 holds. Suppose q divides n without loss of generality. We look at

$$\chi_t^n = \begin{pmatrix} \chi_t^{(1)} \\ \chi_t^{(2)} \\ \vdots \\ \chi_t^{(n/q)} \end{pmatrix} = \begin{pmatrix} \underline{k}^{(1)}(L) \\ \vdots \\ \underline{k}^{(n/q)}(L) \end{pmatrix} \varepsilon_t = \begin{pmatrix} \underline{k}^{(1)}(L) & & \\ & \ddots & \\ & & \underline{k}^{(n/q)}(L) \end{pmatrix} \begin{pmatrix} I_q \\ \vdots \\ I_q \end{pmatrix} \varepsilon_t \, .$$

By the Wold Representation we know that all $k^{(j)}, j = 1, ..., n/q$ are analytic in the open unit disc and $\det \underline{k}^{(j)}(z) \neq 0$ for all $|z| < 1$ and $j = 1, 2, ...$

$$\begin{aligned} \varphi_t^n &:= \begin{pmatrix} \left(\underline{k}^{(1)}\right)^{-1}(L) & & \\ & \ddots & \\ & & \left(\underline{k}^{(n/q)}\right)^{-1}(L) \end{pmatrix} \begin{pmatrix} y_t^{(1)} \\ \vdots \\ y_t^{(n/q)} \end{pmatrix} \\ &= \begin{pmatrix} I_q \\ \vdots \\ I_q \end{pmatrix} \varepsilon_t + \begin{pmatrix} \left(\underline{k}^{(1)}\right)^{-1}(L) & & \\ & \ddots & \\ & & \left(\underline{k}^{(n/q)}\right)^{-1}(L) \end{pmatrix} \begin{pmatrix} \xi_t^{(1)} \\ \vdots \\ \xi_t^{(n/q)} \end{pmatrix} = \begin{pmatrix} I_q \\ \vdots \\ I_q \end{pmatrix} \varepsilon_t + \zeta_t^n \text{ , say.} \end{aligned} \tag{3.7}$$

Clearly, the first term on the RHS is a *q-static* factor sequence, where all eigenvalues diverge (at rate n). Therefore, if the double sequence (ζ_{it}) corresponding to ψ_t^n on the RHS is statically idiosyncratic, we obtain ε_t (up to a rotation) from static averaging over (φ_{it}) by theorem 2.2.7. Consequently, also (ε_t) is causally subordinated to (y_{it}).
To see why (ζ_{it}) is statically idiosyncratic, let $U_j \Sigma_j V_j^* = k^{(j)}(\theta)$ be the singular

value decomposition of $k^{(j)}(\theta)$, where we suppressed the dependence on θ in the notation on the LHS. Let $f_\xi^n(\theta) = P\Lambda P^*$ be the eigen-decomposition of f_ξ^n with orthonormal eigenvectors being the columns of P. Then

$$f_\psi^n(\theta) = \bigoplus_{j=1}^J U_j \underbrace{\bigoplus_{j=1}^J \Sigma_j^{-1} \bigoplus_{j=1}^J V_j^* P\Lambda P^* \bigoplus_{j=1}^J V_j \Sigma_j^{-1} \bigoplus_{j=1}^J}_{A^n} U_j^* \ .$$

The largest eigenvalue of $f_\zeta^n(\theta)$ is equal to the largest eigenvalue of $A^n(\theta)$ which

$$\begin{aligned}
\lambda_1\left(\int_\Theta f_\zeta^n\right) &\leq \int_\Theta \lambda_1\left(f_\zeta^n\right) \\
&\leq 2\pi \operatorname{ess\,sup}_\theta \sup_n \lambda_1\left(f_\zeta^n\right) \\
&\leq 2\pi \operatorname{ess\,sup}_\theta \sup_n \lambda_1\left(f_\xi^n\right)\left(\inf_j \lambda_q\left(\Sigma_j\right)\right)^{-2} \\
&< 2\pi \operatorname{ess\,sup}_\theta \sup_n \lambda_1\left(f_\xi^n\right)(\delta)^{-2} < \infty \ ,
\end{aligned}$$

which completes the proof by theorem 2.2.6. ∎

Note that condition A8 can be relaxed. Firstly, look at the following example:

$$y_{it} = \chi_{it} + \xi_{it} = (1-L)\varepsilon_t + \xi_{it} \ ,$$

where (u_t) is orthonormal white noise and (ξ_{it}) is dynamically idiosyncratic. Consider the cross-sectional average $\bar{y}_t = n^{-1}\lim_n \sum_{i=1}^n y_{it} = (1-L)\varepsilon_t$. The innovations $\varepsilon_t \in \mathbb{H}_{t-1}(y)$ are the innovations of (χ_{it}) and causally subordinated to (y_{it}). More generally, recall that we can "factor out" zeros from an analytic function: So we can write $\det \underline{k}^{(j)} = \underline{g}_j \underline{h}_j$, where $\underline{g}_j$ is a polynomial with all zeros of $\det \underline{k}^{(j)}$ which are on the unit circle. As is well known, the zeros of an analytic function are isolated, so if z_0 is a zero, we have $\underline{h}_j(z) \neq 0$ in a neighbourhood of z_0. Furthermore the degree of a zero can be only finite or the function is zero everywhere. This implies there can be only finitely many different zeros on a compact set (like the unit circle). So we can write $\underline{g}_j(z) = \Pi_{l=1}^{J_j}(z - z_{j1})^{m_l}$ where $|z_{jl}| = 1$. It follows that $g_j^{-1}(\theta) \det k^{(j)}(\theta) = h_j(\theta) \neq 0$ a.e. on Θ, where $g_j(\theta) := \underline{g}_j(e^{-i\theta})$ and

$h_j(\theta) := \underline{h}_j(e^{-i\theta})$. Thus instead we may assume that there exists a sub-sequence of $(k^{(j)} : j \in \mathbb{N})$ such that $h_j(\theta) > \delta$ a.e. on Θ. Therefore equation (3.7) becomes instead:

$$\varphi_t^n = \begin{pmatrix} \underline{g}_1(L)I_q \\ \vdots \\ \underline{g}_{n/q}(L)I_q \end{pmatrix} \varepsilon_t + \begin{pmatrix} (\underline{h}_1)^{-1}(L) & & \\ & \ddots & \\ & & (\underline{h}_{n/q})^{-1}(L) \end{pmatrix} \begin{pmatrix} \xi_t^{(1)} \\ \vdots \\ \xi_t^{(n/q)} \end{pmatrix}. \tag{3.8}$$

Hence instead of A8 it is enough to assume 1. that there exists a sub-sequence of reordered $(k^{(j)} : j \in \mathbb{N})$, such that after factoring out zeros, we have $\inf_j \lambda_q(h_j) > \delta$ a.e. on Θ. 2. that we can compute a static aggregate of the left term on the RHS of (3.7) which has ε_t as its innovations.
Summing up, we may argue that it is the very feature of dynamic factor sequences, as opposed to the "usual" DLRA of double sequences *without* a factor structure, that the innovations of (χ_{it}) are fundamental to the observed double sequence (y_{it}). The causal subordination originates from the fact that the remainder of the rank-q DLRA vanishes under aggregation and thus the second term on the RHS of (3.8) vanishes under aggregation. Thus, we may say that the non-fundamentalness of the shocks arising from DLRA is not a structural feature of dynamic factor sequences but rather a matter of estimation technique.

3.4 A State Space Representation

In this section, we propose how to model factor sequences incorporating strong and weak factors in a state space framework. Linear system representations for factor models without the incorporation of weak factors have been investigated in Anderson and Deistler (2008); Deistler et al. (2010). Let (ε_t) be a q-dimensional orthonormal white noise process of innovations of (χ_{it}). We distinguish between the following observation equations:

$$\text{dynamic CC:} \quad \chi_t^n = \Lambda^n F_t = \begin{bmatrix} \Lambda_s^n & \Lambda_w^n \end{bmatrix} \begin{pmatrix} F_t^s \\ F_t^w \end{pmatrix} = \underbrace{\begin{bmatrix} \Lambda_s^n & \Lambda_w^n & 0 \end{bmatrix}}_{H^n} \underbrace{\begin{pmatrix} F_t^s \\ F_t^w \\ x_t^r \end{pmatrix}}_{x_t} \tag{3.9}$$

$$\text{static factors:} \quad F_t = \begin{bmatrix} I_r & I_{r_\chi^+} & 0 \end{bmatrix} x_t \tag{3.10}$$

$$\text{static CC:} \quad C_t^n = \Lambda_s^n F_t^s = \begin{bmatrix} \Lambda_s^n & 0 & 0 \end{bmatrix} x_t \tag{3.11}$$

$$\text{strong factors:} \quad F_t^s = \begin{bmatrix} I_r & 0 & 0 \end{bmatrix} x_t \; . \tag{3.12}$$

