



Diplomarbeit

Verallgemeinerte dynamische Faktormodelle

Darstellung und Vergleich zweier Schätzmethoden

Ausgeführt am
Institut für Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von Herrn
O. Univ.Prof. Dipl.-Ing. Dr.Techn. Manfred Deistler
durch

Alexander Filler
Weißes-Kreuzgasse 64/12
2340 Mödling

Mödling, am 20.08.2007

.....

Inhaltsverzeichnis

1	Einleitung	1
1.1	Fehler-in-den-Variablen Modelle	3
1.2	Faktormodelle	5
1.2.1	Klassische Faktormodelle	6
1.2.2	Verallgemeinerte Faktormodelle	6
2	Grundlagen	8
2.1	Matrizen	8
2.1.1	Eigenwerte und -vektoren	8
2.1.2	Verallgemeinerte Eigenwerte und -vektoren	8
2.1.3	Matrizenklassen	10
2.1.4	Normale Matrizen	10
2.1.5	Halbordnung von Matrizen	12
2.2	Konvergenzaussagen	13
2.3	Schätzer für Varianz-Kovarianz Matrizen bzw. spektrale Dichten . . .	14
2.4	Principal Components	17
2.4.1	PCA im statischen Fall	17
2.4.2	PCA im dynamischen Fall	18
3	Darstellungstheorie	19
3.1	Grundlagen	19
3.2	Faktorstruktur	21
4	Schätzung und Vorhersage des verallgemeinerten dynamischen Faktor-	28
	modells	
4.1	Die Methode von Forni, Hallin, Lippi und Reichlin	29
4.1.1	Das Modell	29
4.1.2	Der Schätzer	34
4.2	Die Methode von Stock und Watson	40

4.2.1	Das Modell	40
4.2.2	Der Schätzer	45
4.3	Vergleich der beiden Methoden	50
5	Monte Carlo Simulation	53
6	Appendix A	77
7	Appendix B	82
8	R Code	90

Danksagung

Mein spezieller Dank richtet sich an Professor Deistler. Seine Anregungen und Diskussionen zu diesem Thema waren von unschätzbarem Wert, und ermöglichten mir, aufgeworfene Problemstellungen zu verstehen und zu lösen.

Bedanken möchte ich mich auch bei Professor Lippi, welcher mir bei seinem Besuch in Wien die Feinheiten der von ihm mitentwickelten Methode erklärt hat.

Ein aufrichtiges Dankeschön gilt meinen Eltern. Durch ihre volle Unterstützung war es mir möglich, mein Studium so durchzuziehen, wie ich es für richtig hielt.

Kapitel 1

Einleitung

Die Idee von Faktormodellen ist ziemlich simpel. Das Ziel ist es, hochdimensionale Prozesse durch wesentlich niedrigdimensionalere zu erklären. Das heißt, man geht davon aus, dass die vielen verschiedenen Komponenten des beobachteten Prozesses von wenigen gemeinsamen Komponenten, die Faktoren genannt werden, getrieben werden. Beispielsweise könnte man in den beobachtbaren Prozess sämtliche volkswirtschaftlichen Kenngrößen verschiedener Staaten stecken, von denen man aus markoökonomischer Sicht erwarten kann, dass sie gemeinsamen Trends und Schwingungen folgen. Der „Rest“, also der Fehler, der nicht durch die gemeinsamen Faktoren erklärt wird, kann als länder- sektor- regionalspezifische Abweichung interpretiert werden.

Ebenso können finanzwirtschaftliche Daten einem Faktormodell folgen. Seien etwa die Komponenten des hochdimensionalen Prozesses sämtliche Assets eines Index, so liegt die Vermutung nahe, dass sämtliche Assets, manche mehr, manche weniger, einem Markttrend folgen, welcher als Faktor interpretiert wird. Die Fehler wären dann die spezifischen Abweichungen der Assets, bedingt durch Schocks, Spekulationen etc.

Klarerweise gibt es nicht „DAS“ Faktormodell schlechthin, sondern man muss auf die Gegebenheiten Rücksicht nehmen.

So kann man zum einen zwischen statischen und dynamischen Faktormodellen unterscheiden. Ein dynamischer Ansatz entspricht vielleicht des öfteren eher der Realität, doch stellt sich immer die Frage, wie lange die Vergangenheit auf jetzige und zukünftige Werte Einfluss hat und dementsprechend berücksichtigt wird. Unter entsprechenden Annahmen kann man ein dynamisches Modell auch in ein statisches umschreiben, womit man die Vorteile beider Ansätze (Dynamik ist realistisch, Statik vereinfacht die Theorie und den Rechenaufwand) miteinander verbindet.

Zum anderen ist die Frage der Struktur der Fehler von wesentlicher Bedeutung. Geht man davon aus, dass die Fehler untereinander unabhängig sind, oder erlaubt man ihnen

gewisse Abhängigkeiten?

Verallgemeinerte dynamische Faktormodelle gehen wie im Namen bereits erwähnt von einem dynamischen Ansatz aus. Dynamische Faktormodelle mit unabhängigen Fehlern werden auch als Frisch-Modelle¹ bezeichnet, und sind bereits viel diskutiert worden, da es den Ansatz bereits seit langem gibt. Die Charakteristik der Frisch-Modelle, die Unabhängigkeit der Fehler, hat jedoch einen schalen Beigeschmack, da vor allem in makroökonomischen und finanzwirtschaftlichen Modellen die Unabhängigkeit nicht wirklich gewährleistet werden kann, viel mehr die Korrelation so gut wie nicht wegzudenken ist.

Das verallgemeinerte dynamische Faktormodell wird somit durch das Fallenlassen der Unabhängigkeitsannahme der Fehler charakterisiert.

Im weiteren Verlauf des ersten Kapitels möchte ich die Äquivalenz der Faktormodelle zu den allgemein gehaltenen Fehler-in-den-Variablen Modellen aufzeigen.

Das zweite Kapitel beinhaltet grundlegende Definitionen und Sätze aus der linearen Algebra und der Zeitreihenanalyse, welche aber nur als Resultate präsentiert werden.

In Kapitel 3 präsentiere ich die Resultate von Chamberlain und Rothschild aus dem Jahr 1983, welche bereits damals Bedingungen formuliert haben, um die Präsenz eines Faktormodells gewährleisten zu können. Das von ihnen beschriebene Modell erlaubte bereits Abhängigkeiten der Fehler im Querschnitt, auf zeitliche Abhängigkeiten wurde aber nicht eingegangen.

Kapitel 4 stellt zwei Methoden, von Forni, Hallin, Lippi und Reichlin sowie von Stock und Watson, zur konsistenten Schätzung des Faktorraumes und anschließenden Prognose des verallgemeinerten Faktormodells, vor.

Um die beiden Methoden vergleichen zu können, werden in Kapitel 5 abschließend die Ergebnisse einer Monte Carlo Simulation präsentiert.

¹siehe [13] oder [18]

1.1 Fehler-in-den-Variablen Modelle

Das Konzept von linearen dynamischen Fehler-in-den-Variablen Modellen (im Folgenden FiV-Modelle genannt) sieht folgender Maßen aus:

Der wahre aber nicht beobachtbare, n-dimensionale Prozess χ unterliegt gewissen Restriktionen:

$$\omega(z) \chi_t = 0 \quad (1.1)$$

wobei z den Backshiftoperator darstellt und für die Relationsfunktion $\omega(z)$ Folgendes gilt:

$$\omega(z) = \sum_{j=-\infty}^{\infty} \omega_j z^j; \quad \omega_j \in \mathbb{R}^{m \times n}; \quad \sum_{j=-\infty}^{\infty} \|\omega_j\| < \infty \quad (1.2)$$

und $\|\cdot\|$ eine beliebige Matrizennorm auf $\mathbb{R}^{m \times n}$ ist.

Ohne Beschränkung der Allgemeinheit wird angenommen, dass die Zeilen von ω linear unabhängig sind und $m \leq n$ ist. Desweiteren möchte ich mich auf den Fall, dass $\omega(z)$ rational ist beschränken.

Der tatsächlich beobachtbare Prozess x_t setzt sich dann aus den latenten Variablen χ_t und den Messfehlern ε_t zusammen:

$$x_t = \chi_t + \varepsilon_t \quad (1.3)$$

wobei folgende Annahmen gelten:

- a.1 x_t, χ_t und ε_t sind (schwach) stationäre Prozesse mit absolut aufsummierbaren Kovarianzfunktionen und spektralen Dichten $\Sigma^x(\lambda)$, $\Sigma^\chi(\lambda)$ und $\Sigma^\varepsilon(\lambda)$
- a.2 $\mathbb{E} \chi_t = \mathbb{E} \varepsilon_t = 0$
- a.3 $\mathbb{E} \chi_t \varepsilon_s = 0, \forall s, t \in \mathbb{Z}$
- a.4 $\Sigma^x(\lambda) > 0$

a.1 ermöglicht erst die Verwendung sämtlicher Sätze der Spektraltheorie stationärer Prozesse.

a.2 ist keine Einschränkung, sondern lediglich eine Festlegung auf zentrierte Prozesse.

a.3 hingegen ist existenziell wichtig für die Theorie der FiV-Modelle (nicht nur hier), da ohne diese Orthogonalität eine sinnvolle Zerlegung in einen latenten und einen Fehlerterm kaum möglich ist. Zum Beispiel würde jede singuläre Matrix $V \in \mathbb{R}^{n \times n}$ folgende Aufteilung ermöglichen: $x_t = V x_t + (I - V) x_t$ mit latentem Teil $\chi = V x_t$, welcher mit einer Relationsfunktion ω (z.B. Orthogonalkomplement von V) (1.1) erfüllt. a.4 ist keine einschneidende Einschränkung.

Eine Konsequenz von a.3 ist, dass die Kovarianzfunktion von x_t als Summe der Kovarianzfunktionen von χ_t und ε_t geschrieben werden kann:

$$\begin{aligned} \Gamma_s^x &= \mathbb{E} x_{t+s} x_t^* \text{ }^2 \\ &= \mathbb{E} (\chi_{t+s} + \varepsilon_{t+s})(\chi_t + \varepsilon_t)^* \\ &= \mathbb{E} \chi_{t+s} \chi_t^* + \mathbb{E} \varepsilon_{t+s} \varepsilon_t^* \\ &= \Gamma_s^\chi + \Gamma_s^\varepsilon \end{aligned}$$

Natürlich gilt dadurch auch für die spektrale Dichte die Aufspaltung:

$$\begin{aligned} \Sigma^x(\lambda) &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \Gamma_s^x e^{-i\lambda s} \\ &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} (\Gamma_s^\chi + \Gamma_s^\varepsilon) e^{-i\lambda s} = \Sigma^\chi(\lambda) + \Sigma^\varepsilon(\lambda) \end{aligned}$$

Beachtet man nun noch, dass die Voraussetzung $\omega(z)\chi = 0$ äquivalent ist dazu, dass $\omega(e^{-i\lambda})\Sigma^\chi(\lambda) = 0$ ³ gilt (beide Richtungen sind einfach einzusehen), bekommt man eine Formulierung der FiV-Modelle in spektraler Darstellung:

$$\begin{aligned} \Sigma^x(\lambda) &= \Sigma^\chi(\lambda) + \Sigma^\varepsilon(\lambda) \\ 0 &= \omega(e^{-i\lambda})\Sigma^\chi(\lambda) \end{aligned} \tag{1.4}$$

² * bedeutet Konjugation und Transponierung.

³ $\omega(e^{-i\lambda})$ steht in einer Eins-zu-Eins-Beziehung mit $\omega(z) : \omega(e^{-i\lambda}) = \sum_{j=-\infty}^{\infty} \omega_j e^{-i\lambda j}$; $\omega_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \omega(e^{-i\lambda}) e^{i\lambda j} d\lambda$; $\omega(e^{-i\lambda})$ muss vollen Zeilenrang haben.

1.2 Faktormodelle

Eine äquivalente Darstellungsform der FiV-Modelle sind die sogenannten Faktor-Analyse-Modelle oder kurz Faktormodelle. Hier liegt die Annahme zu Grunde, dass der wahre Prozess χ_t von einem (wesentlich) niedrigdimensionalen Faktor(-Prozess) f_t erzeugt wird, oder äquivalent dazu, im Bild einer rationalen Funktion $B(z)$ liegt.

$$\chi_t = B(z) f_t \quad (1.5)$$

Veranschaulichen kann man sich die Gleichwertigkeit der beiden Darstellungsformen folgendermaßen:

$\omega(z)$ hat vollen Zeilenrang m . Somit benötigt man $B(z)$, welche im Rechtskern von $\omega(z)$ liegt, vollen Spaltenrang $n-m$ hat und eine Darstellung besitzt

$$B(z) = \sum_{j=-\infty}^{\infty} b_j z^j, b_j \in \mathbb{R}^{n \times (n-m)}$$

mit $\sum_{j=-\infty}^{\infty} \|b_j\| < \infty$. Das heißt, dass $B(z)$ keine Pole am Einheitskreis $|z| = 1$ besitzt, dies dadurch auch für $B(z)^*$ gilt, und somit $f_t = B(z)^* \chi_t$ definiert werden kann.

So ein $B(z)$ ⁴ existiert:

Betrachtet man $\omega(z)$ als Matrix über dem Körper der rationalen Funktionen, so kann man $n-m$ linear unabhängige Vektoren finden damit $\begin{pmatrix} \omega(z) \\ B(z)^* \end{pmatrix}$ vollen Rang n hat.

Durch das Gram-Schmidt-Verfahren kann man nun $B(z)^*$ dermaßen umformen, so dass $B(z)$ tatsächlich im Rechtskern von $\omega(z)$ liegt. Falls nun ein Element $b_{ij}(z)$ der Matrix $B(z)$ einen k -fachen Pol bei z_i auf dem Einheitskreis besitzt, kann man die entsprechende Spalte von $B(z)$ mit $(z - z_i)^k$ multiplizieren, womit die neue Matrix $B(z)$ keinen Pol am Einheitskreis besitzt und dennoch im Rechtskern von $\omega(z)$ liegt.

Somit gilt

$$\chi_t = (\omega(z)^* \quad B(z)) \begin{pmatrix} \omega(z) \\ B(z)^* \end{pmatrix} \chi_t = (\omega(z)^* \quad B(z)) \begin{pmatrix} 0 \\ f_t \end{pmatrix} = B(z) f_t = \chi_t$$

Die Orthogonalität von χ_t auf ε_t wird durch $B(z)$ auf f_t übertragen (und anschließend natürlich wieder zurück), wodurch die Orthogonalität von χ_t auf ε_t bestehen bleibt,

⁴ $B(z)$ wird auch als Transferfunktion bezeichnet.

und dadurch die Annahmen a.1 - a.4 noch immer erfüllt sind. Das äquivalente Analogon zu (1.4) lautet somit:

$$\begin{aligned}\Sigma^x(\lambda) &= \Sigma^x(\lambda) + \Sigma^\varepsilon(\lambda) \\ \Sigma^x(\lambda) &= B(e^{-i\lambda})\Sigma^f(\lambda)B(e^{-i\lambda})^*\end{aligned}\tag{1.6}$$

Es ist wert daraufhinzuweisen, dass ohne Beschränkung der Allgemeinheit f_t als White-Noise angenommen werden kann, da für rationale spektrale Dichten $\Sigma^x(\lambda)$ immer eine Darstellungsform $\Sigma^x(\lambda) = B(e^{-i\lambda})IB(e^{-i\lambda})^*$ möglich ist, und dadurch der Faktorprozess f_t eine Kovarianzmatrix $2\pi I$ der Größe $(n - m) \times (n - m)$ besitzt.

Der entscheidende Punkt der Faktormodelle (oder FiV-Modelle) ist nun, die Anzahl der Möglichkeiten für $B(z)$ (bzw. $\omega(z)$) so klein wie möglich zu halten. Dazu benötigt es jedoch weitere, teils sehr restriktive Bedingungen an die Fehlerstruktur des Modells.

1.2.1 Klassische Faktormodelle

Bei klassischen Faktormodellen verfolgt man den Ansatz, dass die Fehler zu jedem Zeitpunkt unkorreliert sind. Das ist gleichbedeutend dazu, dass das Fehlerspektrum Diagonalgestalt hat.

Spezielles Interesse gilt hier dem Frisch-Korng einer Matrix $\text{mc}(\Sigma^x)$, welcher der maximale Korng aller unter Σ^x möglichen spektralen Dichten Σ^x ist.⁵ In anderen Worten entspricht $(n - \text{mc}(\Sigma^x))$ der Anzahl der Faktoren, mit denen die diagonale Fehlerstruktur noch erhalten bleibt.⁶

Da die Annahme der unkorrelierten Fehler jedoch für viele Anwendungen nicht angemessen ist (z.b. in makroökonomischen Anwendungen), wird diese immer häufiger fallengelassen und eine andere Vorgehensweise bevorzugt:

1.2.2 Verallgemeinerte Faktormodelle

lassen Abhängigkeiten der Fehler sowohl im Querschnitt als auch über die Zeit zu. Diese Zusammenhänge müssen jedoch exakt präzisiert werden, um eine Identifikation in einen latenten und einen Fehlerterm zu gewährleisten.

⁵Es muss für jede Frequenz λ gelten, deswegen steht auch Σ^x statt $\Sigma^x(\lambda)$.

⁶Zur Erinnerung: die Anzahl der Faktoren war stets $(n - m)$, d.h. $\text{mc}(\Sigma^x)$ ist der Rang der Matrix $\omega(z)$; $\omega(e^{-i\lambda})\Sigma^x(\lambda) = 0$.

Natürlich liegt auch hier das Hauptaugenmerk auf der Bestimmung der Anzahl der Faktoren, doch anders als bei klassischen Faktormodellen sucht man nicht nach der minimalen Anzahl, um ein diagonales Fehlerspektrum zu haben, sondern man möchte die Größenordnung der Fehler möglichst klein im Verhältnis zu den latenten Variablen halten. Dies geschieht meistens mit Hilfe der Varianz-Kovarianz-Matrix der Fehler, zum Beispiel durch die Bedingung der Beschränktheit ihrer Eigenwerte. Das heißt, die Varianz der Linearkombination des Fehlerprozesses ist stets wesentlich kleiner als die des latenten, sie soll sogar durch die unbeschränkte Hinzunahme neuer Assets in den beobachtbaren Prozess x_t weggemittelt werden.

Kapitel 2

Grundlagen

2.1 Matrizen

2.1.1 Eigenwerte und -vektoren

Definition 2.1.1. Ein Skalar $\lambda_i \in K$ heißt Eigenwert einer Matrix $A \in K^{n \times n}$, falls er die Bedingung

$$\begin{aligned} |A - \lambda_i I_n| &= 0 \\ \Leftrightarrow \exists x_i \in K^{n \times 1} : (A - \lambda_i I_n)x_i &= 0 \end{aligned}$$

erfüllt. x_i ist der i -te Eigenvektor der Matrix A .

Lemma 2.1.2. Sei $B \in K^{n \times n}$ positiv semi-definit, dann gilt

$$\lambda_n \leq \frac{x' B x}{x' x} \leq \lambda_1,$$

wobei λ_1 der größte und λ_n der kleinste Eigenwert von B sind.

2.1.2 Verallgemeinerte Eigenwerte und -vektoren

Definition 2.1.3. Zu zwei Matrizen A und $B \in K^{n \times n}$, mit B regulär, heißen die Skalare $\lambda \in K$, für die

$$|A - \lambda B| = 0$$

gilt, die verallgemeinerten Eigenwerte von (A, B) .

Satz 2.1.4. Seien zwei Matrizen A und $B \in \mathbb{R}^{n \times n}$, mit A positiv semi-definit und B positiv definit gegeben, dann existiert eine reguläre Matrix F , sodass

$$F'AF = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} = \Lambda$$

$$F'BF = I$$

mit $\lambda_i \geq \lambda_{i+1} \geq 0$.

Bemerkung1 Für jedes λ_i existiert ein Vektor x_i , sodass

$$(A - \lambda_i B)x_i = 0$$

und

$$x_i' B x_i = 1$$

gelten. Sind weiters alle Eigenwerte verschieden, dann gilt $x_j' A x_i = 0$ und $x_j' B x_i = 0, \forall i \neq j$, und F ist dann (x_1, \dots, x_n) . Tritt ein Eigenwert mit Häufigkeit m auf, kann man m beliebige, unabhängige Vektoren des zugehörigen Eigenraumes verwenden.

Bemerkung2 Λ enthält sämtliche verallgemeinerten Eigenwerte vom Matrizenpaar (A,B), da

$$\begin{aligned} 0 &= \det(F'AF) - \lambda \det(F'BF) \\ &= \det(F'(A - \lambda B)F) \\ &= \det(F') \det(A - \lambda B) \det(F) \\ &= {}^1 \det(A - \lambda B) \end{aligned}$$

Korollar 2.1.5. Seien $A, B \in K^{n \times n}$ mit A positiv semi-definit und B positiv definit, dann gilt

$$\lambda_n \leq \frac{x'Ax}{x'Bx} \leq \lambda_1,$$

wobei λ_1 der größte und λ_n der kleinste verallgemeinerte Eigenwert von (A,B) sind.

¹Da $\det(F) \neq 0$ ist.

2.1.3 Matrizenklassen

Definition 2.1.6. (Ähnlichkeit) Eine Matrix $A \in K^{n \times n}$ heißt ähnlich zu einer Matrix $B \in K^{n \times n}$, falls eine reguläre Matrix P existiert, sodass

$$B = P^{-1}AP$$

gilt.

Definition 2.1.7. (ω -orthogonale Matrizen) Eine reguläre Matrix $A \in K^{n \times n}$ heißt ω -orthogonal, falls ihre Inverse

$$A^{-1} = \omega(A)'$$

erfüllt (ω sei ein Automorphismus über dem Körper K).

Für $\omega = I$ nennt man solche Matrizen schlicht orthogonal, und für $K = \mathbb{C}$, $\omega = (\bar{})$ werden solche Matrizen als **unitär** bezeichnet.

Definition 2.1.8. Eine Matrix $A \in K^{n \times n}$ heißt ω -orthogonal-ähnlich zu einer Matrix $B \in K^{n \times n}$, falls eine ω -orthogonale Matrix P existiert, sodass

$$B = P^{-1}AP = \omega(P)'AP$$

gilt.

2.1.4 Normale Matrizen

Definition 2.1.9. Eine Matrix $A \in K^{n \times n}$ heißt ω -normal, wenn die Bedingung

$$\omega(A)'A = A\omega(A)', \quad \omega \in \text{Aut}(K)$$

erfüllt ist.

Für $K = \mathbb{R}$, und $\omega = I$ bzw. $K = \mathbb{C}$, und $\omega = (\bar{})$ spricht man von einer **normalen** bzw. konjugiert-normalen Matrix.

Folgende Matrizen-Klassen sind Beispiele (konjugiert-)normaler Matrizen:

- unitäre Matrizen, für die laut Definition 2.1.7 Folgendes gilt:

$$A^*A = AA^* = I_n$$

- symmetrische bzw. hermitesche Matrizen, für die laut Definition 2.1.10 Folgendes gilt:

$$A^*A = AA = AA^*$$

- schiefsymmetrische bzw. schiefhermitesche Matrizen, für die laut Definition 2.1.10 Folgendes gilt:

$$A^*A = (-A)A = A(-A) = AA^*$$

Definition 2.1.10. Eine Matrix $A \in K^{n \times n}$ heißt ω -symmetrisch bzw. ω -schiefsymmetrisch, falls

$$A = \omega(A') \text{ bzw. } A = -\omega(A')$$

gilt.

Für $K = \mathbb{R}$, und $\omega = I$ bzw. $K = \mathbb{C}$, und $\omega = (\bar{})$ spricht man von einer symmetrischen / schiefsymmetrischen bzw. hermiteschen / schiefhermiteschen Matrix.

In weiterer Folge wird $K = \mathbb{C}$, und $\omega = (\bar{})$ angenommen. Desweiteren sollen auch konjugiert-normale Matrizen als normale Matrizen bezeichnet werden.

Satz 2.1.11. (Spektralsatz normaler Matrizen) Eine Matrix $A \in \mathbb{C}^{n \times n}$ ist genau dann normal, falls es eine Orthonormalbasis aus Eigenvektoren von A gibt. Das heißt, dass A unitär-ähnlich zu einer Diagonalmatrix ist:

$$A = U'DU, \quad U \dots \text{unitär}, \quad D \dots \text{diagonal}$$

Für reelle Matrizen gilt die etwas abgeschwächtere Form:

Eine Matrix $B \in \mathbb{R}^{n \times n}$ ist genau dann normal, falls sie orthogonal-ähnlich zu einer Jordan-Normalform ist, welche aus einer Diagonalmatrix und aus reellen 2×2 Jordan-Matrizen aufgebaut ist. Das heißt,

$$B = O'JO, \quad O \dots \text{orthogonal}, \quad J \dots \text{reell erweiterte Diagonalform.}$$

Für hermitesche bzw. symmetrische Matrizen gilt sogar:

Satz 2.1.12. (Spektralsatz hermitescher Matrizen) Eine Matrix $A \in \mathbb{C}^{n \times n}$ ist genau dann hermitesch bzw. schiefhermitesch, falls sie unitär-ähnlich zu einer Diagonalmatrix mit **rein** reellen bzw. imaginären Einträgen ist:

$$A = U'DU, \quad U \dots \text{unitär}, \quad D \dots \text{diagonal, rein reell bzw. imaginär}$$

Satz 2.1.13. (Spektralsatz symmetrischer Matrizen) *Eine Matrix $B \in \mathbb{R}^{n \times n}$ ist genau dann symmetrisch, falls sie orthogonal-ähnlich zu einer Diagonalmatrix ist:*

$$B = O'DO, \quad O \dots \text{orthogonal}, D \dots \text{diagonal, reell}$$

Es ist zu beachten, dass die Eigenwerte (welche immer in den Diagonal-Matrizen stehen) der jeweiligen Matrizen keinesfalls unterschiedlich sein müssen.

2.1.5 Halbordnung von Matrizen

Die Ordnung von Matrizen ist alles andere als trivial und benötigt folglich eine Einschränkung der Menge der zu vergleichenden Matrizen. Die folgende Halbordnung ist jedoch recht hilfreich, da in den nachkommenden Sätzen gerade die Ordnung von hermiteschen Matrizen erforderlich ist.

Definition 2.1.14. *Seien A, B zwei hermitesche Matrizen, dann sagen wir, dass $A \geq B$ ist, wenn $A - B$ positiv semi-definit ($A - B \geq 0$) ist.*

Die Ordnungsrelation $A - B \geq 0$ ist reflexiv ($A \geq A$), transitiv ($A \geq B \wedge B \geq C \Rightarrow A \geq C$) und antisymmetrisch ($A \geq B \wedge B \geq A \Rightarrow A = B$) und somit eine Halbordnung. Sie ist jedoch keine Totalordnung, da hierfür für 2 beliebige hermitesche Matrizen A, B stets entweder $A \geq B$ oder $B \geq A$ gelten müsste, und dies das einfache Beispiel widerlegt:

$$\begin{aligned} A &= \begin{pmatrix} - & 1 & 0 \\ & 0 & - & 1 \end{pmatrix}, & B &= \begin{pmatrix} & 2 & - & 2 \\ - & 2 & & 2 \end{pmatrix} \\ A - B &= \begin{pmatrix} - & 3 & & 2 \\ & 2 & - & 3 \end{pmatrix}, & B - A &= \begin{pmatrix} & 1 & - & 2 \\ - & 2 & & 1 \end{pmatrix} \\ (1, 1)(A - B)(1, 1)' &= -2 & (1, 0)(B - A)(1, 0)' &= -3 \end{aligned}$$

Die obige Relation impliziert auch folgende Beziehungen:

$$\begin{aligned} \det(A) &\geq \det(B), \\ \text{sp}(A) &\geq \text{sp}(B) \text{ und} \\ \mu_i(A) &\geq \mu_i(A) \end{aligned}$$

Die letzte Beziehung ist Teil des Resultates des folgenden Satzes.

Satz 2.1.15. Seien $H_1, H_2 \in \mathbb{C}^{n \times n}$ hermitesche Matrizen, für die $H_1 \geq H_2$ gilt. Weiters seien $\mu_1^k \geq \dots \geq \mu_n^k, k = 1, 2$ die Eigenwerte der Matrizen H_1, H_2 und der Rang der Matrix $(H_1 - H_2)$ gleich r , dann gilt

1. $\mu_i^1 \geq \mu_i^2$ für $i = 1, \dots, n$ und
2. $\mu_{i+r}^1 \leq \mu_i^2$ für $i = 1, \dots, n - r$

Beweis (Satz 2.1.15) siehe [15] Seite 301, Theorem 1

Lemma 2.1.16. Sei $A \in \mathbb{R}^{t \times t}$ eine n -Bandmatrix

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n & 0 & \dots & & & & 0 \\ a_2 & a_1 & a_2 & \dots & a_n & 0 & \dots & & & 0 \\ a_3 & a_2 & a_1 & a_2 & \dots & a_n & 0 & \dots & & 0 \\ \vdots & \ddots & \vdots \\ a_n & \dots & & & a_1 & \dots & a_n & 0 & \dots & 0 \\ 0 & a_n & \dots & & & a_1 & \dots & a_n & \ddots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & & & & 0 & a_n & \dots & a_2 & a_1 \end{pmatrix}$$

so ist ihr größter Eigenwert beschränkt durch $a_1 + 2 \sum_{j=2}^n |a_j|$.