The corresponding transition equation is:

$$x_{t+1} = \begin{pmatrix} F_{t+1}^s \\ F_{t+1}^w \\ x_{t+1}^r \end{pmatrix} = \underbrace{\begin{pmatrix} M_{ss} & M_{sw} & M_{sr} \\ M_{ws} & M_{ww} & M_{wr} \\ M_{sr} & M_{sw} & M_{rr} \end{pmatrix}}_{M} \begin{pmatrix} F_t^s \\ F_t^w \\ x_t^r \end{pmatrix} + \underbrace{\begin{pmatrix} G_s \\ G_w \\ G_r \end{pmatrix}}_{G} \varepsilon_{t+1} \; . \tag{3.13}$$

In short we have

$$y_t^n = \chi_t^n + \xi_t^n = C_t^n + e_t^{\chi,n} + \xi_t^n = C_t^n + e_t^n$$

$$\chi_t^n = H^n x_t = \Lambda_s^n F_t^s + \Lambda_w^n F_t^w \tag{3.14}$$

$$x_{t+1} = M x_t + G\varepsilon_{t+1} \; . \tag{3.15}$$

In the finite dimensional case $r_\chi < 0$ the growing cross-sectional dimension is associated with adding rows to H^n or Λ^n respectively. Equations (3.9) to (3.12) are observation or "measurement" equations for the dynamic common component, the static factors, the static common component and the strong static factors respectively. Equation (3.13) is the corresponding transition equation.
The rational transferfunction of (χ_t^n) is uniquely determined on $z \in \mathbb{C}$ and given

by

$$\underline{k}(z) =: \underline{k}^n(z) =: H^n(I_{r_\chi} - Mz)^{-1}G \ ,$$

where z is a complex number. The spectrum of (χ_t^n) is rational and given by $\underline{f}_\chi^n(z) = \underline{k}(z)\underline{k}^*(z)$ where $\underline{k}^*(z) := \overline{\underline{k}(\overline{z}^{-1})}'$. See Deistler and Scherrer (2022) for details.

The transfer function of the strong static factors is

$$\underline{k}^s(z) =: \begin{bmatrix} I_r & 0 & 0 \end{bmatrix} (I_{r_\chi} - Mz)^{-1}G \ .$$

We suppose that the parameters (M, G, H^n) are such that (y_{it}) satisfies assumptions A5 and A2. We assume that $\mathbb{E}\, F_t^s\,(F_t^w)' = 0$ and the loadings Λ_s^n, Λ_w^n are of dimension $(r \times n)$ and $(r_\chi - r \times n)$ respectively and like in theorem 3.2.2.3 which ensures that C_{it} is the static common component of (y_{it}) and spanned by the strong factors (F_t^s).

Furthermore we impose the following standard linear system assumption:

Assumption 9 (Canonical State Space Representation)
The dynamic common component (χ_t^n) is generated by the system (M, G, H^n) for all $n \in \mathbb{N}$ in equations (3.9) to (3.12). In addition we assume that (M, G, H^n) is in canonical state space representation (see e.g. Hannan and Deistler, 2012, for details), which includes

(i) *Minimality of (M, G, H^n) with minimal state dimension m. The state dimension does not increase for all $n \geq N(m)$. We suppose that $n \geq N(m)$.*

(ii) *The system is stable which is satisfied if $\lambda_1(M) < 1$*

(iii) *The system is miniphase.*

(iv) rk $H^n = r_\chi$.

Minimality means that the dimension of the state x_t, say $m \geq r_\chi$, cannot be reduced (for details on linear system theory see e.g. Hannan and Deistler, 2012;

Deistler and Scherrer, 2022, chapters 2 and 7 respectively) and is an important condition for identification and estimation. The coordinates x_t^r in x_t in equation (3.13) are potential "remainder" state dimensions that we may need to describe the full dynamics of the system. The static factors F_t are the minimal static factors for (χ_t^n) in the sense of Anderson and Deistler (2008) which is ensured by $\operatorname{rk} H^n = r_\chi$.

Stability ensures that (χ_{it}) is stationary and the transferfunction does not explode. The miniphase condition ensures the left-invertability of the transfer function meaning that the innovation ε_t can be obtained from multiplying the left-inverse of $\underline{k}(z)$ to χ_t. If

$$\operatorname{rk} \underline{\mathcal{M}}(z) := \operatorname{rk} \begin{bmatrix} I_m - Mz & -G \\ H^n & 0 \end{bmatrix} = m + q \text{ for all } |z| < 1 \,, \tag{3.16}$$

the miniphase condition is satisfied (see e.g. Kailath, 1980; Anderson and Deistler, 2008).

Next, the question naturally arises under which conditions, the dynamic common component (χ_{it}) is identified from the strong static factors (F_t^s) alone. In this case, we could obtain the dynamic common component from SLRA - bypassing DLRA and therefore frequency domain methods which would improve the practical usability. To examine this, we look at the state space system corresponding to the strong static factors, i.e. the system given by equations (3.12) and (3.13). By A9.(ii) we know that also $k^s(z)$ is stable. If the transfer-function of the strong static factors is also miniphase which can be checked by condition (3.16) replacing H^n with $[I_r, 0, 0]$, this implies that $\mathbb{H}_t(\varepsilon) = \mathbb{H}_t(F^s)$ by the Wold representation theorem (see e.g. Deistler and Scherrer, 2022, ch. 2). Consequently, if $\underline{k}^s(z)$ is miniphase, then the dynamic common component is identified from projecting y_{it} on the infinite past of (F_t^s), i.e. $\chi_{it} = \operatorname{proj}(y_{it} \mid \mathbb{H}_t(F_t^s))$. In this case also χ_{it} is causally subordinated to y_{it}, i.e. $\mathbb{H}_t(\chi) \subset \mathbb{H}_t(y)$ since $\mathbb{H}_t(\chi) \subset \mathbb{H}_t(F^s) \subset \mathbb{H}_t(y)$. Note that, for this identification procedure we do not even need to assume that χ_{it} is generated by a state space system. It is sufficient to assume that (F_t^s) and (χ_{it}) have the same innovations.

If we suppose in addition that the system (3.12) and (3.13) is also minimal - even easier, the dynamic common component is the projection of the output variables on the state of the strong static factors. Recall that a state space system is minimal if and only if it is observable and controllable (see e.g. Deistler and Scherrer, 2022, ch. 7). Controllability is a feature of the matrices (M, G) in the transition equation (3.13) and holds by A9.(i). Observability is a feature of the system matrices in the observation equation, i.e. here of $([I_r, 0, 0], M)$ and is satisfied e.g. by the Popov-Belevitch-Hautus-Test (see Kailath, 1980, ch. 2.4) if all right eigenvectors of M are not in the right-kernel of $[I_r, 0, 0]$.
In either case, we have the following relations that allow identification of the dynamic common common component from the strong static factors:

$$\begin{aligned}
\mathbb{H}(C) &= \mathbb{H}(F^s) = \mathbb{G}(y) \\
\mathbb{H}_t(C) &= \mathbb{H}_t(F^s) = \mathbb{H}_t(\varepsilon) = \mathbb{H}_t(\chi) \\
\chi_{it} &= \operatorname{proj}(y_{it} \mid \mathbb{H}(\varepsilon)) = \operatorname{proj}(y_{it} \mid \mathbb{H}_t(\varepsilon)) = \operatorname{proj}(y_{it} \mid \mathbb{H}_t(F^s)) \\
&= \operatorname{proj}(y_{it} \mid \operatorname{sp}(x_t)) \, .
\end{aligned}$$

We summarise the discussion above:

Theorem 3.4.1 (Identifiability from Strong Factors)
Suppose A1, A2, A5 hold.
1. If in addition A7, A8 hold, (F_t^s) is generated by the system (3.12, 3.13) and $\underline{k}^s(z)$ is miniphase, then the dynamic common component is identified from the strong static factors via $\chi_{it} = \operatorname{proj}(y_{it} \mid \mathbb{H}_t(F^s))$.
2. If A6 and A9 hold and $([I_r, 0, 0], M)$ is observable, then $\chi_{it} = \operatorname{proj}(y_{it} \mid \operatorname{sp}(x_t))$.