Beweis (Lemma 2.1.16)

Sei $x \in \mathbb{R}^t$ mit $\|x\| = 1$ und n der Einfachheit halber 2. So gilt

$$x'Ax = x'(a_1 I_t + a_2 I_t^{-1} + a_2 I_t^{+1})x = a_1 x'x + a_2 (x' I_t^{-1} x + x' I_t^{+1} x) \leq a_1 + 2|a_2|.$$

Wobei I_t^{-1} Einsen in der ersten unteren Nebendiagonalen und sonst überall Nuller hat.

□

2.2 Konvergenzaussagen

Definition 2.2.1. Sei $(x_t | t \in \mathbb{N})$ eine Folge von Zufallsvariablen. Dann konvergiert $(x_t | t \in \mathbb{N})$ im quadratischen Mittel gegen x_0 , wenn

$$\begin{aligned} \mathbb{E}x_0^* x_0 &< \infty \\ \lim_{t \rightarrow \infty} \mathbb{E}(x_t - x_0)^*(x_t - x_0) &= 0 \end{aligned}$$

$(x_0 = l.i.m_{t \rightarrow \infty} x_t)$

Definition 2.2.2. Sei $(x_t|t \in \mathbb{N})$ eine Folge von Zufallsvariablen. Dann konvergiert $(x_t|t \in \mathbb{N})$ in Wahrscheinlichkeit gegen x_0 wenn für jedes $\epsilon > 0$

$$\lim_{t \rightarrow \infty} P(|x_t - x_0| > \epsilon) = 0$$

gilt. ($x_0 = \text{plim}_{t \rightarrow \infty} x_t$)

Lemma 2.2.3. (Slutzky) Sei $(x_t|t \in \mathbb{N})$ eine Folge von n -dimensionalen Zufallsvektoren mit $\text{plim}_{t \rightarrow \infty} x_t = x_0$, und g eine reelle, in x_0 stetige Funktion $g : \mathbb{R}^n \mapsto \mathbb{R}$, dann gilt

$$\text{plim}_{t \rightarrow \infty} g(x_t) = g(\text{plim}_{t \rightarrow \infty} x_t) = g(x_0)$$

2.3 Schätzer für Varianz-Kovarianz Matrizen bzw. spektrale Dichten

Lemma 2.3.1. Sei x_t ein n -dimensionaler stochastischer Prozess mit der Darstellung

$$x_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j}$$

mit $\sum_{j=-\infty}^{\infty} |c_j(i, k)| < \infty$ für alle $i, k = 1, \dots, n$, und z_t ist ein linear unabhängiger White-Noise Prozess mit endlichen vierten Momenten, dann ist

$$\hat{\Gamma}_h = \frac{1}{T} \sum_{t=1}^{T-h} (x_{t+h} - \bar{x}_T)(x_t - \bar{x}_T)^*, \text{ für } h \geq 0 \quad (2.1)$$

bzw.

$$\hat{\Gamma}_h = \hat{\Gamma}_{-h}^*, \text{ für } h < 0$$

mit $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$, ein konsistenter Schätzer für die Varianz-Kovarianz Matrix $\Gamma_h = \mathbb{E}(x_{t+h} x_t')$.

Definition 2.3.2.

$$I(\theta_j)^T = 1/(2\pi T) \left| \sum_{t=1}^T x_t e^{-it\theta_j} \right|^2$$

heißt **Periodogramm** des n -dimensionalen Zufallsvektors x_t und ist auf den Fourierfrequenzen $\theta_j = \frac{2\pi j}{T}$, $j=0, \dots, T-1$ der Beobachtungen $[x_1, \dots, x_T]$ definiert. Eine andere Darstellungsmöglich ist folgende:

$$I(\theta_j)^T = \frac{1}{2\pi} \sum_{h=-(T-1)}^{T-1} \hat{\Gamma}_h e^{-i\theta_j h}$$

für Fourierfrequenzen $\theta_j = \frac{2\pi j}{T}$, $j=1, \dots, T-1$, wobei $\hat{\Gamma}_h$ der Schätzer der Varianz-Kovarianz Matrizen aus (Lemma 2.3.1) ist, und

$$I(0)^T = 1/T \bar{x}_T \bar{x}_T'$$

für die Frequenz 0.

Bemerkung Das Periodogramm ist die intuitive Schätzung der spektralen Dichte $\Sigma(\theta)$, da $\Sigma(\theta) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_h e^{-i\theta h}$ gilt, falls die spektrale Dichte existiert. Das Periodogramm ist auch asymptotisch erwartungstreu, doch die Konsistenz der Folge von Schätzfunktionen $(I(\theta)^T)_{T \in \mathbb{N}}$ für $\Sigma(\theta)$ kann nicht gewährleistet werden. Somit definiert man folgende Schätzer der spektralen Dichte:

Lemma 2.3.3. Sei x_t eine n -dimensionaler stochastischer Prozess mit der Darstellung

$$x_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j}$$

mit $\sum_{j=-\infty}^{\infty} |c_j(i, k)| |j|^2 < \infty$ für alle $i, k = 1, \dots, n$, z_t ein linear unabhängiger White-Noise Prozess mit endlichen vierten Momenten (genauer „cumulants“²), und $\theta_j = \frac{2\pi j}{T}$ die Fourierfrequenzen der Beobachtungen $[x_1, \dots, x_T]$, dann ist

$$\hat{\Sigma}(\theta) = \frac{2\pi}{T} \sum_{t=1}^{T-1} W_T \left(\theta - \frac{2\pi t}{T} \right) I \left(\frac{2\pi t}{T} \right)^T \quad (2.2)$$

mit

$$W_T(\alpha) = \sum_{j=-\infty}^{\infty} \frac{1}{B_T} W \left(\frac{\alpha + 2\pi j}{B_T} \right)$$

²Definition aus [4] Kap. 2.6

ein konsistenter Schätzer der spektralen Dichte $\Sigma(\theta)$, wobei $I\left(\frac{2\pi t}{T}\right)^T$ das Periodogramm aus Definition(2.3.2) ist, $0 < B_T \rightarrow 0$ und $B_T T \rightarrow \infty$ für $T \rightarrow \infty$ gelten muss und $W(\alpha)$ eine Gewichtsfunktion mit beschränkter Variation, welche

$$\begin{aligned} W(\alpha) &= W(-\alpha), \\ \int_{-\infty}^{\infty} W(\alpha) d\alpha &= 1, \\ \int_{-\infty}^{\infty} |W(\alpha)| d\alpha &< \infty \end{aligned}$$

erfüllt.

Ebenso ein konsistenter, jedoch für die Implementierung öfters praktischerer Schätzer der spektralen Dichte liefert das nächste Lemma.

Lemma 2.3.4. Sei x_t eine n -dimensionaler stochastischer Prozess mit der Darstellung

$$x_t = \sum_{j=-\infty}^{\infty} c_j z_{t-j}$$

mit $\sum_{j=-\infty}^{\infty} |c_j(i, k)| |j|^2 < \infty$ für alle $i, k = 1, \dots, n$, z_t ein linear unabhängiger White-Noise Prozess mit endlichen vierten Momenten (genauer „cumulants“³), und $\theta_j = \frac{2\pi j}{T}$ die Fourierfrequenzen der Beobachtungen $[x_1, \dots, x_T]$, dann ist

$$\hat{\Sigma}(\theta) = \frac{1}{2\pi} \sum_{h=-M_T}^{M_T} w(M_t^{-1}h) \hat{\Gamma}_h e^{-i\theta h} \quad (2.3)$$

ein konsistenter Schätzer der spektralen Dichte $\Sigma(\theta)$, falls $M_T > 0$, $M_T \rightarrow \infty$ und $\frac{M_T}{T} \rightarrow 0$ für $T \rightarrow \infty$ gilt, $\hat{\Gamma}_h$ ein konsistenter Schätzer der Varianz-Kovarianz Matrix (z.B. (2.1)) ist und w eine stückweise stetige, differenzierbare Funktion deren ersten drei Ableitungen beschränkt sind, mit

$$\begin{aligned} w(0) &= 1, \\ w(x) &= w(-x), \\ |w(x)| &\leq 1 \text{ und} \\ w(x) &= 0, \text{ für } |x| > 1 \end{aligned}$$

³Definition aus [4] Kap. 2.6

ist.

Bemerkung Spezielles Augenmerk in (2.2) bzw. (2.3) gilt der Wahl von M_T bzw. B_T und $W(\alpha)$ bzw. $w(x)$, wofür es viele Möglichkeiten gibt, wovon ein paar beispielsweise in ([5], 10.4 Seite 349 ff) besprochen werden.

2.4 Principal Components

x_t sei ein n -dimensionaler stochastischer Vektor, welcher die Bedingung a.1 und

$$\mathbb{E}x_t = 0$$

erfüllt.

So lautet das PCA-Problem

$$\min_{rk(A(z)) \leq q} \mathbb{E}(x_t - A(z)x_t)'(x_t - A(z)x_t), \quad (2.4)$$

wobei $A(z) = B(z)W(z)$ ist, mit $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$, $b_j \in \mathbb{R}^{n \times q}$, $\sum_{j=-\infty}^{\infty} \|b_j\| < \infty$, und

$$W(z) = \sum_{j=-\infty}^{\infty} w_j z^j; w_j \in \mathbb{R}^{q \times n}; \sum_{j=-\infty}^{\infty} \|w_j\| < \infty.$$

Dadurch besitzt x_t folgende Darstellung:

$$x_t = \chi_t + \varepsilon_t; \quad \text{mit} \quad \chi_t = B(z)f_t = B(z)W(z)x_t \quad (2.5)$$

2.4.1 PCA im statischen Fall

Der statische Fall zeichnet sich dadurch aus, dass die Zielfunktion (2.4) ihre Dynamik verliert, sprich $B(z) = B$, $W(z) = W$ und $A(z) = BW$ gilt und die Zielfunktion dadurch lautet:

$$\min_{rk(A) \leq q} \mathbb{E}(x_t - Ax_t)'(x_t - Ax_t). \quad (2.6)$$

Die Lösung dieses Problems ergibt

$$B = [o_1, \dots, o_q]; \quad W = B^*,$$

wobei o_i der i -te normalisierte Eigenvektor der Matrix $\Gamma_0^x = \mathbb{E}(x_t x_t') = O\Lambda O'$ ist.

Ein anderer Ansatz, um Principal Components zu charakterisieren, ist der Folgende:

Gesucht sind $\beta_j \in \mathbb{R}^n, j = 1, \dots, q$ mit

$$\begin{aligned} \max_{\|\beta_1\|=1} \text{Var}(\beta_1' x_t) &= \max_{\|\beta_1\|=1} \beta_1' \Gamma_0^x \beta_1 \\ \max_{\|\beta_i\|=1, \beta_i' \beta_j=0, j < i} \text{Var}(\beta_i' x_t) &= \max_{\|\beta_i\|=1, \beta_i' \beta_j=0, j < i} \beta_i' \Gamma_0^x \beta_i \end{aligned}$$

oder in Matrixschreibweise

$$\max_{B'IB=I} \text{tr}(B' \Gamma_0^x B) \quad (2.7)$$

mit $B = [\beta_1, \dots, \beta_q]$. Die Lösung für B ist die Matrix $B = [o_1, \dots, o_q]$.

Einen guten Überblick über die Eigenschaften von Principal Components findet man in [16].

2.4.2 PCA im dynamischen Fall

Im dynamischen Fall formt man die Zielfunktion (2.4) auf folgende Form um:

$$\min_{rk(A(\lambda)) \leq q} \int_{-\pi}^{\pi} \text{tr}(I - A(\lambda)) \Sigma^x(\lambda) (I - A(\lambda))^* d\lambda, \quad (2.8)$$

wobei man dies für jedes $\lambda \in (-\pi, \pi]$ löst.

Die Lösung ist somit

$$\begin{aligned} B(\lambda) &= [p_1(\lambda), \dots, p_q(\lambda)], \\ W(\lambda) &= B(\lambda)^*, \end{aligned}$$

wobei $p_i(\lambda)$ den i-ten normalisierten Eigenvektor von $\Sigma^x(\lambda) = P(\lambda)\Lambda(\lambda)P(\lambda)^*$ mit $\Lambda(\lambda) = \text{diag}(\mu_1(\lambda), \dots, \mu_n(\lambda)); \mu_1(\lambda) > \dots > \mu_n(\lambda)$ bezeichnet.

Für die fouriertransformierten b_j und w_j gilt somit:

$$\begin{aligned} b_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} B(\lambda) e^{i\lambda j} d\lambda, \\ w_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} W(\lambda) e^{i\lambda j} d\lambda. \end{aligned}$$

Kapitel 3

Darstellungstheorie

Bevor ich im nächsten Kapitel zwei Methoden für die Schätzung verallgemeinerter Faktormodelle vorstelle, möchte ich mit diesem Kapitel die Resultate der Paper [6] und [7] vorführen, in welchen die Autoren Chamberlain und Rothschild [CR] bereits vor über zwei Jahrzehnten hinreichende Bedingungen für Faktorstrukturen angeführt haben. Chamberlain und Rothschild gingen aber nicht auf zeitliche Abhängigkeiten ein. Da aber die beiden nachfolgenden Schätzungen darauf beruhen, die dynamischen Faktormodelle statisch zu formulieren, und überdies sich die Annahmen an die Strukturen stark ähneln, macht es durchwegs Sinn, diese Resultate zu präsentieren.

3.1 Grundlagen

x_i seien Zufallsvariablen aus dem $L_2(P)$ und sollen die zufälligen Returns von Assets darstellen, welche alle einen Preis von 1 haben. Der stochastische Vektor $\mathbf{x}_n = (x_1, \dots, x_n)$ repräsentiert also die Returns von n Assets, und ein Portfolio kann somit als $p_n = \alpha'_n \mathbf{x}_n$ dargestellt werden. Darüberhinaus soll bereits

$$\text{b.1 } \mathbb{E}(\mathbf{x}_n) = 0$$

gelten. CR verlangen dies nicht a-priori, da sie in ihrem Paper speziell auf die Existenz eines risikolosen Portfolios mit Erwartungswert ungleich 0 eingehen. Dies ist aber für die Darstellungstheorie nicht weiter von Bedeutung, da es durch die notwendige Mittelung nicht mehr auftritt, und daher von mir nicht weiter besprochen wird.

P_n bezeichne nun den von $[x_1, \dots, x_n]$ aufgespannten, linearen Unterraum des $L_2(P)$, und $P = \bigcup_{n=1}^{\infty} P_n$ die Vereinigung über alle n . Daraus folgt, dass \overline{P} , als Abschluss eines

linearen Unterraumes des Hilbertraumes $L_2(P)$ mit innerem Produkt $(x, y) = \mathbb{E}(xy)$, ebenfalls ein Hilbertraum ist.