Theorem 3.4.1 provides another structural interpretation of weak factors: We can regard dynamic factor sequences as a “natural extension” of static factor sequences: The dynamic common component is the part of y_{it}, that is driven not only by the contemporaneous strong static factors but also by their infinite past or state dimensions respectively.
The vast majority of “finite dimensional” approximate factor models suggested in the econometrics literature can be cast in the state space form above. For example

a very common model is to incorporate dynamics *within* the strong static factors as (see e.g. Stock and Watson, 2005, 2011; Bai and Wang, 2016; Bai and Ng, 2007):

$$y_{it} = \lambda_{i0}^s f_t^s + \lambda_{i1}^s f_{t-1}^s + \cdots + \lambda_{i,p_s} f_{t-p_C}^s + e_{it} \tag{3.17}$$
$$C_{it} = \Lambda_i^s F_t^s \tag{3.18}$$

where (f_t^s) is $q \times 1$ and are called "dynamic factors" (note that these dynamic factors do in general not coincide with the definition of dynamic factors that we use in this thesis as the common orthonormal white noise shocks which span the dynamic aggregation space) and

$$\lambda_i^s = \begin{bmatrix} \lambda_{i0}^s & \lambda_{i1}^s & \cdots & \lambda_{i,p_C}^s \end{bmatrix} \text{ and } F_t^s = \begin{bmatrix} f_t^s \\ f_{t-1}^s \\ \vdots \\ f_{t-p_C}^s \end{bmatrix} .$$

So λ_i^s is the i-th row of Λ_s in equation (3.9). Furthermore the factors (f_t^s) are modelled as a VAR(p_f) process:

$$f_t^s = A_1^{f,s} f_{t-1}^s + A_2^{f,s} f_{t-1}^s + \cdots + A_p^{f,s} f_{t-p_f}^s + \varepsilon_t .$$

It follows that also the stacked vector of strong static factors (F_t^s) can be represented as solution of a singular VAR(p) system, so

$$F_t^s = A_1^F F_{t-1}^s + A_2^F F_{t-2}^s + \cdots + A_p^s F_{t-p}^s + b\varepsilon_t , \tag{3.19}$$

where b is $r \times q$ and $p = \max\{p_C + 1, p_f\}$. A VAR system is called singular if the innovation variance matrix is singular. The properties of singular VAR systems also in connection to factor models have extensively been studied e.g. in Deistler et al. (2010, 2011); Anderson et al. (2012); Chen et al. (2011). It is easy to see, how to put such a system into the state space representation (3.14), (3.15). Minimality of that system can always be achieved by reducing the state dimension

(Hannan and Deistler, 2012, ch. 2).
It is also common to *only* suppose that the strong static factors follow a VAR(p)-system (see e.g. Doz et al., 2011; Ruiz et al., 2022), i.e. only suppose (3.19) - without the additional restrictions given by (3.17) but usually with regular innovation variance matrix.

Remark 3.4.2 (Most Dynamic Factor Models are not as Dynamic as They Think) *Models like (3.17), (3.18), (3.19) are also referred to as "dynamic factor models" and the representation (3.18) together with (3.17) is also referred to as "casting the dynamic factor model in static form" (see e.g. Stock and Watson, 2011). Clearly, these models are "dynamic" in the sense that the strong factors are modelled in a dynamic way. However, they are not dynamic in the sense that the strong static factors in equation (3.18) emerge from static rather than from dynamic aggregation. As a consequence the model (3.17) is structurally not comparable to the dynamic factor sequences from Forni et al. (2000); Forni and Lippi (2001) as is often stated.*

Chapter 4

Implications for Factor Analysis

The structural results discussed in the previous two sections have a number of important implications for the theory and practice of factor analysis. Let us first turn to a structural interpretation.

4.1 Structural Implications

Given our discussion in section 3.3, the innovations (ε_t) of the dynamic common component (χ_{it}) are to be interpreted as the *common innovation process* or the *common fundamental shocks* or the *structural dynamic shocks* (see e.g. Stock and Watson, 2005) of the economy. The dynamic common component is the projection of the observed variables on the infinite past of these structural shocks.
Consequently, if we are interested in finding the part that is driven by the structural shocks of the economy, we make in general a structural error, which does not vanish for $(n, T) \to \infty$, if we would merely use the static common component and ignore the impact of weak static factors. The size of this error in general varies with the cross-sectional units and depends on the data generating process. In other words, if the cross-sectional unit we are interested in, say y_{it} has a non-trivial weak common component, the projection onto the static strong factors alone $C_{it} = \text{proj}(y_{it} \mid F_t^s)$ does not tell the whole story - but only represents the part that is *contemporaneously common*. We have to carefully distinguish between

contemporaneously- and *dynamically* common. The common component based on SLRA of the American school captures the *contemporaneous* co-movement whereas the DLRA of the Italian school captures the *dynamic* co-movement of the variables. Both parts might be of structural interest for the researcher but have to be kept separate when interpreting and analysing them.
Nonetheless, we still may specify a time series model for the strong static factors but also this does not make the corresponding common component to be the dynamic common component (see Remark 3.4.2), i.e. capture the dynamic co-movement.

Remark 4.1.1 (Impulse Responses to Structural Dynamic Shocks and Factors as Intstrument Variables)
The consideration of a non-trivial weak common component may be important e.g. when using the common component by means of removing measurement error (see Lippi, 2021). Literally, a measurement error is reintroduced when using the static rather than the dynamic common component. Also, if we consider the impulse responses to structural shocks like in Forni et al. (2009) or in terms of a factor augmented VAR Stock and Watson (2005), we induce a "population error" whenever the weak common component is non-trivial but we estimate the common component via SLRA. These papers are all correct as they contain the implicit assumption that $C_{it} = \chi_{it}$ for all $i \in \mathbb{N}$. However, this is not innocent. We induce an asymptotically non-vanishing error to impulse response analysis by not controlling for the presence of weak factors.
Another important application of factor analysis is to use the strong static factors as instrumental variables Bai and Ng (2010). If the dependent variable in the regression equation also depends on weak factors, incorporating them as instruments will reduce variance of the parameter estimates while maintaining instrument-exogeneity.

Remark 4.1.2 (Reconsidering Integer Parameters)
Under the assumptions of theorem 3.1.1, if we furthermore assume that the strong static factors (F_t^s) are purely non-deterministic, the spectrum of (F_t^s) has rank $q_C \leq r$ almost everywhere on Θ: Since $\mathbb{S}_t \subset \mathbb{G}$, the innovations of (F_t^s) are of

dimension $q_C \leq q$.
We may also assume that $\sup_n r_\chi(n) = \dim \overline{\mathrm{sp}}(\chi_t) = r_\chi < \infty$ *is finite dimensional which implies a static factor structure where* $r_\chi \leq r$ *(see theorem 3.2.2). In summary, we distinguish the following integer parameters* q_C, q, r, r_χ *with*

$$q_C \leq r \leq r_\chi$$
$$q_C \leq q \leq r_\chi \ .$$

Note that methods which determine the number of dynamic factors via the dynamic dimension of the strong factors (F_t^s) *(see e.g. Bai and Ng, 2007) target* q_C *rather than* q. *However, it might be that* $q_C < q$ *only happens for very pathological cases (see example 3.1.3).*
Theorem 3.2.2 also implies that in general we cannot use methods like in Bai and Ng (2002); Ahn and Horenstein (2013) which are designed for determining the number of strong static factors, for estimating $r_\chi(n)$ *the dimension of the dynamic common component unless we assume that* $e_{it}^\chi = 0$ *for all* $i \leq n$. *This is common practice (see e.g. Forni et al., 2005, 2009; Barigozzi and Luciani, 2019). For example Forni et al. (2005) use first DLRA to estimate* Γ_χ^n *and in a second step approximate* χ_t^n *with a static factor structure using an optimisation procedure based on "generalised principal components". The proposed algorithm enforces* $e_{it}^\chi = 0$ *for all* $i \leq n$ *which makes it - from a structural point of view - unnecessary to estimate* Γ_χ^n *with frequency domain methods in the first place.*

Remark 4.1.3
It is often assumed that (F_t^s) *and* (e_{it}) *are uncorrelated for all leads and lags (see e.g. Bai and Ng, 2006, Assumption D). Again, this is only the case if* $e_{it}^\chi = 0$ *for all* i, t *which is in general not the case.*

4.2 Estimation of the Dynamic Common Component

By Onatski (2012) it was shown that the method of static principal components is not a consistent estimator for weak factors. While Onatski (2012) considered weak factors in general, the same holds true for weak factors that live in the dynamic aggregation space and are part of the dynamic common component. In Forni et al. (2000, 2004) it was shown that sample DLRA is consistent for the dynamic common component for $(n, T) \to \infty$. In the following we demonstrate by means of a Monte-Carlo simulation that sample DLRA can capture the weak common component wheres sample SLRA cannot.

For this, consider the following model: Let $\varepsilon_t \sim \mathcal{N}(0, 1)$ be scalar Gaussian white noise. We construct an idiosyncratic component with cross-sectional correlation as follows: Let $w_i = (1, 1 + 1/20, 1 + 2/20, ..., 1 + n/20)$ be a vector of weights. We draw independently $\lambda_i^\xi \sim w_i^{-1} \times \mathcal{N}(0, 1)$. Now set $\xi_{it} = \lambda_i^\xi \varepsilon_t^{\xi,1} + \varepsilon_{it}^{\xi,2}$, where $\varepsilon_t^{\xi,1} \sim \mathcal{N}(0, 1)$ and $\varepsilon_{it}^{\xi,2} \sim \mathcal{N}(0, 1)$ are drawn independently, also independent from (ε_t).