Seien nun die Komponenten von \mathbf{x}_n folgendermaßen darstellbar

$$x_i = \beta_{i,1}f_1 + \dots + \beta_{i,r}f_r + e_i \quad (3.1)$$

mit $Var(e_i) \leq E < \infty, \forall i$, für die weiters gilt

b.2 $\mathbb{E}(\mathbf{f}) = 0; \mathbb{E}(\mathbf{e}_n) = 0$

b.3 $\mathbb{E}(\mathbf{f}\mathbf{e}'_n) = \mathbb{E}((f_1, \dots, f_r)'(e_1, \dots, e_n)) = 0$

b.4 $\sup_n \mu_{1n}^e \leq c$

gilt. b.4 heißt, dass der größte Eigenwert der Matrizen $\Gamma_n^e = \mathbb{E}(\mathbf{e}_n\mathbf{e}'_n)$ uniform beschränkt ist.

Definition 3.1.1. (approximative r-Faktorstruktur) *Die Folge $\{\Gamma_n\}$ besitzt eine approximative r-Faktorstruktur falls eine Folge $\{\beta_{i,1}, \dots, \beta_{i,r}\}_{i=1}^\infty$ existiert, sodass die Darstellung*

$$\Gamma_n = C_n C'_n + R_n \quad (3.2)$$

mit $C_{n,ij} = \beta_{i,j}$ möglich ist, und $\{R_n\}$ eine Folge von positiv semi-definiten Matrizen mit

$$\sup_n g_1(R_n) < \infty \quad (3.3)$$

ist. $g_1(R_n)$ sei der größte Eigenwert der Matrix R_n .

$$\sup_n g_r(\Gamma_n) = \sup_n \mu_{rn} = \infty \quad (3.4)$$

Es sei erwähnt das CR (3.4) nicht in ihrer Definition enthalten haben. Da die Annahme jedoch in weiterer Folge von zentraler Bedeutung ist, macht es durchwegs Sinn diese bereits in die Definition einzubauen.

Das heißt die Folge $\{\Gamma_n^x\} = \mathbb{E}(\mathbf{x}_n\mathbf{x}'_n)$, für die \mathbf{x}_n die Annahmen b.1 - b.4 erfüllt und die Darstellung (3.1) besitzt, erfüllt die Bedingungen der Definition 3.1.1, falls man noch

b.5 $\sup_n g_r(C_n C'_n) = \infty$

voraussetzt, da

$$\Gamma_n^x = \mathbb{E}(\mathbf{x}_n\mathbf{x}'_n) = B_n \mathbb{E}(\mathbf{f}\mathbf{f}') B'_n + \mathbb{E}(\mathbf{e}_n\mathbf{e}'_n) = B_n \Gamma_n^f B'_n + \Gamma_n^e = C_n C'_n + \Gamma_n^e$$

mit $C_n = B_n O_n \Lambda_n^{\frac{1}{2}}$, $\Gamma_n^f = O_n \Lambda_n O_n'$, stets erreicht werden kann, und $g_r(\Gamma_n^x) = \mu_{rn} > g_r(C_n C_n')$ gilt.

Die Annahme b.4 bzw. (3.3) stellt die größte Innovation der Paper von CR dar, da durch sie die zentrale Annahme der klassischen Frisch-Faktormodelle (CR verweisen speziell auf die Paper von Ross, welcher eine strikte Faktorstruktur voraussetzt), die Unkorreliertheit der idiosynkratischen Fehler, stark abgeschwächt wird.

Das Ziel ist somit, sowohl notwendige als auch hinreichende Bedingungen zu finden, um eine approximative Faktorstruktur gewährleisten zu können.

3.2 Faktorstruktur

Chamberlains und Rothschilds Bedingungen geht eine zentrale Annahme voraus: Es seien beliebig viele Assets vorhanden, sprich $n \rightarrow \infty$. Dies hat zwar auf der einen Seite den Nachteil, dass dies in der Realität nicht erreicht werden kann, auf der anderen Seite ist es ein Modell, mit dem man beliebig viele Assets beschreiben kann, und zwar mit wesentlich geringerem Rechenaufwand als zum Beispiel multivariate ARX-Modelle.

Mit Hilfe der Unbeschränktheit von n formulieren Chamberlain und Rothschild folgenden Satz:

Satz 3.2.1. *Angenommen $\sup_n \mu_{rn} = \infty$, $\mu_{r+1} < \infty$ und $\mu_\infty = \inf_n \mu_{nn} > 0$, wobei μ_{in} der i -te größte Eigenwert der Matrix $\Gamma_n^x = \mathbb{E}(\mathbf{x}_n \mathbf{x}_n')$ ist, dann*

(i) *folgt $\{\Gamma_n^x\}$ einer approximativen r -Faktorstruktur, sodass*

$$\Gamma_n^x = C_n C_n' + R_n$$

gilt, mit $C_{n,ij} = \beta_{ij}$, und die Folge $\{R_n\}$ besteht aus positiv semi-definiten Matrizen, deren Eigenwerte uniform beschränkt sind durch μ_{k+1} für alle n .

(ii) $\lim_{M \rightarrow \infty} C_{nM} C_{nM}' = C_n C_n'$ für alle n ,
wobei die j -te Spalte von C_{nM} die ersten n Einträge des j -ten Eigenvektors von Γ_M^x mit $M \geq n$, also $o_{jM} \mu_{jM}^{1/2}$, mit $\Gamma_M^x = (o_{1M}, \dots, o_{MM}) \text{diag}(\mu_{1M}, \dots, \mu_{MM}) (o_{1M}, \dots, o_{MM})'$, sind.

(iii) Die approximative r -Faktorstruktur aus (i) ist eindeutig. Das heißt, falls eine Folge $\{\gamma_{i1}, \dots, \gamma_{ir}\}_{i=1}^{\infty}$ existiert, sodass

$$\Gamma_n^x = G_n G_n' + W_n$$

mit $G_{n,ij} = \gamma_{i,j}$ gilt, und $\{W_n\}$ eine Folge von positiven semi-definiten Matrizen mit für n uniform beschränkten Eigenwerten ist, so folgt daraus, dass

$$G_n G_n' = C_n C_n', \text{ und } W_n = R_n$$

gelten muss.

Zu (i) und (iii) muss man glaube ich nicht viel sagen. (i) ist die zentrale Erkenntnis, nämlich, dass die Bedingungen $\sup_n \mu_{rn} = \infty$, $\mu_{r+1} < \infty$ und $\mu_{\infty} = \inf_n \mu_{nn} > 0$ ausreichen, um die Existenz einer approximativen Faktorstruktur zu gewährleisten. Und (iii) besagt, dass die Aufspaltung der Varianz-Kovarianz-Matrix aus (i) eindeutig ist.

(ii) legt den Grundstein für die nachfolgenden konsistenten Schätzmethoden, da es zeigt, dass die PCA punktweise gegen die tatsächliche Ladungsmatrix konvergiert.

Bemerkung Bevor ich jedoch den Beweis obigen Satzes bringe, möchte ich kurz auf die Umkehrung des Satzes eingehen. Das heißt, gelten die Annahmen $\sup_n \mu_{rn} = \infty$, $\mu_{r+1} < \infty$ und $\mu_{\infty} = \inf_n \mu_{nn} > 0$ falls \mathbf{x}_n einer Faktorstruktur folgt?

- Die Aussage $\mu_{\infty} = \inf_n \mu_{nn} > 0$ ist offensichtlich äquivalent dazu, dass Γ_n^x regulär ist für alle n . Da man die Regularität sowieso immer stillschweigend voraussetzt, ist dies kein Problem.
- $\mu_{r+1} < \infty$ folgt aus Satz(2.1.15) mit $H_1 = \Gamma_n^x$ und $H_2 = R_n$:

$$\mu_{r+1} \leq g_1(R_n) = \mu_1^e < c < \infty.$$

- $\sup_n \mu_{rn} = \infty$ setzt man durch (3.4) voraus.

Um Satz(3.2.1) zu beweisen, benötigt man folgendes Lemma.

Lemma 3.2.2. *Unter den Annahmen $\sup_n \mu_{rn} = \infty$, $\mu_{r+1} < \infty$ und $\mu_\infty = \inf_n \mu_{nn} > 0$ gilt:*

(i) *Es existiert eine Seminorm $\|\cdot\|^*$ auf \bar{P} , für die gilt: Falls*

$$p_n = \alpha' \mathbf{x}_n \rightarrow p,$$

so gilt

$$\lim_{n \rightarrow \infty} \|p_n\|^* = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^n \alpha_{in}^2 \right)^{1/2} = \|p\|^*.$$

(ii) *$F := \{p \in \bar{P} \mid \|p\|^* = 0\}$ ist ein r -dimensionaler linearer Unterraum von \bar{P} .*

(iii) *Sei $p_1 \in F$ und $p \in \bar{P}$, dann gilt $\|p_1 + p\|^* = \|p\|^*$.*

(iv) *Sei $p \in F^\perp$, dann gilt $\text{Var}(p) \leq \mu_{r+1} \|p\|^{*2}$.*

(v) *Sei $p|F$ die Projektion von p auf F und $p|F_n$ die Projektion von p auf die normierten Principal Components von \mathbf{x}_n*

$$\begin{aligned} p|F_n &= \sum_{i=1}^r \text{Cov}(p, r_{in}) r_{in} \text{ mit} \\ r_{in} &= (o'_{in} \mathbf{x}_n) / \mu_{in}^{1/2} \text{ wobei} \\ \Gamma_n^x &= O_n \text{diag}(\mu_{1n}, \dots, \mu_{nn}) O_n' \end{aligned} \tag{3.5}$$

dann gilt

$$\lim_{n \rightarrow \infty} p|F_n = p|F.$$

Den vollständigen Beweis des Lemmas findet man in [6]. Ich möchte mich auf (ii) konzentrieren, da dies die Kernaussage des Lemmas ist. Dadurch wird nämlich gezeigt, dass der Faktorraum tatsächlich r -dimensional ist.

Beweis (Lemma 3.2.2(ii))

- $\dim(F) \leq r$

Angenommen $\dim(F) > r \Rightarrow \exists p_1, \dots, p_r \in F : (p_j, p_i) = \delta_{ij}$ ¹. Weiters $\exists \{p_{jn}\} :$

¹ $\delta_{ij} = 1$ für $i = j$, 0 sonst.

$p_{jn} \rightarrow p_j$, für die gilt:

$$\begin{aligned}
(p_{jn}, p_{kn}) &= (\alpha'_{jn} \mathbf{x}_n, \alpha'_{kn} \mathbf{x}_n) = \alpha'_{jn} \Gamma_n^x \alpha_{kn} \\
&= \alpha'_{jn} \Gamma_{1n}^x \alpha_{kn} + \alpha'_{jn} \Gamma_{2n}^x \alpha_{kn} \text{ mit} \\
\Gamma_{1n}^x &= \sum_{j=1}^r \mu_{jn} o_{jn} o'_{jn} \text{ und } \Gamma_{2n}^x = \Gamma_n^x - \Gamma_{1n}^x.
\end{aligned} \tag{3.6}$$

Da aber

$$\begin{aligned}
\alpha'_{jn} \Gamma_{2n}^x \alpha_{kn} &\leq \alpha'_{jn} \Gamma_{2n}^x \alpha_{jn} \alpha'_{kn} \Gamma_{2n}^x \alpha_{kn} \\
&\leq \mu_{r+1} \alpha'_{jn} \alpha_{jn} \mu_{r+1} \alpha'_{kn} \alpha_{kn} = \mu_{r+1}^2 \|p_{jn}\|^{*2} \|p_{kn}\|^{*2} \rightarrow 0
\end{aligned}$$

gilt, und deswegen für $n \rightarrow \infty$ (3.6) gegen (p_j, p_k) konvergiert, gilt dies auch für $\alpha'_{jn} \Gamma_{1n}^x \alpha_{kn}$.

Folglich muss

$$A'_n \Gamma_{1n}^x A_n \rightarrow I_{r+1} \text{ konvergieren, mit } A_n = (\alpha_{1n}, \dots, \alpha_{rn}).$$

Da jedoch $\det(A'_n \Gamma_{1n}^x A_n) = 0$, wegen $rk(\Gamma_{1n}^x) = r$, gilt, ist dies ein Widerspruch in sich.

- $\dim(F) \geq r$

Seien r_{in} die normierten Principal Components aus (3.5), so gilt

$$\|r_{in}\|^* = \|(o'_{in} \mathbf{x}_n) / \mu_{in}^{1/2}\|^* = \|o_{in}\| / \mu_{in}^{1/2} = 1 / \mu_{in}^{1/2}$$

Weiters betrachte man die Aufspaltung $r_{in} = r_{1in} + r_{2in}$ mit $r_{1in} \in F$. Dann gilt für $i \leq r$

$$\|r_{2in}\|^2 \leq \mu_{r+1} \|r_{2in}\|^{*2} = \mu_{r+1} \|r_{in}\|^{*2} = \mu_{r+1} / \mu_{in} \rightarrow 0 \tag{3.7}$$

Die obige Ungleichung stimmt wegen (3.2.2)(iv), und die nachfolgende Gleichheit folgt aus (3.2.2)(iii).

Definiert man $R_n = (r_{1n}, \dots, r_{rn})' = (r_{11n}, \dots, r_{1rn})' + (r_{21n}, \dots, r_{2rn})'$, so kann man $R_n R'_n = R_{1n} R'_{1n} + R_{2n} R'_{2n}$ schreiben. Da $R_n R'_n = I_r$ und $R_{2n} R'_{2n} \rightarrow 0$ gilt, konvergiert $R_{1n} R'_{1n} \rightarrow I_r$. Daraus folgt, dass für jedes ϵ ein M existiert, sodass alle Elemente von $R_{1n} R'_{1n}$ bereits in einer ϵ -Umgebung der jeweiligen Elemente von I_r sind. Wählt man ϵ so klein, dass $R_{1n} R'_{1n}$ strikt diagonal dominant ist, folgt daraus deren Regularität und dadurch die Unabhängigkeit der $(r_{11n}, \dots, r_{1rn})$, welche alle aus F sind. \square

Beweis (Satz 3.2.1) Aus Lemma(3.2.2) weiß man bereits dass F r -dimensional ist. Seien daher f_1, \dots, f_r eine Orthonormalbasis von F , dann lässt sich x_i anschreiben als die Summe der Projektionen auf F und dessen Orthogonalraum

$$x_i = x_i|F + x_i|F^\perp = \beta_{i1}f_1 + \dots + \beta_{i,r}f_r + e_i$$

mit $\beta_{ij} = Cov(x_i, f_j)$.

(i) Definiere $C_{n,ij} := \beta_{ij}$ und $R_{n,ij} := Cov(e_i, e_j)$. So gilt

$$\alpha' R_n \alpha = Var(\alpha' \mathbf{e}_n) \leq \mu_{r+1} \|\alpha' \mathbf{e}_n\|^{*2} = \mu_{r+1} \|\alpha' \mathbf{x}_n\|^{*2} = \mu_{r+1} \alpha' \alpha$$

die Ungleichung wegen Lemma(3.2.2)(iv) und die nächste Gleichheit wegen Lemma(3.2.2)(iii).

$\Rightarrow g_1(R_n) \leq \mu_{r+1}$ und die gewünschte Faktorstruktur existiert.

(ii) z.z.: $\lim_{M \rightarrow \infty} \sum_{k=1}^r \beta_{ik,M} \beta_{jk,M} = \sum_{k=1}^r \beta_{ik} \beta_{jk}$

rechte Seite:

$$\begin{aligned} Cov(x_i, x_j|F) &= Cov\left(\sum_{k=1}^r \beta_{ik} f_k + e_i, \sum_{l=1}^r \beta_{jl} f_l\right) = Cov\left(\sum_{k=1}^r \beta_{ik} f_k, \sum_{l=1}^r \beta_{jl} f_l\right) \\ &= \sum_k Cov(\beta_{ik} f_k, \sum_{l=1}^r \beta_{jl} f_l) = \sum_k Cov(\beta_{ik} f_k, \beta_{jk} f_k) = \sum_{k=1}^r \beta_{ik} \beta_{jk} \end{aligned}$$

linke Seite:

betrachte zuerst

$$\begin{aligned} Cov(x_j, r_{kM}) &= Cov(x_j, o'_{kM} \mathbf{x}_M / \mu_{kM}^{1/2}) = \mu_{kM}^{-1/2} Cov(x_j, \sum_{i=1}^M o_{ikM} x_{iM}) \\ &= \mu_{kM}^{-1/2} \sum_{i=1}^M Cov(x_j, o_{ikM} x_{iM}) = \mu_{kM}^{-1/2} \sum_{i=1}^M o_{ikM} Cov(x_j, x_{iM}) \\ &= \mu_{kM}^{-1/2} o'_{kM} (\Gamma_{Mi}^x)' = \mu_{kM}^{-1/2} o'_{kM} O_M \Lambda o_{iM} = \mu_{kM}^{1/2} o_{kM,i}, \end{aligned}$$

wobei Γ_{Mi}^x die i -te Zeile von Γ_M^x ist, und anschließend

$$\begin{aligned} Cov(x_i, x_j|F_M) &= Cov(x_i, \sum_{k=1}^r Cov(x_j, r_{kM}) r_{kM}) = \sum_{k=1}^r Cov(x_i, Cov(x_j, r_{kM}) r_{kM}) \\ &= \sum_{k=1}^r Cov(x_i, r_{kM}) Cov(x_j, r_{kM}) = \sum_{k=1}^r \mu_{kM}^{1/2} o_{kM,i} \mu_{kM}^{1/2} o_{kM,j} = \sum_{k=1}^r \beta_{ik,M} \beta_{jk,M}. \end{aligned}$$

Da nun dank Lemma(3.2.2)(v) $\lim_{M \rightarrow \infty} Cov(x_i, x_j | F_M) = Cov(x_i, x_j | F)$ gilt, ist auch dies gezeigt.

(iii) Da dieser Beweis noch technischer ist als der vorige, möchte ich den interessierten Leser auf [7] verweisen. \square

Verallgemeinerte PCA

Zu guter letzt zeigen Chamberlain und Rothschild noch, dass Satz(3.2.1)(ii) ebenfalls gilt, falls man die Linearkombinationen von x nicht wie üblich mit dem kanonischen inneren Produkt I_n sucht

$$\max_{B' I_n B = I} tr(B' \Gamma_n^x B),$$

sondern stattdessen eine beliebige, positiv-definite Matrix Ω_n mit $g_1(\Omega_n) < \infty$ und $\inf_n g_n(\Omega_n) > 0$ betrachtet (verallgemeinerte PCA)

$$\max_{B' \Omega_n B = I} tr(B' \Gamma_n^x B). \quad (3.8)$$

Das heißt, asymptotisch liefern die beiden Ladungsmatrizen dasselbe.

Einer der Unterschiede der folgenden Schätzmethoden beruht auf eben dieser Tatsache, dass die PCA bzw. verallgemeinerte PCA asymptotisch dasselbe liefern. Jene Autoren, die die verallgemeinerte PCA verwenden, erhoffen sich, durch zuvorige konsistente Schätzung von Γ_n^e , welche sie als Ω_n verwenden, die Effizienz der Methode zu verbessern.

Dynamische Darstellungstheorie

Die Autoren Forni und Lippi haben in [9] die Darstellungstheorie auf dynamische Faktormodelle erweitert.

$$x_{i,t} = b_{1i}(z)u_{1t} + \dots + b_{qi}(z)u_{qt} + \varepsilon_{it} \quad (3.9)$$

Dadurch werden die Annahmen der Eigenwerte der Varianz-Kovarianz-Matrix auf die der spektralen Dichte-Matrix umgeändert.

Definition 3.2.3. (q - dynamische Faktorfolge) Sei $q > 0, \in \mathbb{N}$. Eine Folge

$$\{(x_{it}, \dots, x_{nt})', t \in \mathbb{Z}\}_{n \in \mathbb{N}}$$

heißt q - dynamische Faktorfolge, falls ein q -dimensionaler orthonormaler White-Noise Prozess

$$\{(u_{it}, \dots, u_{qt})', t \in \mathbb{Z}\}$$

und eine Folge

$$\{\varepsilon_{it}, t \in \mathbb{Z}\}_{i \in \mathbb{N}}$$

existieren die stationär sind und deren spektrale Dichte existiert, sodass gilt:

- (i) $x_{it} = b_{1i}(z)u_{1t} + \dots + b_{qi}(z)u_{qt} + \varepsilon_{it}, \forall i \in \mathbb{N}$ mit $b_{ij} \in L_2(P)$,
- (ii) $\mathbb{E} \varepsilon_{it}u_{js} = 0, \forall s, t \in \mathbb{Z}, i \in \mathbb{N}, j = 1, \dots, q$,
- (iii) $\lambda_1^\varepsilon(\theta) < c, \theta$ -fast überall,
- (iv) $\lambda_q^x(\theta) = \infty, \theta$ -fast überall.

$\lambda_i^\varepsilon(\theta)$ bezeichnet den i -t größten Eigenwert der spektralen Dichte $\Sigma^\varepsilon(\theta)$.

Die Existenz einer solchen q - dynamischen Faktorfolge lässt sich mit folgendem Satz festlegen.

Satz 3.2.4. Die Folge $\{(x_{it}, \dots, x_{qt})', t \in \mathbb{Z}\}_{n \in \mathbb{N}}$ ist eine q - dynamische Faktorfolge dann und nur dann falls

- (i) $\lambda_{q+1}^x(\theta) < c, \theta$ -fast überall
- (ii) $\lambda_q^x(\theta) = \infty, \theta$ -fast überall

gelten.

Eine mögliche Schätzung ähnlich zu Satz(3.2.1) wird dann ausgiebig in [11] besprochen. Da jedoch die Schätzung nur von theoretischer Natur ist, geht man einen Schritt zurück, indem man die Dynamik etwas einschränkt. Das heißt, die linearen Filter $b_{j,i}$ aus (3.9) sollen zum einen nur noch kausal, zum anderen nur beschränkt in die Vergangenheit reichen, und der q -dimensionale Faktorprozess soll eine VAR-Darstellung besitzen. Somit sind die $b_{j,i}$ Polynome und (3.9) besitzt eine statische Darstellung, womit man sich in gewisser Weise wieder in einem statischen Fall befindet.

Kapitel 4

Schätzung und Vorhersage des verallgemeinerten dynamischen Faktormodells

In diesem Kapitel, stelle ich zwei Methoden vor, die unter geeigneten Annahmen eine konsistente Schätzung des latenten Prozesses ermöglichen. Eine wichtige Annahme ist hier, dass sowohl der Zeithorizont als auch die Dimension des beobachtbaren Prozesses unbeschränkt sind. Das heißt, man verwendet die vorhandene Beobachtung des n -dimensionalen Prozesses $x_t : X = (x_{it})_{i=1,\dots,n;t=1,\dots,T}$ und zeigt, dass unter der Annahme, dass $\lim_{n,T \rightarrow \infty} (n, T) = (\infty, \infty)$, ohne dass die Wachstumsraten in einer bestimmten Relation zueinander stehen, die propagierten Schätzer konsistent sind.

Weiters wird in Folge angenommen, dass die Lag-Polynome des Faktorprozesses beschränkte Ordnung haben, und der Faktorprozess eine VAR-Darstellung besitzt. Das hat den Vorteil, dass das Modell auch statisch formuliert werden kann. Durch die Festlegung auf Polynome gewährleistet man ebenso die Unabhängigkeit von zukünftigen Werten des Faktorprozesses, man möchte ja schließlich auch prognostizieren können, wobei Abhängigkeiten von der Zukunft naturgemäß fehl am Platz wären.

4.1 Die Methode von Forni, Hallin, Lippi und Reichlin

Die erste Methode, das bereits mehrfach angesprochene verallgemeinerte, dynamische Faktormodell, konsistent zu schätzen, ist jene der Autoren Forni, Hallin, Lippi und Reichlin [FHLR].

Forni, Hallin, Lippi und Reichlin haben bereits in ihrem Paper [11] (*The generalized factor model: identification and estimation*) eine konsistente Schätzung vorgeschlagen, und diese, auf Grund praktischer Probleme bei der Prognosenberechnung durch ihr Paper [12] (*The generalized dynamic factor model: one-sided estimation and forecasting*) verbessert.

In diesem Abschnitt werde ich die Annahmen und Lemmata des Papers [12] durchgehen und die zur Konsistenz notwendigen Beweise im Appendix nachbringen.

Die FHLR-Methode schätzt den von den Faktoren aufgespannten Raum (Faktorraum) in zwei Stufen.

Zuerst werden die spektralen Dichten $\Sigma^x(\lambda)$, $\Sigma^\chi(\lambda)$ und $\Sigma^\varepsilon(\lambda)$ (für n fest) konsistent geschätzt, beginnend mit dem Schätzer für $\Sigma^x(\lambda)$ mit (2.3) und endend damit, dass die Aufspaltung $\hat{\Sigma}^x(\lambda) = \hat{\Sigma}^\chi(\lambda) + \hat{\Sigma}^\varepsilon(\lambda)$ konsistente Schätzer für $\Sigma^\chi(\lambda)$ und $\Sigma^\varepsilon(\lambda)$ liefert.

Anschließend werden die durch diese Zerlegung implizit bestimmten Schätzer der Varianz - Kovarianzfunktionen berechnet, und schlussendlich der Faktorraum so bestimmt, dass die Varianzen der Linearkombinationen des Fehlerprozesses möglichst klein werden, besser gesagt, für $n \rightarrow \infty$ verschwinden.

4.1.1 Das Modell

Um die Konsistenz der obigen Schätzungen gewährleisten zu können, benötigt man folgende Annahmen:

A1 X ist eine Beobachtung des reellen, stationären Prozesses

$\{x_t = (x_{1t}, \dots, x_{nt})'; t \in \mathbb{Z}\}_{n \in \mathbb{N}}$, welcher Erwartungswert 0 und endliche zweite Momente $\Gamma_k = \mathbb{E}x_{t+k}x_t^*$ besitzt.

A2 Für alle $n \in \mathbb{N}$ kann der Prozess x_t folgendermaßen dargestellt werden: $x_t =$

$\sum_{j=0}^{\infty} c_j w_{t-j}$, mit $w_t = (w_{1t}, \dots, w_{nt})'$; $aw_t \neq 0$, $\forall a \neq \mathbf{0} \in \mathbb{R}^{1 \times n}$; w_t ist White-

Noise mit endlichen vierten Momenten, und für die $c_j = (c_{ik,j})_{i,k=1,\dots,n}$ gilt

$\sum_{j=0}^{\infty} \|c_{ik,j}\| j^{\frac{1}{2}} < \infty, \forall i, j, n \in \mathbb{N}$.

Die Existenz von absolut summierbaren Kovarianzfunktionen und daraus resultierenden spektralen Dichten gewährleistet A2¹. A2 mag in dieser Hinsicht übertrieben erscheinen, doch durch diese Bedingung folgt auch die Konsistenz des später verwendeten Schätzers der spektralen Dichte.

Weiters soll für den latenten Teil

$$\chi_{it} = b_{i1}(z)f_{1t} + \dots + b_{iq}(z)f_{qt} \quad (4.1)$$

gelten, wodurch das Modell die gewohnte Form bekommt:

$$x_t = \chi_t + \varepsilon_t = B(z)f_t + \varepsilon_t; \quad (4.2)$$

wobei f_t eine vektorautoregressive Darstellungsform besitzt:

$$f_t : A(z)f_t = u_t$$

und folgende weitere Annahmen an den Faktorprozess f_t gelegt werden:

- B1 (a) Die Lagpolynome $(b_{ij}(z))_{j=1,\dots,q}$ aus (4.1) sollen Ordnung s haben. Das heißt, dass $B(z) = B_0 + \dots + B_s z^s$ eine Polynommatrix mit $B_s \neq 0 - Matrix$ ist.
- (b) $A(z) = A_0 + \dots + A_S z^S$ soll eine Polynommatrix der Ordnung $S \leq s$ sein, wobei wieder $A_S \neq 0 - Matrix$ gelten soll.
- (c) $\det(A(z)) \neq 0, \forall z \in \mathbb{C}$ mit $|z| \leq 1$.
- B2 $u_t = (u_{1t}, \dots, u_{qt})$ sei ein orthonormaler White-Noise Prozess, welcher orthogonal zum Fehlerprozess ε_t ist.

B1(a) ist eine Einschränkung auf die Darstellungsform (4.2), da die Abhängigkeit auf den Faktorprozess auf endlich viele vergangene Zeitpunkte beschränkt ist. Man beachte jedoch, dass B1(c) eine Darstellungsform von $f_t = (A(z))^{-1}u_t = \sum_{k=0}^{\infty} a_k u_{t-k}$ ermöglicht, sodass sich u_t wieder als Faktorprozess f_t von (1.5) interpretieren lässt.