To obtain data generated from a state space system, we consider $r_\chi = 2$ with one strong factor F_t^s and one weak factor F_t^w:

$$y_{it} = \chi_{it} + \xi_{it} \tag{4.1}$$

$$\chi_{it} = F_t^w \quad \text{for} \quad i = 1, ..., 10 \quad \text{and} \quad \chi_{it} = F_t^s \quad \text{for} \quad i = 11, ..., n \tag{4.2}$$

$$\begin{aligned}\underbrace{\begin{pmatrix} F_{t+1}^s \\ F_{t+1}^w \end{pmatrix}}_{x_{t+1}} &= \underbrace{\begin{bmatrix} M_{ss} & M_{sw} \\ M_{ws} & M_{ww} \end{bmatrix}}_{M} \begin{pmatrix} F_t^s \\ F_t^w \end{pmatrix} + \underbrace{\begin{bmatrix} G_s \\ G_w \end{bmatrix}}_{G} \varepsilon_{t+1} \\ &= \underbrace{\begin{bmatrix} 0.1945375 & -0.3842384 \\ 0.2702844 & 0.9054625 \end{bmatrix}}_{M} \begin{pmatrix} F_t^s \\ F_t^w \end{pmatrix} + \underbrace{\begin{bmatrix} 0.9025054 \\ 0.3272368 \end{bmatrix}}_{G} \varepsilon_{t+1} \,,\end{aligned} \tag{4.3}$$

where the parameters were chosen such that $\mathbb{V}\, x_t = I_2$. It is easy to see that the first eigenvalue of the spectrum of χ_t^n diverges a.e. on the frequency band.

Furthermore $\lambda_1(\Gamma_\chi^n) = n - 10$ diverges with rate n, whereas the second eigenvalue of $\lambda_2(\Gamma_\chi^n) = 10$ is bounded.
We compare three different methods:

(i) Estimation with sample DLRA. We estimate the spectrum of (y_t^n) using the lag-window estimator

$$\hat{f}_y^n(\theta) = (2\pi)^{-1} \sum_{k=-M(T)}^{M(T)} \kappa(k/M(T)) e^{-\iota k\theta} \hat{\Gamma}_y^n(k) \ ,$$

where $\hat{\Gamma}_y^n(k) = T^{-1} \sum_{t=|k|+1} y_t^n y_{t-|k|}^{n}{}'$ and $\kappa(\cdot)$ is the Bartlett kernel where $M(T) = 0.75\sqrt{T}$. We compute DLRA from the estimated spectrum with $q = 1$ as in Forni et al. (2000).

(ii) Estimation with SLRA for $r = 1$, where we compute the SLRA from the sample variance matrix $\hat{\Gamma}_y^n$.

(iii) Estimation with SLRA for $r = 2$, estimates are computed as in (ii).

As a performance measure, we evaluate the average mean squared error (AMSE) - *averaging over the cross-sectional index set* I with cardinality $\#I$,

$$AMSE = \frac{1}{\#I} \sum_{i \in I} \frac{1}{T} \sum_{t=1}^{T} (\chi_{it} - \hat{\chi}_{it})^2 \ , \tag{4.4}$$

where χ_{it} in (4.4) is the true common component and $\hat{\chi}_{it}$ is the estimated common component. As the final performance measure, we take the average of (4.4) over all replications.
The results of (4.4) for $i = 1, ..., 10$ are shown in table 4.1. Table 4.2 shows the results for $i = 11, ..., n$, and table 4.3 shows the results for the whole cross-section, i.e. $I = \{1, 2, ..., n\}$. Some observations in order: Table 4.1 reveals that in our example, DLRA can estimate the weak common component better if n and T gets larger, and has difficulties to estimate χ_{it} - especially while T is still small. In contrast, estimates for the weak common component of SLRA do not improve for increasing n, T as expected.

	d,T=60	s,T=60	s2,T=60	d,T=120	s,T=120	s2,T=120	d,T=240	s,T=240	s2,T=240	d,T=480	s,T=480	s2,T=480	d,T=960	s,T=960	s2,T=960
n=30	0.586	1.054	0.696	0.533	1.022	0.648	0.519	1.016	0.684	0.51	1.009	0.694	0.517	1.007	0.697
n=30	(0.162)	(0.259)	(0.408)	(0.128)	n= (0.165)	(0.396)	(0.101)	(0.102)	(0.377)	(0.077)	(0.075)	(0.324)	(0.059)	(0.049)	(0.256)
n=60	0.571	1.049	0.894	0.541	1.003	1.001	0.511	1	1.066	0.514	1.006	1.142	0.51	1.007	1.172
n=60	(0.151)	(0.216)	(0.443)	(0.116)	(0.143)	(0.388)	(0.084)	(0.101)	(0.327)	(0.064)	(0.07)	(0.255)	(0.052)	(0.048)	(0.178)
n=120	0.547	1.043	1.119	0.495	1.013	1.201	0.482	1.005	1.287	0.476	0.995	1.329	0.478	1.001	1.351
n=120	(0.14)	(0.202)	(0.392)	(0.113)	(0.145)	(0.298)	(0.077)	(0.105)	(0.185)	(0.056)	(0.072)	(0.115)	(0.048)	(0.05)	(0.08)
n=240	0.48	1.022	1.219	0.423	1.013	1.304	0.394	1.005	1.357	0.395	1.004	1.378	0.398	1.004	1.386
n=240	(0.128)	(0.218)	(0.328)	(0.089)	(0.148)	(0.239)	(0.076)	(0.105)	(0.14)	(0.055)	(0.074)	(0.094)	(0.042)	(0.052)	(0.066)
n=480	0.425	1.024	1.238	0.319	0.998	1.314	0.279	1.011	1.37	0.276	1.002	1.389	0.285	1.003	1.393
n=480	(0.126)	(0.214)	(0.291)	(0.074)	(0.141)	(0.199)	(0.062)	(0.107)	(0.129)	(0.057)	(0.071)	(0.088)	(0.043)	(0.051)	(0.061)

Table 4.1: Units influenced by weak factors only: AMSE (equation 4.4) over 500 replications for $I = \{1, ..., 10\}$ of model (4.1)-(4.3): d = DLRA, s = SLRA, $r = 1$, s2 = SLRA with $r = 2$. Standard errors in parentheses.

	d,T=60	s, T=60	s2,T=60	d, T=120	s,T=120	s2,T=120	d,T=240	s,T=240	s2,T=240	d,T=480	s,T=480	s2,T=480	d,T=960	s,T=960	s2,T=960
n=30	0.288	0.286	0.31	0.251	0.15	0.236	0.221	0.088	0.214	0.203	0.07	0.205	0.193	0.064	0.199
n=30	(0.082)	(0.267)	(0.17)	(0.059)	(0.15)	(0.12)	(0.045)	(0.055)	(0.102)	(0.031)	(0.021)	(0.085)	(0.024)	(0.012)	(0.066)
n=60	0.179	0.057	0.169	0.163	0.038	0.16	0.136	0.03	0.153	0.125	0.027	0.155	0.114	0.025	0.155
n=60	(0.04)	(0.031)	(0.067)	(0.028)	(0.012)	(0.055)	(0.02)	(0.006)	(0.045)	(0.015)	(0.004)	(0.035)	(0.011)	(0.002)	(0.025)
n=120	0.121	0.033	0.123	0.11	0.021	0.108	0.09	0.015	0.103	0.08	0.013	0.1	0.072	0.012	0.099
n=120	(0.03)	(0.011)	(0.033)	(0.018)	(0.004)	(0.024)	(0.011)	(0.002)	(0.016)	(0.007)	(0.001)	(0.01)	(0.006)	(0.001)	(0.007)
n=240	0.082	0.026	0.087	0.075	0.014	0.069	0.059	0.009	0.062	0.051	0.007	0.057	0.044	0.006	0.055
n=240	(0.026)	(0.007)	(0.017)	(0.015)	(0.002)	(0.011)	(0.008)	(0.001)	(0.007)	(0.005)	(0)	(0.004)	(0.003)	(0)	(0.003)
n=480	0.061	0.022	0.063	0.055	0.011	0.045	0.039	0.007	0.036	0.033	0.004	0.032	0.028	0.003	0.03
n=480	(0.023)	(0.006)	(0.01)	(0.013)	(0.001)	(0.005)	(0.007)	(0.001)	(0.003)	(0.004)	(0)	(0.002)	(0.002)	(0)	(0.001)

Table 4.2: Units influenced by strong factors only: AMSE (equation 4.4) over 500 replications for $I = \{11, ..., n\}$ of model (4.1)-(4.3): d = DLRA, s = SLRA, $r = 1$, s2 = SLRA with $r = 2$. Standard errors in parentheses.