Durch B1(a) hat man somit die Möglichkeit, das Modell (4.2) umzuschreiben:

¹siehe [5] Theorem 4.4.1

$$x_t = CF_t + \varepsilon_t \quad (4.3)$$

mit $F_t = (f'_t, \dots, f'_{t-s})'$ und $C = (B_0 \cdots B_s)$. Dieses Modell kann nun als statisches Faktormodell mit $r = q(s+1)$ statischen anstatt wie bisher mit q dynamischen Faktoren interpretiert werden.²

Die Bedingung B2 impliziert die Orthogonalität von χ_t auf ε_t , da

$$\begin{aligned} \mathbb{E}\chi_t \varepsilon'_t &= \mathbb{E}(B(z)(A(z))^{-1}u_t)\varepsilon'_t \\ &= \mathbb{E}(B_0 \sum_{k=0}^{\infty} a_k u_{t-k} + \dots + B_s \sum_{k=0}^{\infty} a_k u_{t-k-s})\varepsilon'_t \\ &= B_0 \sum_{k=0}^{\infty} a_k \mathbb{E}u_{t-k}\varepsilon'_t + \dots + B_s \sum_{k=0}^{\infty} a_k \mathbb{E}u_{t-k-s}\varepsilon'_t = 0 \end{aligned}$$

gilt. B1(b) ist weniger eine Einschränkung als viel mehr eine Festlegung. Man möchte lediglich, dass der Faktorraum, der für die Prognose verwendet werden soll, bereits durch die in (4.1) und (4.2) verwendeten Lags des Faktorprozesses bestimmt ist:

Möchte man den besten linearen Schätzer für χ_{T+h} basierend auf $F_{jt}, j = 1, \dots, r; t \in \mathbb{Z}$, $\chi_{T+h|T}$ bestimmen, so ist dies die Projektion von χ_{T+h} auf $F_{jt}, j = 1, \dots, r$. Das heißt, man benötigt die Prognosen des Faktorprozesses F_{jT+h} , bedingt man kennt den Prozess bis zum Zeitpunkt T. Da nun f_t einem VAR-Prozess folgt, kann man seine Prognosen einfach bestimmen:

$$f_{T+1|T} = -A_1 f_T - \dots - A_S f_{T-S}$$

da u_{T+1} orthogonal zu $\{f_t, t \leq T\}$ ($f_t = \sum_{j=0}^{\infty} k_j u_{t-j}$) ist. Wobei

$$f_{T+2|T} = -A_1 f_{T+1|T} - \dots - A_S f_{T-(S-1)}; \text{ usw. ist.}$$

Wie man sieht, ist der weitest zurückliegende Zeitpunkt, den man benötigt, (T-S). Das heißt, der durch $\{f_{T1}, \dots, f_{Tq}, \dots, f_{T-s,1}, \dots, f_{T-s,q}\}$ aufgespannte Raum reicht für

² $r = q(s+1)$ gilt nur dann, falls jeder Lag jedes Faktors auch tatsächlich gebraucht wird. Das tatsächliche r soll die Anzahl der f_{it} (inklusive all ihrer Lags) aus (4.2) sein. Das heißt, der erste dynamische Faktor f_{1t} muss nicht gleichviele Lags in der Darstellung (4.2) beinhalten wie zum Beispiel f_{2t} . Dies ist möglich, da an die Matrizen B_j keine Restriktionen gelegt wurden (außer, dass $B_s \neq 0$ -Matrix ist). Die Relation $r = q(s+1)$ ist somit nicht zwingend und wird in weiterer Folge auch nicht benötigt, sie soll daher nur als Erklärung für den Übergang dienen.

die bestmögliche Prognose aus. Würde man auf B1(b) verzichten wollen, würde dies also lediglich eine formale Änderung der Beweise bedeuten.

Durch die Bedingung B ist die Äquivalenz der beiden Darstellungsmöglichkeiten (4.2) und (4.3) gegeben. In weiterer Folge bezeichne nun $L(F|t)$ den durch die r statischen Faktoren $F_{jt}, j = 1, \dots, r$ aufgespannten Raum, welchen ich schlicht mit Faktorraum bezeichnen werde.

Waren die Annahmen A und B³ bis jetzt dafür verantwortlich, dass eine statische Formulierung möglich ist, so sichert die folgende Annahme zum einen die Existenz einer dynamischen q -Faktorstruktur (siehe [9]) und zum anderen die konsistente Schätzung der spektralen Dichten des latenten und des Fehlerprozess durch dynamische PCA.

- C1 (a) $\lambda_{q,n}^x(\theta) \rightarrow \infty$ für $n \rightarrow \infty$, θ -fast überall auf $[-\pi, \pi]$.
 (b) $\lambda_{i,n}^x(\theta) > \lambda_{i+1,n}^x(\theta)$, θ -fast überall auf $[-\pi, \pi]$, für $i=1, \dots, q$.
- C2 $\lambda_{1,n}^\varepsilon(\theta) \leq M$, M fest, $\forall \theta \in [-\pi, \pi], \forall n \in \mathbb{N}$.

$\lambda_{i,n}^x$ bzw. $\lambda_{i,n}^\varepsilon$ bezeichnen hier den i -ten Eigenwert der $n \times n$ Matrizen Σ^x bzw. Σ^ε und werden in weiterer Folge ohne den Index n geschrieben. Sofern nicht anders angegeben versteht man unter $\lambda_q^x(\theta) \rightarrow \infty$ für $n \rightarrow \infty$ die Divergenz der Folge $\{\lambda_{q,n}^x\}_{n \in \mathbb{N}}$. Das heißt, C1(b) verlangt die strikte Ungleichheit der größten Eigenwerte. Dies ist zwar nicht notwendig, aber da es keine einschneidende, dafür aber sehr hilfreiche Annahme ist, soll sie gelten.

C1(a) verlangt, dass die q größten Eigenwerte der spektralen Dichte des latenten Prozesses divergieren. Diese Annahme ist die Idee dieser Methode und hat einen dementsprechend großen Stellenwert. Zusammen mit C2 welche die Beschränktheit der Eigenwerte der spektralen Fehlerdichte verlangt, wird die Trennung möglich.

Es ist zu beachten, wie relativ wenige Bedingungen an die Struktur der Fehler verlangt werden. Es sollen lediglich die Eigenwerte der spektralen Dichte „brav“, sprich gleichmäßig beschränkt, sein. Das heißt, schwache Kreuz-Korrelationen und Abhängigkeiten über die Zeit sind zulässig! Der Fall der linearen Unabhängigkeit der Fehler ist natürlich nicht ausgeschlossen, aber eben keineswegs eine Bedingung.

Die Annahme C sorgt zwar dafür, dass die spektralen Dichten des latenten und des Fehlerprozesses eindeutig getrennt werden können, jedoch reicht sie nicht aus, um die

³Wenn bei den Annahmen nur die Großbuchstaben stehen, sind dadurch sämtliche Annahmen, die mit dem jeweiligen Großbuchstaben beginnen, gemeint.

Divergenz der r größten Eigenwerte der Kovarianzmatrix $\Gamma_0^x = \mathbb{E}(\chi_t \chi_t')$ zu garantieren, welche jedoch im Beweis der Konsistenz benötigt wird. Deswegen die weitere Annahme

$$D \quad \mu_{r,n}^x \rightarrow \infty, \text{ für } n \rightarrow \infty.$$

$\mu_{i,n}^x$ sei der i -te Eigenwert der $n \times n$ Kovarianzmatrix $\Gamma^x(0)$. Ebenso wie für die Eigenwerte der spektralen Dichte, wird in weiter Folge auf den Index n verzichtet, solange keine Verwechslungsgefahr besteht.

Ein simples, aber dennoch anschauliches Beispiel für die Unabhängigkeit der Annahme D von $C1(a)$ ist folgendes

Beispiel Sei $\chi_{1t} = u_{t-1}$ und $\chi_{it} = u_t, \forall i \geq 2$, mit $\text{Var}(u_t) = \text{Var}(u_{t-1}) = 1$. Dann ist

$$\Gamma_0^x = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 1 \\ 0 & 1 & \ddots & & \\ \vdots & \vdots & & & \\ 0 & 1 & & & 1 \end{pmatrix}$$

und

$$\Sigma^x(\theta) = \frac{1}{2\pi} [\Gamma_{-1}^x e^{i\theta} + \Gamma_0^x + \Gamma_1^x e^{-i\theta}] = \begin{pmatrix} 1 & e^{-i\theta} & \dots & e^{-i\theta} \\ e^{i\theta} & 1 & \dots & 1 \\ \vdots & \vdots & & \\ e^{i\theta} & 1 & & 1 \end{pmatrix}$$

Deren charakteristische Polynome sind dann:

$$\begin{aligned} \text{char}(\Gamma_0^x) &= x^{n-2}(1-x)(n-1-x) \\ \text{char}(\Sigma^x(\theta)) &= x^{n-1}(n-x) \end{aligned}$$

Wie man sieht divergiert jeweils nur der größte Eigenwert, während der zweite in beiden Fällen beschränkt bleibt. Bei dem unteren charakteristischen Polynom ist das erwünscht, jedoch sollte der zweitgrößte Eigenwert der Kovarianzmatrix Γ_0^x auch divergieren, damit man eine Möglichkeit hat herauszufinden, dass 2 statische Faktoren benötigt werden. Dieses Beispiel gilt sogar auch, wenn endlich viele $\chi_{it} = u_{t-1}$ sind⁴.

⁴Das charakteristische Polynom der Kovarianzmatrix Γ_0^x lautet dann $x^{n-2}(n-m-x)(m-x)$, wobei $m = |\{\chi_{it} = u_{it}\}|$.

Die Annahme E ist rein technischer Natur. Es sei nur soviel gesagt, dass sie keine einschneidende Beschränkung ist, jedoch im Beweis der konsistenten Schätzung des Faktorraumes eine Vereinfachung darstellt.

- E (a) $\check{\mu}_k^\chi > \check{\mu}_{k+1}^\chi$, $k = 1, \dots, r$.
 (b) $\check{\mu}_n^\varepsilon > \delta$ für $n \rightarrow \infty$.

4.1.2 Der Schätzer

Die Tatsache, dass sich der Prozess x_t in zwei unkorrelierte Prozesse χ_t und ε_t aufspalten lässt, führt dazu, dass man einen Schätzer für x_t durch Schätzer $\hat{\chi}_t$ und $\hat{\varepsilon}_t$ für χ_t und ε_t bekommt.

Das Interesse gilt hier einzig dem Schätzer $\hat{\chi}_t$, welcher durch die Projektion von x_t auf den Faktorraum, der eine Linearkombination der x 's ist, entsteht.

Der Faktorraum seinerseits wird so bestimmt, dass der Anteil des idiosynkratischen Fehlers ε_t , welcher nur schwach korreliert ist, mit wachsendem n verschwindet.

Das heißt, man sucht eine Linearkombination der x 's: $a'x_t = a_1x_{t1} + \dots + a_nx_{tn}$, welche unter Berücksichtigung der Restriktion $Var(ax_t) = 1$ möglichst nahe am tatsächlichen Faktorraum liegt. Sprich

$$\min_a Var(a'\varepsilon_t), \text{ s.t. } Var(x_t) = 1 \quad (4.4)$$

oder äquivalent dazu

$$\max_a Var(a'\chi_t), \text{ s.t. } Var(\varepsilon_t) = 1 \quad (4.5)$$

Gesucht werden somit r unabhängige Kombinationen $\hat{V}_{jt} = \hat{Z}'_j x_t$, die rekursiv durch folgende Bedingungen konstruiert werden:

$$\begin{aligned} \hat{Z}_j &= \arg \max_{a \in \mathbb{R}^n} a' \hat{\Gamma}_0^\chi a \\ \text{s.t. } a' \hat{\Gamma}_0^\varepsilon a &= 1, \quad a' \hat{\Gamma}_0^\varepsilon \hat{Z}_i = 0, i < j \end{aligned} \quad (4.6)$$

wobei $\hat{\Gamma}_0^\chi$ bzw. $\hat{\Gamma}_0^\varepsilon$ Schätzer der Varianz-Kovarianzmatrizen $\Gamma_0^\chi = \mathbb{E}(\chi_t \chi_t')$ bzw. $\Gamma_0^\varepsilon = \mathbb{E}(\varepsilon_t \varepsilon_t')$ mit n endlich sind.

Dies entspricht dem Maximierungsproblem einer verallgemeinerten PCA (2.7) und praktischerweise kann man diese Formulierung durch ein paar Umformungen auf ein verallgemeinertes Eigenwertproblem wie in (Satz 2.1.4) mit $A = \hat{\Gamma}_0^\chi$ und $B = \hat{\Gamma}_0^\varepsilon$ zurückführen:

$$\max_{a \in \mathbb{R}^n} a' \hat{\Gamma}_0^\chi a, \quad \text{s.t. } a' \hat{\Gamma}_0^\varepsilon a = 1$$

$$\begin{aligned}
&\Leftrightarrow \max_{a \in \mathbb{R}^n} L(a, \lambda) = a' \hat{\Gamma}_0^\chi a - \lambda(a' \hat{\Gamma}_0^\varepsilon a - 1) \\
&\Rightarrow \frac{\partial L(a, \lambda)}{\partial a} = 2\hat{\Gamma}_0^\chi a - 2\lambda \hat{\Gamma}_0^\varepsilon a = 0 \\
&\Leftrightarrow \hat{\Gamma}_0^\chi a - \lambda \hat{\Gamma}_0^\varepsilon a = 0
\end{aligned}$$

Die gesuchten $\hat{Z}_j, j = 1, \dots, r$ sind somit die zu den r größten verallgemeinerten Eigenwerten gehörigen Eigenvektoren von $(\hat{\Gamma}_0^\chi, \hat{\Gamma}_0^\varepsilon)$.

Satz 4.1.1. *Sei x_t ein stochastischer Prozess für den Annahme A zutrifft, und welcher in der Form von (4.2) darstellbar ist, sodass die Annahmen B, C, D und E erfüllt sind, dann sind*

$$\hat{\chi}_t = [\hat{\Gamma}_0^\chi \hat{Z} (\hat{Z}' \hat{\Gamma}_0 \hat{Z})^{-1}] \hat{Z}' x_t \quad (4.7)$$

$$\hat{\chi}_{T+h|T} = [\hat{\Gamma}_h^\chi \hat{Z} (\hat{Z}' \hat{\Gamma}_0 \hat{Z})^{-1}] \hat{Z}' x_T \quad (4.8)$$

ein konsistenter in-sample-Schätzer bzw. ein konsistenter out-of-sample-Prädiktor für den latenten Prozess χ_t .

Wobei $\hat{Z} = [\hat{Z}_1, \dots, \hat{Z}_r]$ ist.

Bemerkung 1 $\hat{Z}x_t$ spannt den geschätzten Faktorraum auf, auf den die zukünftigen⁵ χ_t projiziert werden.

Bemerkung 2 Unter der Voraussetzung, dass man den optimalen Faktorraum bestimmt hat, kommt man auf (4.7) und (4.8), indem man sich zuerst den OLS-Schätzer von $(\chi_{T+h,i} \sim F)$ für $i = 1, \dots, n$ ansieht, wobei $F = XZ$, mit $\dim(F) = (T \times r)$ die (in Wirklichkeit unbeachtbare!) „Beobachtung“ des tatsächlichen Faktorprozesses ist. Dies ergibt:

$$\hat{\chi}_{t+h} = \chi_h' F [F' F]^{-1} F_t,$$

wobei $F_t = Z' x_t$ der Faktorprozess zum Zeitpunkt t ($r \times 1$), und χ_h die „Beobachtungen“ des um h Zeitpunkte verschobenen latenten Prozesses ($T \times n$) sind.

Da $\frac{1}{T} \chi_h' F$ den selben Varianz-Kovarianz-Matrix-Schätzer darstellt wie $\frac{1}{T} \chi_h' (\chi Z)$ ⁶, erhält man die Form

$$\begin{aligned}
\hat{\chi}_{t+h} &= \frac{1}{T} \chi_h' \chi Z \left[\frac{1}{T} (Z' X' X Z) \right]^{-1} Z' x_t \\
&= \left(\frac{1}{T} \chi_h' \chi \right) Z \left[Z' \left(\frac{1}{T} X' X \right) Z \right]^{-1} Z' x_t^7 \\
&= \hat{\Gamma}_h^\chi \hat{Z} [\hat{Z}' \hat{\Gamma}_0 \hat{Z}]^{-1} \hat{Z} x_t.
\end{aligned}$$

⁵Es müssen nicht notwendigerweise zukünftige χ_t sein. In (4.7) z.B. werden die geschätzten latenten Variablen **innerhalb** des bekannten Zeitraumes geschätzt.