For the cross-sectional units influenced only by the strong factors, i.e. $i = 11, ..., n$, SLRA with $r = 1$ outperforms DLRA especially if n/T is large. Asymptotically DLRA catches up, but does not reach the performance of SLRA with $r = 1$ in our setup. This suggests SLRA benefits from being the more parsimonious and stable procedure compared to DLRA. However if n/T is small DLRA is slightly better than the "correctly specified" SLRA with $r = 1$ which indicates that the identification of the factors benefits from dynamic averaging if the spectrum is estimated more precisely (with larger T). Of course this effect would not occur if the factors were serially uncorrelated. For $r = 2$ SLRA seems also to be consistent (see Barigozzi and Cho, 2020) but has higher variance.
Finally, the results in table 4.3 of AMSE evaluated over the whole index set are not really different from the results in table 4.2. When it comes to evaluating whether the weak common component is captured well, AMSE is not a suitable evaluation criterion since the contribution of the weak common component is "averaged out" by taking the cross-sectional mean. Instead we shall rather look at each series

	d,T=60	s,T=60	s2,T=60	d,T=120	s,T=120	s2,T=20	d,T=240	s,T=240	s2,T=240	d,T=480	s,T=480	s2,T=480	d,T=960	s,T=960	s2,T=960
n=30	0.387	0.542	0.439	0.345	0.44	0.374	0.32	0.397	0.371	0.305	0.383	0.368	0.301	0.379	0.365
n=30	(0.103)	(0.195)	(0.225)	(0.079)	(0.112)	(0.204)	(0.062)	(0.048)	(0.192)	(0.045)	(0.028)	(0.164)	(0.034)	(0.018)	(0.129)
n=60	0.245	0.223	0.29	0.226	0.199	0.3	0.199	0.192	0.305	0.189	0.19	0.319	0.18	0.189	0.324
n=60	(0.054)	(0.045)	(0.127)	(0.041)	(0.026)	(0.109)	(0.029)	(0.017)	(0.091)	(0.023)	(0.012)	(0.071)	(0.018)	(0.008)	(0.05)
n=120	0.157	0.117	0.206	0.142	0.104	0.199	0.123	0.098	0.202	0.113	0.095	0.203	0.106	0.094	0.203
n=120	(0.034)	(0.02)	(0.061)	(0.024)	(0.013)	(0.045)	(0.016)	(0.009)	(0.029)	(0.011)	(0.006)	(0.017)	(0.009)	(0.004)	(0.012)
n=240	0.099	0.067	0.134	0.09	0.056	0.121	0.073	0.051	0.115	0.065	0.048	0.113	0.059	0.047	0.111
n=240	(0.026)	(0.011)	(0.028)	(0.016)	(0.006)	(0.019)	(0.01)	(0.004)	(0.011)	(0.006)	(0.003)	(0.007)	(0.005)	(0.002)	(0.005)
n=480	0.068	0.043	0.088	0.061	0.032	0.071	0.044	0.028	0.064	0.038	0.025	0.061	0.033	0.024	0.058
n=480	(0.022)	(0.007)	(0.014)	(0.013)	(0.003)	(0.008)	(0.008)	(0.002)	(0.005)	(0.005)	(0.001)	(0.004)	(0.003)	(0.001)	(0.002)

Table 4.3: All: AMSE (equation 4.4) over 500 replications for all cross-sectional units, i.e. $I = \{1, ..., n\}$ of model (4.1)-(4.3): d = DLRA, s = SLRA, $r = 1$, s2 = SLRA with $r = 2$. Standard errors in parentheses.

individually.

Summing up, we can cast the trade-off between SLRA and DLRA as follows. Let $C_{it} = \hat{C}_{it} + \hat{\nu}_{it}^{SLRA}$ where $\hat{C}_{it}$ is the estimate of C_{it} with SLRA and $\hat{\nu}_{it}^{SLRA}$ is the corresponding estimation error. And let $\chi_{it} = \hat{\chi}_{it} + \hat{\nu}_{it}^{DLRA}$. We have

$$\chi_{it} = C_{it} + e_{it}^{\chi} = \hat{C}_{it} + e_{it}^{\chi} + \hat{\nu}_{it}^{SLRA}$$

$$\text{so we compare} \quad (\chi_{it} - \hat{C}_{it}) = \hat{\nu}_{it}^{SLRA} + e_{it}^{\chi}$$
$$\text{and} \quad (\chi_{it} - \hat{\chi}_{it}) = \hat{\nu}_{it}^{DLRA} \; .$$

The simulation results indicate that $\hat{\nu}_{it}^{SLRA}$ has smaller variance than $\hat{\nu}_{it}^{DLRA}$ - except if the dynamics is strong and T is large relative to n. On the other hand if e_{it}^{χ} is large this can dominate the stability advantage of SLRA.

4.3 Implications for Forecasting

Although the contemporaneous influence of weak factors might be important for *individual* cross-sectional units, we know that it vanishes under static aggregation and is therefore small "on average". On the other hand, this situation might change when looking at the contribution of weak factors to subsequent periods. The potential gains of considering weak factors for forecasting becomes already apparent in the extreme examples 3.1.3, 3.1.4. In this section we further investigate the role of weak factors in forecasting within the state space framework. Consider again the simple model from section 4.2. The population projection of

$y_{i,t+1}$ on the infinite past of all variables $\mathbb{H}_t(y)$ is given by:

$$\begin{aligned}
\text{proj}(y_{i,t+1} \mid \mathbb{H}_t(y)) &= \text{proj}(\chi_{i,t+1} + \xi_{i,t+1} \mid \mathbb{H}_t(\varepsilon) \oplus \mathbb{H}_t(\xi)) = \text{proj}(\chi_{i,t+1} \mid \mathbb{H}_t(\varepsilon)) \\
&= \text{proj}(\Lambda_{i,s} F^s_{t+1} + \Lambda_{i,w} F^w_{t+1} \mid \text{sp}(F^s_t) \oplus \text{sp}(F^w_t)) \\
&= \text{proj}(\Lambda_{i,s}(M_{ss} F^s_t + M_{sw} F^w_t + G_s \varepsilon_{t+1}) \\
&\qquad + \Lambda_{i,w}(M_{ws} F^s_t + M_{ww} F^w_t + G_w \varepsilon_{t+1}) \mid \text{sp}(F^s_t) \oplus \text{sp}(F^w_t)) \\
&= (\Lambda_{i,s} M_{ss} + \Lambda_{i,w} M_{ws}) F^s_t + (\Lambda_{i,s} M_{sw} + \Lambda_{i,w} M_{ww}) F^w_t \;, \quad (4.5)
\end{aligned}$$

where we used that (ξ_{it}) is orthogonal to (χ_{it}) for all leads and lags, that $\xi_{it} \perp \xi_{js}$ for all i, j and $t \neq s$ and $\mathbb{E} F^s_t F^{w\prime}_t = 0$. Equation (4.5) reveals the following: When considering factors for a forecasting regression model, it is not vital whether a factor is strong or weak, but rather how it enters the *dynamics of the common component*. For most i, we expect $\Lambda_{i,s}$ to be "large". So "in most cases" the weak factors enter the one-step ahead prediction via M_{sw}. If M_{sw} is large - even if $\Lambda_{i,w} = 0$ - part of the variation of $y_{i,t+1}$ is explained by the weak factor F^w_t. Clearly, considering the system in (3.14)(3.15), the remainder state dimensions x^r_t influence χ_{it} not contemporaneously but in a time lag. In this sense they are "super weak factors" and can improve forecasting even more.
The factor augmented auto-regression suggested in the seminal work of (see Stock and Watson, 2002a,b; Bai and Ng, 2006) is probably the most common method for forecasting with factor models:

$$y_{i,t+h} = \beta F^s_t + \underset{\sim}{\alpha}(L) y_{i,t} + \nu_{t+h} \quad \text{for} \;\; h \geq 1 \;, \qquad (4.6)$$

where $\underset{\sim}{\alpha}(L)$ is a lag-polynomial to incorporate lags of the output variable in order to account for individual dynamics.
To relate (4.6) to (4.5), consider for example the population projection

$$\begin{aligned}
\text{proj}(y_{i,t+1} \mid \text{sp}(F^s_t, y_{it})) &= \text{proj}(y_{i,t+1} \mid \text{sp}(F^s_t) \oplus \text{sp}(e_{it})) \\
&= (\Lambda_{i,s} M_{ss} + \Lambda_{i,w} M_{ws}) F^s_t \\
&\qquad + (\Lambda_{i,s} M_{sw} + \Lambda_{i,w} M_{ww}) \,\text{proj}(F^w_t \mid \text{sp}(e_{it})) \;, \quad (4.7)
\end{aligned}$$

which has a larger population forecasting error due to the fact that we project F_t^w on $e_{it} = \Lambda_{i,w} F_t^w + \xi_{it}$, which is a linear combination of F_t^w contaminated with "noise", rather than on F_t^w itself as in equation (4.5). Of course, if we add further lags of y_{it} to the projection in (4.7) the prediction error can be reduced and we obtain more complicated calculations but the rationale stays the same. What the potential gains are from including weak factors in a forecasting regression is ultimately an empirical question and varies from unit to unit. Though model (4.6) is probably mostly used in practice, (Stock and Watson, 2002b, see equation (2.5)) already suggested to include also lags of F_t^s into the forecasting model which is quite anticipatory in light of our discussion above: Given the strong factors follow a VAR system and the conditions of theorem 3.4.1 are satisfied if that VAR system is put into state space representation, including lags of F_t^s in the forecasting regression is equivalent to including weak factors.
Next, we investigate these thoughts in a Monte-Carlo simulation: For this consider three competing models:

(i) "b": regressing $y_{i,t+1}$ on both, i.e. F_t^s and F_t^w,

(ii) "S": regressing $y_{i,t+1}$ only on the strong factors F_t^s,

(iii) "SW": the Stock and Watson (2002a) method, i.e. regressing $y_{i,t+1}$ on the strong factors and $y_{i,t}$ plus lags, where lag order is selected for each regression individually by AIC.