⁶ $F = XZ$, X kann man darstellen als $X = \chi + \varepsilon$ und zeilenweise gilt a.3, wodurch die Gleichheit folgt.

Beweis (Satz 4.1.1)

Schritt 1 Der erste Schritt folgt den Überlegungen von [11]. In diesem Paper wird die Konsistenz des Schätzers $\hat{\chi}_t = B(z)B(z)^*x_t$, welcher durch die Transferfunktion $B(z) = \sum_{j=-\infty}^{\infty} b_j z^j$ mit b_j aus (2.5) festgelegt ist, gezeigt. Der aufmerksame Leser stellt sich hier die berechnete Frage, wozu ein neuer Schätzer konstruiert wird, obwohl bereits ein konsistenter zur Verfügung steht. Die Antwort darauf lautet, dass der Schätzer aus [11] zwar konsistent, jedoch in der Praxis nicht anzuwenden ist, da im allgemeinen $B(z)$ zweiseitig ist. Dadurch würde man für die Prognose auch zukünftige Beobachtungen benötigen, was für einen Prognosenschätzer das Ende bedeutet. Der Weg des obigen Schätzers führt aber auch zur Schätzung der spektralen Dichte des latenten Prozesses, welche in [12] den Grundstein zum Erfolg legt. Man definiere:

$$\hat{\Sigma}^x(\theta) = \hat{\lambda}_1(\theta)\hat{p}_1(\theta)\hat{p}_1(\theta)^* + \dots + \hat{\lambda}_q(\theta)\hat{p}_q(\theta)\hat{p}_q(\theta)^*, \quad (4.9)$$

$$\hat{\Sigma}^\varepsilon(\theta) = \hat{\lambda}_{q+1}(\theta)\hat{p}_{q+1}(\theta)\hat{p}_{q+1}(\theta)^* + \dots + \hat{\lambda}_n(\theta)\hat{p}_n(\theta)\hat{p}_n(\theta)^*, \quad (4.10)$$

wobei λ_i und p_i die Eigenwerte bzw. die Eigenvektoren des Schätzers $\hat{\Sigma}(\theta)$ für die spektrale Dichte des Prozesses x_t (mit festem n !) sind.

Das heißt, somit, dass $\check{\Sigma}^x(\theta)$ und $\check{\Sigma}^\varepsilon(\theta)$ konsistente Schätzer für

$$\check{\Sigma}^x(\theta) = \lambda_1(\theta)p_1(\theta)p_1(\theta)^* + \dots + \lambda_q(\theta)p_q(\theta)p_q(\theta)^* \quad (4.11)$$

$$\check{\Sigma}^\varepsilon(\theta) = \lambda_{q+1}(\theta)p_{q+1}(\theta)p_{q+1}(\theta)^* + \dots + \lambda_n(\theta)p_n(\theta)p_n(\theta)^* \quad (4.12)$$

mit n fest sind, da $\hat{\Sigma}(\theta)$ ein konsistenter Schätzer für $\check{\Sigma}(\theta)$ ist (Vergleiche Lemma(2.3.3) und Lemma(2.3.4)), die Annahme C1(b) die Identifikation von (4.11) und (4.12) gewährleistet, und die Eigenwerte und Eigenvektoren stetige Funktionen der Einträge der zugehörigen Matrizen sind⁸. In weiterer Folge sind $\hat{\Gamma}_k^x$ und $\hat{\Gamma}_k^\varepsilon$ ebenso konsistente Schätzer für

$$\check{\Gamma}_k^x = \int_{-\pi}^{\pi} e^{ik\theta} \check{\Sigma}^x(\theta) d\theta \text{ und} \quad (4.13)$$

$$\check{\Gamma}_k^\varepsilon = \int_{-\pi}^{\pi} e^{ik\theta} \check{\Sigma}^\varepsilon(\theta) d\theta \quad (4.14)$$

Schritt 2 Um die Notation zu vereinfachen bezeichne ab nun

$K_h = [\Gamma_h^x Z (Z' \Gamma_0 Z)^{-1}] Z'$, und dadurch \hat{K}_h die Projektion aus (4.7) bzw. (4.8).

⁸Für stetige Funktionen gilt der Satz von Slutsky (Satz 2.2.3).

Da durch die Annahme E(a), und die Bemerkung 1 von Theorem 2.1.4 die Identifikation der Matrix F (hier Z) garantiert ist, und zusätzlich $\hat{\Gamma}_0$ und $\hat{\Gamma}_k^\chi$ konsistente Schätzer sind, ist \hat{K}_h ein konsistenter Schätzer für \check{K}_h für $T \rightarrow \infty$. Das heißt, für festes n , $\epsilon > 0$ und $\eta > 0$, existiert ein $T_1(n, \epsilon, \eta)$, sodass für $T > T_1$

$$P \left(\sum_{j=1}^n |\hat{K}_{h,ij} - \check{K}_{h,ij}| > \epsilon \right) < \eta$$

gilt. Sei $M(n, \eta)$ für n und η fest so bestimmt, dass $P(\max_j |x_{jt}| \geq M(n, \eta)) < \eta$ gilt, dann existiert für gegebene n , ϵ und η ein $T_2(n, \epsilon, \eta)$ sodass für $T > T_2$

$$P(|\hat{\chi}_{i,T+h|T} - \check{\chi}_{i,T+h|T}| > \epsilon) = P \left(\left| \sum_{j=1}^n (\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt} \right| > \epsilon \right) < \eta \quad (4.15)$$

gilt und somit die Konsistenz von $\hat{\chi}_{T+h|T}$ gegen $\check{\chi}_{T+h|T}$ für $T \rightarrow \infty$ gegeben ist.

Beweis (4.15)

$$\begin{aligned} & P \left(\left| \sum_{j=1}^n (\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt} \right| > \epsilon \right) \leq P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij}) x_{jt}| > \epsilon \right) \\ &= P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij})| |x_{jt}| > \epsilon \right) \leq P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij})| \max_j |x_{jt}| > \epsilon \right) \\ &\leq P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij})| M(n, \eta/2) > \epsilon \text{ und } \max_j |x_{jt}| < M(n, \eta/2) \right) \\ &+ P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij})| \max_j |x_{jt}| > \epsilon \text{ und } \max_j |x_{jt}| \geq M(n, \eta/2) \right) \\ &\leq P \left(\sum_{j=1}^n |(\hat{K}_{h,ij} - \check{K}_{h,ij})| M(n, \eta/2) > \epsilon \right) + P \left(\max_j |x_{jt}| \geq M(n, \eta/2) \right) < \eta \end{aligned}$$

für $T_2(n, \epsilon, \eta) = T_1(n, \epsilon/M(n, \eta/2), \eta/2)$. □

Schritt 3 Der dritte Schritt ist der Kern des Beweises der FHLR-Methode, da in ihm die konsistente Schätzung des Faktorraumes gezeigt wird. Dazu sind einige Hilfslemmata nötig:

Lemma 4.1.2. *Es seien $\{A_n\}_{n=k,k+1,\dots}$ und $\{B_n\}_{n=k,k+1,\dots}$ (k fest) zwei Folgen von symmetrischen, positiv (semi-)definiten, $n \times n$ Matrizen (A_n positiv semi-definit, B_n positiv definit) für die gilt:*

- (i) *der k -t größte Eigenwert von A_n : μ_{nk} divergiert für $n \rightarrow \infty$,*
- (ii) *der größte Eigenwert von B_n ist nach oben beschränkt mit M .*

Dann divergiert auch der k -t größte verallgemeinerte Eigenwert von (A_n, B_n) : λ_{nk} für $n \rightarrow \infty$.

Lemma 4.1.3. *Seien $\check{\Gamma}_0^\varepsilon$ und $\check{\Gamma}_0^\chi$ die Matrizen aus (4.14) und (4.13) und $\check{\mu}_k^\varepsilon$ bzw. $\check{\mu}_k^\chi$ ihre zugehörigen Eigenwerte. Dann gilt unter den Annahmen A,B,C und D dass*

- (i) *$\check{\mu}_r^\chi \rightarrow \infty$, für $n \rightarrow \infty$ und*
- (ii) *$\check{\mu}_1^\varepsilon < M$ für $n \rightarrow \infty$,*

wobei $r = q(s + 1)$ die Anzahl der statischen Faktoren aus (4.3) sein soll.

Die beiden obigen Lemmata legen die Grundsteine, auf deren Resultate nun die beiden folgenden Lemmata aufbauend, die letzten Schwierigkeiten meistern.

Lemma 4.1.4. *Sei $a_n = (a_{n1}, \dots, a_{nn})$, $n \in \mathbb{N}$ eine „untere Dreiecksmatrix“ die $\lim_{n \rightarrow \infty} \sum_{i=1}^n a_{ni}^2 = 0$ erfüllt und gelten weiters die Annahmen von Satz (4.1.1), dann konvergiert*

$$a_n \varepsilon_t = (a_{n1}, \dots, a_{nn})(\varepsilon_{t1}, \dots, \varepsilon_{tn})' \rightarrow 0, \text{ für } n \rightarrow \infty$$

im quadratischen Mittel.

Bemerkung Lemma 4.1.4 spielt eine zentrale Rolle im Beweis der Konsistenz der Schätzer (4.7) und (4.8), da es die Konvergenz von $a_n x_t \rightarrow L(F|t)$ für $n \rightarrow \infty$ im quadratischen Mittel garantiert! (Vorausgesetzt a_n konvergiert nicht zu schnell gegen 0, sodass $a_n \chi_t$ nicht ebenso weggemittelt wird.)

Lemma 4.1.5. *Sei H ein Hilbertraum von zentrierten, quadratisch integrierbaren Zufallsvariablen, mit innerem Produkt: $\langle x, y \rangle = \mathbb{E}(xy)$. Weiters sei K ein Unterraum von H welcher durch das unabhängige k -Tupel $\mathbf{V} = (v_1, \dots, v_k)'$, $v_i \in H$ aufgespannt wird. Ebenso sei angenommen, dass $\{(\mathbb{V}_n = (v_{n1}, \dots, v_{nk})'), n \in \mathbb{N}\}$ eine Folge von ortho-normalen k -Tupeln aus H ist, sodass $(v_{nj} - \text{Proj}(v_{nj}|K))$, für $n \rightarrow \infty$ gegen 0 strebt. Dann gilt für $v \in H$:*

$$\text{Proj}(v|K_n) \rightarrow \text{Proj}(v|K), \text{ für } n \rightarrow \infty$$

im quadratischen Mittel, wobei K_n den durch \mathbb{V}_n aufgespannten Unterraum von H bezeichnet.

Durch die Lemmata (4.1.2) und (4.1.3) wird gezeigt, dass die r größten Eigenwerte $\check{\zeta}_i, i = 1, \dots, r$ von $(\check{\Gamma}_0^\chi, \check{\Gamma}_0^\varepsilon)$ für $n \rightarrow \infty$ divergieren.

Dadurch erfüllen $\left(\frac{\check{Z}_i}{\sqrt{1+\check{\zeta}_i}}\right)_{i=1, \dots, r}$ die Voraussetzung von Lemma (4.1.4), da

$$1 = \check{Z}'_i \check{\Gamma}_0^\varepsilon \check{Z}_i \geq \mu_n^\varepsilon \check{Z}'_i \check{Z}_i,$$

gilt.

Außerdem ist $\frac{\check{Z}_i}{\sqrt{1+\check{\zeta}_i}}$ so beschaffen, dass die $\text{Var}\left(\frac{\check{Z}_i}{\sqrt{1+\check{\zeta}_i}} \chi_t\right)$ nicht gegen 0 geht und somit

$\check{v}_{jt} = \frac{\check{Z}'_j x_t}{\sqrt{1+\check{\zeta}_j}}$ wegen Lemma (4.1.4) für $n \rightarrow \infty$ gegen $L(F|t)$ im quadratischen Mittel konvergiert.

Setzt man in Lemma (4.1.5) $v = \check{\chi}_{i,T+h|T}, K = L(F|t), K_n$ den durch $\check{v}_{jt}, j = 1, \dots, r$ aufgespannten Raum, so sieht man, dass $\check{\chi}_{i,T+h|T}$ gegen $\chi_{i,T+h|T}$ im quadratischen Mittel und daher auch in Wahrscheinlichkeit konvergiert. Das heißt, für gegebene ϵ und η existiert ein $N(\epsilon, \eta)$ sodass für $n > N$

$$P\left(|\check{\chi}_{i,T+h|T} - \chi_{i,T+h|T}| > \epsilon\right) < \eta \quad (4.16)$$

gilt.

Schritt 4 Durch (4.15) und (4.16) folgt nun, dass

$$P\left(|\hat{\chi}_{i,T+h|T} - \chi_{i,T+h|T}| > \epsilon\right) < \eta, \text{ für } n, T \rightarrow \infty \quad (4.17)$$

gilt, und somit Satz 4.1.1 bewiesen ist. \square

Bemerkung Es sei erwähnt, dass das Matrizenpaar $(\hat{\Gamma}_0^\varepsilon, \check{\Gamma}_0^\varepsilon)$ durch ein anderes, zum Beispiel (\hat{D}_n, \check{D}_n) ersetzt werden kann, solange die Bedingungen, die an das ursprüngliche Paar gesetzt wurden, auch weiterhin erfüllt bleiben: $\hat{D}_n - \check{D}_n$ konvergiert in Wahrscheinlichkeit gegen 0 für n gegeben und $T \rightarrow \infty$, und die Eigenwerte von \check{D}_n erfüllen die Annahme E. Der Grund dafür ist, dass die PCA und die verallgemeinerte PCA (Ω_n aus (3.8) frei wählbar!) asymptotisch dasselbe liefern.

4.2 Die Methode von Stock und Watson

Stock und Watson [SW] verfolgen in ihrem Paper [21] (*Forecasting using principal components from a large number of predictors*) den Plan, den Faktorraum mittels PCA zu schätzen. Die PCA dient dazu, die Linearkombinationen mit den größten Varianzen des stochastischen Vektors x_t zu finden. Damit kann eine „intuitive“ Aufteilung von x_t in den latenten und den Fehlerterm erreicht werden, indem man die r Linearkombinationen mit den größten Varianzen als Schätzer für die Faktoren des latenten Prozesses verwendet.

Nachdem der Faktorraum bestimmt ist, erfolgt die Prognose mittels linearer Regression des zu prognostizierenden Vektors auf den Faktorraum.

4.2.1 Das Modell

Es sei (x_t, y_t) ein beobachtbarer $(n+1)$ -dimensionaler Prozess, wobei x_t n -dimensional ist und der eindimensionale Prozess y_{t+h} prognostiziert werden soll (das schließt natürlich nicht aus, dass $y_{t+h} = x_{i,t+h}$, $i = 1, \dots, n$ sein kann).