We draw 500 replications from the data generating process presented in section 4.2 and evaluate Mean Squared Forecast Error (MSFE) performance defined as

$$MSFE := n^{-1} \sum_{i=1}^{n} (y_{i,T} - \hat{y}_{i,T})^2 ,$$

where $\hat{y}_{i,T}$ is the prediction of one of the competing methods estimated with data from $t = 1, ..., T-1$. The results presented in table 4.4 demonstrate that in all circumstances the model regressing on strong *and* weak factors outperforms the others. They also indicate that including lags of the output can account partly for

	b,T=60	S,T=60	SW,T=60	b,T=120	S,T=120	SW,T=120	b,T=240	S,T=240	SW,T=240	b,T=480	S,T=480	SW,T=480	b,T=960	S,T=960	SW_960
n=30	1.815	2.298	2.103	1.701	2.247	2.015	1.666	2.178	1.964	1.721	2.266	2.058	1.66	2.297	2.016
n=30	(0.023)	(0.027)	(0.025)	(0.022)	(0.027)	(0.024)	(0.02)	(0.025)	(0.023)	(0.022)	(0.027)	(0.025)	(0.021)	(0.027)	(0.024)
n=60	1.634	2.179	1.909	1.507	2.21	1.863	1.488	2.126	1.835	1.477	2.187	1.859	1.443	2.114	1.811
n=60	(0.015)	(0.018)	(0.016)	(0.013)	(0.018)	(0.016)	(0.013)	(0.018)	(0.016)	(0.013)	(0.018)	(0.016)	(0.013)	(0.017)	(0.015)
n=120	1.429	1.992	1.718	1.38	2.117	1.73	1.363	2.08	1.731	1.3	2.007	1.667	1.309	2.027	1.702
n=120	(0.009)	(0.012)	(0.01)	(0.008)	(0.012)	(0.01)	(0.008)	(0.012)	(0.01)	(0.008)	(0.012)	(0.01)	(0.008)	(0.012)	(0.01)
n=240	1.365	2.039	1.671	1.294	2.054	1.659	1.236	1.902	1.584	1.228	2.004	1.614	1.194	2.03	1.616
n=240	(0.006)	(0.008)	(0.007)	(0.006)	(0.009)	(0.007)	(0.005)	(0.008)	(0.007)	(0.005)	(0.008)	(0.007)	(0.005)	(0.008)	(0.007)
n=480	1.306	1.953	1.613	1.247	1.992	1.573	1.211	2.021	1.603	1.197	2.026	1.608	1.169	1.942	1.568
n=480	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.006)	(0.005)

Table 4.4: MSFE for one step ahead forecasts for model (4.1), (4.3): “b” regressing on strong and weak factors, “S” regressing on strong factors only and “SW” Stock Watson method from equation (4.6). Standard errors in parentheses.

the influence of weak factors as the SW model outperforms regressing on strong factors alone.

4.4 Infinite-Dimensional Static Aggregation Space

In the literature it is commonly thought that the main advantage of DLRA/the Italian School over SLRA/the American school is the ability to deal with “infinite dimensional factor spaces” (see Forni et al., 2015). In this section we investigate this idea in more detail.

The following example is from Forni et al. (2015): Let u_t be scalar white noise and α_i be drawn independently from a uniform distribution over $(-1, 1)$ and suppose the dynamic common component is given by

$$\chi_{it} = (1 - \alpha_i L)^{-1} u_t \ . \tag{4.8}$$

In light of the discussion in section 3.2, we need to distinguish between $\overline{\mathrm{sp}}(\chi_t)$ and $\mathbb{S}_t$ being infinite dimensional. Clearly, $\overline{\mathrm{sp}}(\chi_t)$ is infinite dimensional but this does not imply that also $\mathbb{S}_t$ is infinite dimensional and a static factor model would be mis-specified as is claimed in Forni et al. (2015). Unfortunately, we were not able to prove that $\mathbb{S}_t$ is infinite dimensional. However with a small adaptation of this example using the “infinite blocks” idea, we can provide an example for $\mathbb{S}_t$ being infinite dimensional:

Example 4.4.1 (Infinite Dimensional Static Aggregation Space)

Let α_j and u_t and (χ_{it}) be as in equation (4.8) above. Define the dynamic common component $(\tilde{\chi}_{it})$ as follows:

$$\begin{pmatrix} \tilde{\chi}_{1t} \\ \hline \tilde{\chi}_{2t} \\ \tilde{\chi}_{3t} \\ \hline \tilde{\chi}_{4t} \\ \tilde{\chi}_{5t} \\ \tilde{\chi}_{6t} \\ \hline \vdots \end{pmatrix} = \begin{pmatrix} \chi_{1t} \\ \hline \chi_{1t} \\ \chi_{2t} \\ \hline \chi_{1t} \\ \chi_{2t} \\ \chi_{3t} \\ \hline \vdots \end{pmatrix} = \begin{pmatrix} (1-\alpha_1 L)^{-1} \\ \hline (1-\alpha_1 L)^{-1} \\ (1-\alpha_2 L)^{-1} \\ \hline (1-\alpha_1 L)^{-1} \\ (1-\alpha_2 L)^{-1} \\ (1-\alpha_3 L)^{-1} \\ \hline \vdots \end{pmatrix} u_t \quad \text{and} \quad y_{it} = \tilde{\chi}_{it} + \xi_{it} \ , \tag{4.9}$$

for some dynamically idiosyncratic double sequence (ξ_{it}). It is straightforward to see, that the result from Hallin and Liška (2011) holds analogously for static factor sequences: In particular, given two SFS $(y_{it}^{(1)})$ and $(y_{jt}^{(2)})$ with dimensions r_1 and r_2, the blended double sequence, say $(\tilde{y}_{it})$, is again a r-SFS with dimension $\max(r_1, r_2) \leq r \leq r_1 + r_2$.
In (4.9) we create an infinite amount of linearly independent static aggregates $F_{1t} = \chi_{1t}, F_{2t} = \chi_{2t}, \ldots$ by computing the average over the respective cross-sectional sub-sequence in $(\tilde{\chi}_{it})$. Consequently $\mathbb{S}_t$ is infinite dimensional and by theorem 2.2.7 there exists no finite r, such that $\sup_n \lambda_{r+1}(\Gamma_y^n) < \infty$. However, we still have $\chi_{it} = \operatorname{proj}(y_{it} \mid \mathbb{G}) = \operatorname{proj}(y_{it} \mid \mathbb{S}_t)$. So (y_{it}) from equation (4.9) is a 1-DFS.

From theorem 3.2.2.1 we know that $\dim \overline{\operatorname{sp}}(\chi_t) < \infty$ implies that $\dim \mathbb{S}_t < \infty$. However the converse is not true as the following example 4.4.2 illustrates.

Example 4.4.2 (1-SFS, 1-DFS, $r_\chi = \infty$)
Let (u_t) be scalar white noise, define the common component by

$$\chi_{it} = \Lambda_{i,s} u_t + \Lambda_{i,w}(1-\alpha_i L)^{-1} u_{t-1} \ , \tag{4.10}$$

while we suppose that $(\Lambda_{i,s})$ is a sequence of strong loadings, i.e. $\sum_{i=1}^{\infty} \Lambda_{i,s}^2 = \infty$ and $(\Lambda_{i,w})$ is a sequence of weak loadings, i.e. $\sum_{i=1}^{\infty} \Lambda_{i,w}^2 < \infty$ and $\Lambda_{i,w} > 0$ for all i and the α_i's are as in (4.8). We have one strong factor, so $\dim \mathbb{S}_t = 1$ and infinitely many weak factors, so $\dim \overline{\operatorname{sp}}(\chi_t) = \infty$.

Finally, even though $\mathbb{S}_t$ is infinite dimensional, it might be still approximated well by SLRA with a finite number of static aggregates. We investigate this question by a simulation study: We compute average mean squared error AMSE, see equation (4.4), for 500 replications of model (4.9) and compare DLRA and SLRA. The idiosyncratic component is drawn independently from a standard normal distribution. The number of factors is chosen for each replication separately by the IC_{p_2} criterion proposed in Bai and Ng (2002). Results are shown in table 4.5. The

	d,T=60	$\hat{r}$,T=60	s,T=60	d,T=120	$\hat{r}$, T=120	s,T=120	d,T=240	$\hat{r}$,T=240	s,T=240	d,T=480	$\hat{r}$,T=480	s,T=480	d,T=960	$\hat{r}$,T=960	s,T=960
n=30	0.17	1.514	0.218	0.13	1.714	0.181	0.103	1.862	0.154	0.086	1.964	0.138	0.076	1.996	0.131
n=30	(0.02)	(0.5)	(0.03)	(0.013)	(0.452)	(0.03)	(0.007)	(0.345)	(0.024)	(0.005)	(0.186)	(0.015)	(0.003)	(0.063)	(0.007)
n=60	0.145	1.95	0.147	0.101	1.994	0.111	0.075	2	0.094	0.059	2	0.086	0.049	2	0.083
n=60	(0.019)	(0.218)	(0.018)	(0.009)	(0.077)	(0.008)	(0.006)	(0)	(0.005)	(0.003)	(0)	(0.003)	(0.002)	(0)	(0.002)
n=120	0.138	1.994	0.128	0.091	2	0.096	0.063	2	0.082	0.046	2	0.075	0.036	2	0.071
n=120	(0.018)	(0.077)	(0.012)	(0.009)	(0)	(0.007)	(0.005)	(0)	(0.005)	(0.003)	(0)	(0.004)	(0.002)	(0)	(0.002)
n=240	0.134	2	0.118	0.086	2	0.087	0.057	2.002	0.074	0.04	2.11	0.063	0.03	2.664	0.044
n=240	(0.018)	(0)	(0.012)	(0.008)	(0)	(0.007)	(0.004)	(0.045)	(0.005)	(0.003)	(0.313)	(0.008)	(0.002)	(0.473)	(0.012)
n=480	0.132	2	0.113	0.085	2.004	0.086	0.055	2.492	0.056	0.037	3	0.031	0.027	3	0.026
n=480	(0.018)	(0)	(0.013)	(0.009)	(0.063)	(0.008)	(0.004)	(0.5)	(0.013)	(0.003)	(0)	(0.001)	(0.002)	(0)	(0.001)

Table 4.5: AMSE for 500 replications for the dynamic "d" and static "s" low rank approximations, together with the average of the number of chosen static factors "$\hat{r}$" by IC_{p_2} from Bai and Ng (2002). Standard deviation of AMSE is reported in parentheses.

results indicate that overall, the infinite dimensions in $\mathbb{S}_t$ can be approximated well by a finite number of static factors, since the performance of SLRA is not so much worse compared to DLRA. If T is small relative to n, the SLRA can outperform DLRA even in the infinite dimensional case with only 2 factors, e.g. for $T = 60$ and $n = 480$. On the other hand DLRA outperforms $SLRA$ if T is large and n is small. For both n and T being large the performance of SLRA and DLRA is very similar. Thus we may conclude that in general even if $\mathbb{S}_t$ is infinite dimensional this does not imply that we should use $DLRA$ for estimation of the common component, since in terms of AMSE we might approximate the common component well by a finite number of static factors.

4.5 Empirical Indication for Weak Factors?

Often it is implicitly assumed that $e_{it}^{\chi} = 0$ for all $i \in \mathbb{N}$, e.g. in Forni et al. (2005), assumption D, or in Forni et al. (2009), assumption 4 (b), or Bai and Ng (2006) assumption D. In this section, we aim to check whether we can find empirical indication to falsify such an assumption. The focus of this thesis is on structure theory, so we leave detailed treatment of estimation and statistical inference for weak factors for future research.

If $e^{\chi}_{it} = 0$ for all $i \in \mathbb{N}$ holds, this would imply zero-correlation between the (strong) static factors and the static idiosyncratic terms at all time lags/leads. We consider the macroeconomic time series data on the US-economy published and maintained by the Federal Reserve Bank[1]. Following McCracken and Ng (2016), we transformed the time series to stationarity and use $r = 8$ (strong) static factors. Using the estimated static factors $\hat{F}_{jt}$ for $j = 1, ..., 8$ and estimated idiosyncratic terms $\hat{e}_{i,t-h}$ we compute sample correlations

$$Corr(F_{jt}, e_{i,t-h}), \qquad i = 1, ..., n \text{ and } \ h = 1, 2, 3 \ , \tag{4.11}$$

at three time lags. Figure 4.1 shows the distribution of these correlation coefficients across i where each distribution corresponds to a j, h pair in (4.11). The vertical lines indicate the critical values for rejecting the Null hypothesis of zero correlation if the underlying data would be normally distributed. Clearly, we would need a statistical test to make a robust a statement here since factors and idiosyncratic terms are estimated and not observed data. What we can observe is that even though there is a lot of mass concentrated around zero (indicating $C_{it} = \chi_{it}$), for some (j, h) combinations we have correlations are considerably large which is a hint for the presence of weak factors/ a non-trivial weak common component.

[1]See https://research.stlouisfed.org/econ/mccracken/fred-databases/

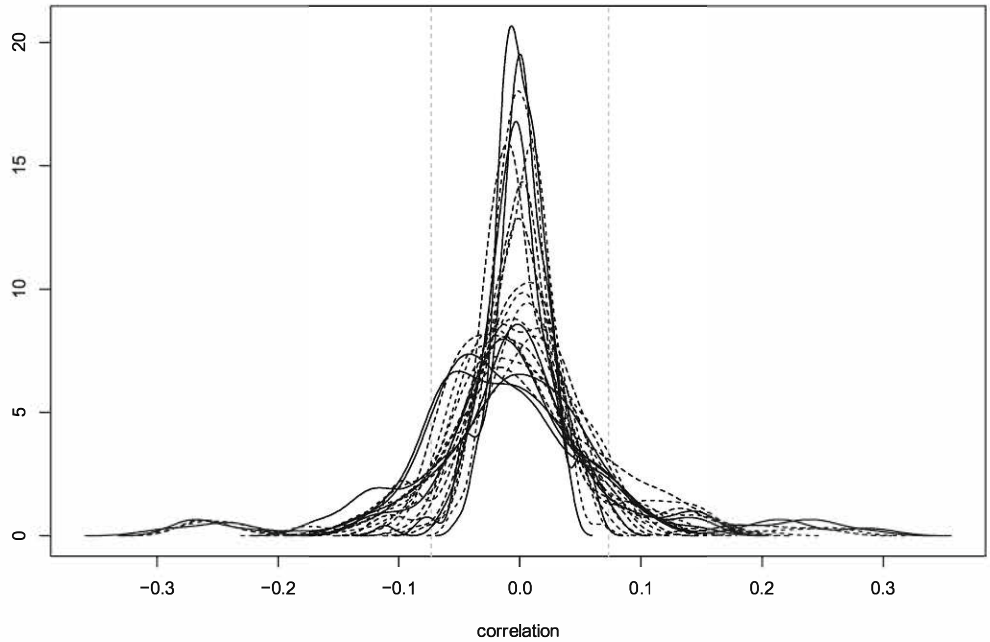

Figure 4.1: The graph shows the distribution of sample correlations of eight estimated strong static factors with 1st, 2nd and 3rd time lag of the estimated static idiosyncratic component.

Chapter 5

Summary and Outlook

In the first part, we introduced the distinction between static and dynamic factor sequences. In particular, we rephrased the results from Chamberlain and Rothschild (1983a) and re-proved them using the arguments and proof techniques from Forni and Lippi (2001) operating with variance matrices instead of spectral densities. We saw that dynamic factor sequences (DFS) emerge from dynamic aggregation, and decompose a stationary double sequence into a dynamic common and a dynamic idiosyncratic component. On the other hand, static factor sequences (SFS) emerge from static aggregation and decompose a double sequence into a static common and a static idiosyncratic component. The static aggregation space is different for every time period and contained in the dynamic aggregation space. We also proved that the static common component can be estimated consistently via static low rank approximation (SLRA). This consistency result makes static factor sequences as defined in this thesis relatable to the "static factor models" which are most commonly used in practice based on Stock and Watson (2002a); Bai and Ng (2002). The assumption for the idiosyncratic component is simplified (compared to Stock and Watson, 2002a; Bai and Ng, 2002) and embeds static factor sequences into the larger theoretical framework presented in this thesis while providing a new asymptotic framework. We also discussed heteroscedasticity and non-stationarity.

In the second part, we showed that we can reconcile both schools into one model by a decomposition of the output into three terms: The static common component, the weak common component and the dynamic idiosyncratic component. The weak common component makes the difference between the American and the Italian school and is spanned (by a potentially infinite number of) weak factors. It can or cannot be dynamically idiosyncratic but always lives in the dynamic aggregation space and is therefore associated with the common structural shocks of the double sequence/the economy. We also provided a canonical representation for double sequences which have a dynamic and a static factor structure and investigated state space representations. We showed that under certain conditions, we can identify the dynamic common component/factors from the strong static factors, and therefore obtain the dynamic common component without using frequency domain techniques.

The dynamic common component is the projection of the output on the infinite past of the common innovations/structural shocks of the economy. This is justified by the fact that under general conditions, the innovations of a purely non-deterministic dynamic common component are causally subordinated to the output as we have shown.

Consequently in the presence of weak factors, by the static common component, we only capture the part of contemporaneous co-movement but not the entire dynamic co-movement. This implies structural errors when looking e.g. at impulse responses. Furthermore we showed that weak factors can have a big influence on forecasting performance - not only for those variables influenced by weak factors but for all. In particular, the impact of weak factors for subsequent periods depends also on how important they are in the dynamics of the strong factors. This can of course vary substantially over the cross-sectional units.

Finally, we did a first examination whether we can find an empirical indication for the presence of weak factors in a large data set of macroeconomic time series of the U.S. economy. We find that a considerable amount of correlation coefficients between estimated strong static factors and time lags of estimated idiosyncratic terms are non-zero and potentially large. This is aligned to the theoretical notion of weak factors elaborated in the previous sections.

The results presented in this thesis trigger in our view a lot of questions that are to be investigated in future research:

- Estimation and inference accounting for weak factors, e.g. via SLRA.
- Empirical study - which variables are influenced by weak factors in macroeconomic data?
- Design and apply new forecasting methods incorporating weak factors.
- We have seen that not every r-SFS is a q-DFS (example 3.0.1). Can we embed the set of static factor sequences into the set of dynamic factor sequences given some regularity conditions?
- We may introduce generalised dynamic factor sequences allowing divergence of the spectral eigenvalues only on a subset of the frequency band. Is such a model empirically relevant?
- We have seen that in pathological cases we can have $q > r$ (example 3.1.3). Under which regularity conditions do we know that $q \leq r$? In particular, when are the innovations of the strong static factors the same as the innovations of the dynamic common component (see remark 4.1.2)?
- We need a procedure to determine $r_\chi(n)$, the total number of static factors.
- Can we use the ideas from section 2.2.6 to provide a new estimation theory for factor models in the non-stationary case?

Appendix A

Auxiliary Results

Proof of Theorem 2.2.6. 1. $\Rightarrow$ 2.: Assume that $\lambda_1(\Gamma_z^n) \to \infty$ for $n \to \infty$, then

$$\left(\hat{c}^{(k)} = \frac{1}{\sqrt{\lambda_1(\Gamma_z^k)}}(p_{1,k}, 0, \dots)\right)_k \in \mathcal{S},$$

where $p_{1,k}$ is the first normalized eigenvector of Γ_z^k. Now, $\hat{c}^{(k)}\Gamma_z^k \left(\hat{c}^{(k)}\right)^* = 1$ for any k which contradicts the presupposition that (z_{it}) is idiosyncratic.
2. $\Rightarrow$ 1.: Suppose that $(\hat{c}^{(k)}) \in \mathcal{S}$, we then have to show that

$$\lim_k \lim_n \left(\hat{c}^{(k)}\right)^{\{n\}} \Gamma_z^n \left(\left(\hat{c}^{(k)}\right)^{\{n\}}\right)^* = 0, \tag{A.1}$$

which is equivalent to $z_t = \underline{\lim_k \lim_n} \sum_{i=1}^n \hat{c}_i^{(k)} z_{it} = 0$. Now, equation (A.1) follows from the fact that for any $a_n \in \mathbb{C}^{1\times n}$

$$a_n \Gamma_z^n a_n^* \leq \lambda_1(\Gamma_z^n) a_n a_n^* \ .$$

■

Fact A.0.1
Let $A, B \geq 0$ be Hermitian matrices of dimension n:

1. $\lambda_s(A + B) \leq \lambda_s(A) + \lambda_1(B)$, *so in particular* $\lambda_1(A + B) \leq \lambda_1(A) + \lambda_1(B)$

for all $1 \leq s \leq n$

2. $\lambda_s(A+B) \geq \lambda_s(A)$ *and* $\lambda_s(A+B) \geq \lambda_s(B)$ *for all* $1 \leq s \leq n$ *therefore*

3. *If* $A \geq B$, *we have* $\lambda_s(A-B+B) = \lambda_s(A) \geq \lambda_s(B)$ *for all* $1 \leq s \leq n$.

Fact A.0.2
For $k = 1, 2$, *suppose* (f_n) *is a sequence in* $L_k(\mathcal{P}, \mathbb{C})$. *If* $f_n \xrightarrow{L_k} f$ *for* $n \to \infty$, *then there exists a subsequence* (f_{s_n}) *of* (f_n) *that converges a.e. to* f *for* $n \to \infty$.

Proof. See (Klenke, 2020, Korrolar 6.13) . ■

Lemma A.0.3 (L^2 convergence implies convergence of cross-spectra)
Suppose $(A_{nt}), (B_{nt})$ *are sequences in* $\mathbb{H}(y)$ *with* $A_{nt} \xrightarrow{L^2} A_t \in \mathbb{H}(y)$, $B_{nt} \xrightarrow{L^2} B_t$ *then* $\operatorname{spec}(A_{nt}, B_{nt}) \to \operatorname{spec}(A_t, B_t)$ *a.e. for* $n \to \infty$.

Proof. By continuity of the inner product, we know that

$$\begin{aligned} \lim_n \langle A_{nt}, B_{nt} \rangle = \lim_n \int_{-\pi}^{\pi} \operatorname{spec}(A_{nt}, B_{nt}) &= \langle \lim_n A_{nt}, \lim_n B_{nt} \rangle \\ = \mathbb{E}\, A_t B_t &= \int_{-\pi}^{\pi} \operatorname{spec}(A_t, B_t) \ . \end{aligned}$$

Consequently, the cross-spectrum $\operatorname{spec}(A_{nt}, B_{nt})$ converges in L^1 on the frequency band to $\operatorname{spec}(A_t, B_t)$ and by Fact A.0.2, there exists a subsequence where it converges *a.e.* on $[-\pi, \pi]$. ■

Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Anderson, B. D. and Deistler, M. (2008). Generalized linear dynamic factor models-a structure theory. In *2008 47th IEEE Conference on Decision and Control*, pages 1980–1985. IEEE.

Anderson, B. D., Deistler, M., Chen, W., and Filler, A. (2012). Autoregressive models of singular spectral matrices. *Automatica*, 48(11):2843–2849.

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. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Ng, S. (2010). Instrumental variable estimation in a data rich environment. *Econometric Theory*, 26(6):1577–1606.

Bai, J. and Ng, S. (2020). Simpler proofs for approximate factor models of large dimensions. *arXiv preprint arXiv:2008.00254*.

Bai, J. and Ng, S. (2021). Approximate factor models with weaker loadings. *arXiv preprint arXiv:2109.03773*.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Barigozzi, M. (2018). Dynamic factor models. *Lecture notes. London School of Economics.*

Barigozzi, M. and Cho, H. (2020). Consistent estimation of high-dimensional factor models when the factor number is over-estimated. *Electronic Journal of Statistics.*

Barigozzi, M. and Luciani, M. (2019). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the em algorithm. *arXiv preprint arXiv:1910.03821.*

Brillinger, D. R. (2001). *Time series: data analysis and theory.* SIAM.

Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods.* Springer science & business media.

Chamberlain, G. and Rothschild, M. (1983a). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, pages 1281–1304.

Chamberlain, G. and Rothschild, M. (1983b). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, pages 1281–1304.

Chen, W., Anderson, B. D., Deistler, M., and Filler, A. (2011). Solutions of Yule-Walker equations for singular AR processes. *Journal of Time Series Analysis*, 32(5):531–538.

Deistler, M., Anderson, B. D., Filler, A., Zinner, C., and Chen, W. (2010). Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control*, 16(3):211–224.

Deistler, M., Filler, A., and Funovits, B. (2011). AR systems and AR processes: The singular case. *Communications in Information and Systems*, 11(3):225–236.

Deistler, M. and Scherrer, W. (2022). *Time Series Models*, volume Lecture Notes in Statistics 224. Springer.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2004). The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, 119(2):231–255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American statistical association*, 100(471):830–840.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of econometrics*, 185(2):359–371.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric theory*, 17(6):1113–1141.

Forni, M. and Reichlin, L. (1996). Dynamic common factors in large cross-sections. *Empirical economics*, 21(1):27–42.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic processes and their applications*, 123(7):2678–2695.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, 163(1):29–41.

Hannan, E. J. and Deistler, M. (2012). *The statistical theory of linear systems.* SIAM.

Jolliffe, I. T. (2002). *Principal component analysis.* Springer.

Kailath, T. (1980). *Linear systems*, volume 156. Prentice-Hall Englewood Cliffs, NJ.

Klenke, A. (2020). Wahrscheinlichkeitstheorie. 4th edt.

Lancaster, P. and Tismenetsky, M. (1985). *The theory of matrices: with applications.* Elsevier.

Lippi, M. (2021). Validating DSGE models with SVARS and high-dimensional dynamic factor models. *Econometric Theory*, pages 1–19.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Ruiz, E., Poncela, P., et al. (2022). Factor extraction in dynamic factor models: Kalman filter versus principal components. *Foundations and Trends® in Econometrics*, 12(2):121–231.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In *The Oxford Handbook of Economic Forecasting*. Oxford University Press.