Beiden Prozessen soll laut Voraussetzung die selbe Faktorstruktur zu Grunde liegen, welche sowohl dynamisch als auch statisch formuliert werden kann. Die Überlegungen basieren wieder auf der Idee, die Dynamik auf eine maximale Lagstruktur einzuschränken, um damit eine statische Formulierung zu ermöglichen.

A1 X ist eine endliche Realisierung des reellen, stationären Prozesses

$\{x_t = (x_{1t}, \dots, x_{nt})'; t \in \mathbb{Z}, n \in \mathbb{N}\}$, welcher Erwartungswert 0 und endliche zweite Momente $\Gamma_k = \mathbb{E}x_{t+k}x_t^*$ besitzt.

Ausgehend von der dynamischen Formulierung

$$x_t = B(z)f_t + e_t \quad (4.18)$$

$$y_{t+h} = \beta_f(z)'f_t + \beta_\omega' \omega_t + \varepsilon_{t+h} \quad (4.19)$$

wobei ω_t ein $m \times 1$ Vektor mit beobachtbaren Variablen, welche für die Vorhersage von y_{t+h} nützlich sind, ist, gelangt man durch folgende Annahme zur statischen Darstellungsform:

B1' (a) Die Lagpolynome $(b_{ij}(z))_{(i=1, \dots, n, j=1, \dots, q)}$ ⁹ und $\beta_f(z)$ aus (4.18) und (4.19) sollen Ordnung s haben. Das heißt, dass $B(z) = B_0 + \dots + B_s z^s$ eine Polynommatrix mit $B_s \neq 0$ – Matrix ist und $\beta_f(z) = \beta_0 + \beta_1 z + \dots + \beta_s z^s$ mit $\beta_s \neq 0$ gilt.

⁹ $b_{ij}(z)$ ist das (i,j) -Element von $B(z)$

(b) f_t folge einem VAR-Prozess:

$$A(z)f_t = u_t$$

mit $A(z) = A_0 + \dots + A_S z^S$ soll eine Polynommatrix der Ordnung $S \leq s$ sein, wobei wieder $A_S \neq 0$ – *Matrix* gelten soll.

(c) $\det(A(z)) \neq 0, \forall z \in \mathbb{C}$ mit $|z| \leq 1$.

SW setzen die Annahme B1' nicht so deutlich wie hier, da diese Annahme aber Sinn macht um eine verallgemeinerte dynamische Faktorstruktur zu gewährleisten, wird sie hier verlangt.

Dadurch kann man die gelagten Faktoren zu einem neuen "gestackten" Faktor $F_t = (f_t', f_{t-1}', \dots, f_{t-s}')'$ zusammenfassen.

$$x_t = CF_t + e_t, \tag{4.20}$$

$$y_{t+h} = \beta_F' F_t + \beta_\omega' \omega_t + \varepsilon_{t+h}, \tag{4.21}$$

wobei F_t der r-dimensionale ¹⁰ Faktorprozess und e_t bzw. ε_t die n- bzw. ein-dimensionalen idiosynkratischen Fehlerprozesse sind, für die

$$\begin{aligned} \mathbb{E}(F_{it}e_{jt}) &= 0, \forall i = 1, \dots, r; j = 1, \dots, n \\ \mathbb{E}(F_{it}\varepsilon_t) &= 0, \forall i = 1, \dots, r \\ \mathbb{E}(\omega_{lt}\varepsilon_t) &= 0, \forall l = 1, \dots, m \end{aligned}$$

gilt, womit die Orthogonalität von $\chi_t = B(z)f_t = CF_t$ auf e_t beziehungsweise ε_t gegeben ist. Stock und Watson verlangen diese Orthogonalität nicht, da sie in ihrem Paper auch den Fall eines stochastischen Drifts der Ladungsmatrizen über die Zeit betrachten, wodurch unter gewissen Annahmen die Faktoren mit den Fehlern leicht korrelieren dürfen. Da ich dies hier aber nicht bespreche, kann man die Orthogonalität voraussetzen.

Für diese statische Darstellungsform sollen nun folgende Annahmen zutreffen:

F1 $(C'C/n) \rightarrow I$.

F2 $\mathbb{E}(F_t F_t') = \Gamma_0^F$, wobei Γ_0^F Diagonalgestalt hat, und die Einträge der Größe nach, absteigend geordnet, und alle größer 0 sind.

F3 $|c_{ij}| \leq \bar{c} < \infty$.

¹⁰r = q(s+1), wobei q die Dimension des dynamischen Faktorprozesses f_t ist.

F4 $\frac{1}{T} \sum_{t=1}^T F_t F_t' \rightarrow \Gamma_0^F$ in Wahrscheinlichkeit.

Während die Annahmen F3 und F4 rein technischer Natur und nicht weiter einschneidend sind, liegt das Hauptaugenmerk auf den Annahmen F1 und F2. Zum einen garantieren die Regularitätsannahmen den Beitrag eines jeden Faktors an der durchschnittlichen Varianz von $x_{i,t}$, und zum anderen werden die Faktoren bis auf das Vorzeichen identifiziert.

Erstes sieht man folgendermaßen (sei μ die vorhin erwähnte durchschnittliche Varianz):

$$\begin{aligned} \mu &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \text{Var}(x_{i,t}) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\text{Var}(\chi_{i,t}) + \text{Var}(e_{i,t})) \\ &\geq \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \mathbb{E}(F_t' C_i' C_i F_t) \xrightarrow{n \rightarrow \infty (F1)} \frac{1}{T} \sum_{t=1}^T \mathbb{E}(F_t' F_t) \xrightarrow{T \rightarrow \infty (F4)} \sum_{j=1}^r \text{Var}(F_j) \end{aligned}$$

□

Zweites überlegt man sich in zwei Schritten. Sei $x_t = CR^{-1}RF_t + e_t$ eine äquivalente Darstellungsform von (4.20) mit beliebiger regulärer Matrix R und r-dimensionalem Faktorprozess $\tilde{F}_t = RF_t$, dann schränkt F1 zuerst R auf die Klasse der orthogonalen Matrizen ein¹¹:

$$\begin{aligned} \tilde{C} = CR^{-1} &\Rightarrow \tilde{C}'\tilde{C}/n = (R^{-1})'C'CR^{-1}/n \xrightarrow{F1} I \\ &\Rightarrow (R^{-1})'R^{-1} \rightarrow I \Leftrightarrow R \dots \text{orthogonal} \end{aligned}$$

und anschließend legt F2, R auf eine Diagonalmatrix mit (+/- 1)-Elementen fest:

Sei $\Gamma_0^{\tilde{F}}$... diag mit

$$\Gamma_0^{\tilde{F}} = \mathbb{E}(\tilde{F}_t \tilde{F}_t') = \mathbb{E}(RF_t F_t' R') = R\Gamma_0^F R' \text{ und}$$

man betrachte die Orthogonalität der Zeilen von R (r_i sei die i-te Zeile):

Da $\Gamma_{0;(i,j)}^{\tilde{F}} = r_i \Gamma_0^F r_j' = 0$ für $i \neq j$ gilt, wird die Orthogonalität der r_i durch die lineare Transformation: $r_i \mapsto r_i (\Gamma_0^F)^{\frac{1}{2}}$ übertragen. Dies kann aber nur zutreffen, wenn die r_i kanonisch sind. Wären zum Beispiel r_i und r_j nicht kanonisch, so müsste $(\Gamma_0^F)^{\frac{1}{2}}$

¹¹Dies gilt bereits im endlichen (und benötigt nicht F1): Sei A eine symmetrische, positiv semi-definite $n \times n$ Matrix mit Rang r, so existieren Matrizen $O = (O_1, O_2)$ und Λ mit O orthogonal, $\dim(O_1) = n \times r$, und Λ diagonal, sodass $A = O \Lambda O'$ gilt. $C := O_1$ erfüllt das gewünschte.

entweder eine orthogonale Rotation oder $c \cdot I$ sein. Da beide Varianten nicht zutreffen können folgt die Behauptung.

Durch diese Identifizierung der Faktoren (bis auf das Vorzeichen) muss man aber rückwirkend den Übergang von der dynamischen auf die statische Darstellungsform noch einmal kurz überdenken. Es kann nämlich nicht angenommen werden, dass der gestackte Faktorprozess F_t aus (4.20) auch tatsächlich jener ist, der Annahme F2 erfüllt. Wenn jedoch $\check{F}_t = (f'_t, f'_{t-1}, \dots, f'_{t-s})'$ aus (4.20) linear unabhängig und der dynamische Faktorprozess f_t stationär ist, so existiert $\Gamma_0^{\check{F}}$ und ist zugleich positiv definit. Das heißt jedoch, dass eine orthogonale Matrix R existiert, sodass $F_t = R\check{F}_t$ die Annahme F erfüllt. In weiterer Folge sind die statischen Faktoren als eben jene zu verstehen, die Annahme F erfüllen.

Bis jetzt wurden noch keine Annahmen an den Fehlerprozess $\{(e_{1t}, \dots, e_{nt}), t \in \mathbb{Z}\}_{n \in \mathbb{N}}$ getroffen, und würde man voraussetzen, dass dieser iid (im Querschnitt) und normalverteilt wäre, würde man im klassischen Faktormodell landen. Um genau das zu verhindern, lässt man nun folgende Verallgemeinerungen zu:

$$\text{G1} \quad \limsup_{n \rightarrow \infty} \sum_t \sum_{s=-\infty}^{\infty} |\gamma_{n,t}(s)| \leq \infty, \text{ mit } \mathbb{E}(e'_t e_{t+s})/n = \gamma_{n,t}(s).$$

$$\text{G2} \quad \limsup_{n \rightarrow \infty} \sum_t \sum_{i=1}^n \sum_{j=1}^n |\tau_{ij,t}|/n \leq \infty, \text{ mit } \mathbb{E}(e_{it} e_{jt}) = \tau_{ij,t}.$$

$$\text{G3} \quad \limsup_{n \rightarrow \infty} \sum_{t,s} \sum_{i=1}^n \sum_{j=1}^n |\text{cov}(e_{is} e_{it}, e_{js} e_{jt})|/n \leq \infty.$$

Diese Annahmen sehen auf den ersten Blick sehr technisch aus, auf den zweiten Blick relativiert sich dies jedoch. G1 und G2 lassen zum einen Abhängigkeiten über die Zeit (G1) und zum anderen Kreuzkorrelationen der Fehler zu (G2). Diese beiden Annahmen sind derart fundamental, dass sie das verallgemeinerte Faktormodell charakterisieren.

Nicht beunruhigen sollte einen die Tatsache, dass in den beiden Annahmen ein Supremum über die Zeit inkludiert ist. Dies ist für e_t stationär natürlich überflüssig, Stock und Watson zeigen aber weiters, dass selbst im Falle eines stochastischen Drifts der Ladungsmatrizen über die Zeit, die folgenden Sätze gelten. Für eben diesen Fall ist dieses Supremum derart wichtig dass ich die Annahmen so stehen lassen möchte, gerade weil es in den folgenden Beweisen nicht benötigt wird, aber keine großartige Einschränkung darstellt.

Weitaus technischer ist Annahme G3, welche beschränkte, vierte Momente der Fehler verlangt.

G2 erinnert in einer gewissen Weise bereits an die Annahme der Beschränktheit des größten Eigenwertes der Fehler-Varianz-Kovarianzmatrix:

$$\sum_{i=1}^n \sum_{j=1}^n |\tau_{ij,t}|/n = \mathbf{1}'|\Gamma_0^e|\mathbf{1}/n = \mathbf{1}'/\sqrt{n}|\Gamma_n^e|\mathbf{1}/\sqrt{n}' \leq \infty.$$

Es fehlt jedoch die Tatsache, dass dies für alle normierten Vektoren γ mit $\|\gamma\| = 1$ gilt, sondern nur für den konstanten Vektor $\mathbf{1}/\sqrt{n}$. Ein Resultat als Folge des Beweises der konsistenten Schätzung der Faktoren, wird dies aber liefern.

Die Unbeschränktheit der r größten Eigenwerte von Γ_0^x ist einfacher einzusehen. Für den latenten Prozess gilt:

$$\mathbb{E}(CF_t F_t' C') = C\Gamma_0^F C' = O\Lambda O',$$

wobei O aus den normierten Eigenvektoren besteht. Da jedoch aus Annahme F1 die Normierung $C/\sqrt{n} = O$ folgt, muss auch $n\Gamma_0^F = \Lambda$ und dadurch die Divergenz der Eigenvektoren des latenten Prozesses gelten, was natürlich die Divergenz der Eigenwerte des tatsächlichen Prozesses impliziert.

Bis jetzt lag die Konzentration der Annahmen auf den Faktoren und den Ladungsmatrizen. Um einen konsistenten Prädiktor für die Gleichung (4.21) gewährleisten zu können, werden folgende Annahmen benötigt:

H1 $\mathbb{E}((F_t', w_t')'(F_t', w_t')) = \mathbb{E}(z_t z_t') = \Gamma_0^z$ sei positiv definit.

H2 $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T z_t z_t' = \Gamma_0^z.$

H3 $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T z_t \varepsilon_{t+h} = 0.$

H4 $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \varepsilon_{t+h}^2 = \sigma^2.$

H5 $\|(\beta_F', \beta_w')\| < \infty.$

Dies sind keine außergewöhnlichen Annahmen, vorallem da H1, H2 und H3 übliche OLS - Annahmen sind (unter Voraussetzung das F_t bekannt sei). Die beiden übrigen Annahmen sind eigentlich rein technischer Natur und stellen keine wirklichen Einschränkungen dar.

4.2.2 Der Schätzer

Der Prädiktor für y_{t+h} aus (4.21) erfolgt, wie die Gleichungen (4.20) und (4.21) bereits erahnen lassen, in zwei Schritten. Da weder F_t noch C aus (4.20) bekannt sind, werden zuerst die den Zeitreihen (x_t, y_{t+h}) gemeinsamen Faktoren F_t geschätzt, und anschließend die Prognose y_{t+h} durch OLS-Regression von y_{t+h} auf F_t , und gegebenenfalls bekannte Zeitreihen w_t , bestimmt.

Die zu minimierende Zielfunktion lautet:

$$(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (x_{it} - C_i F_t)^2 = (nT)^{-1} \text{tr} [(X - FC')(X - FC)'] \quad (4.22)$$

mit

$$X = (x_1, \dots, x_T)' \text{ und } F = (F_1, \dots, F_T)'$$

Löst man dieses Problem für F unter der Annahme, dass C beliebig aber fest ist, indem man (4.22) nach F ableitet und den Gradienten anschließend 0 setzt, so bekommt man

$$F = XC(C'C)^{-1} = XC/n(C'C/n)^{-1} = XC/n. \quad (4.23)$$

Setzt man dies wiederum in die Zielfunktion (4.22) ein und führt ein paar einfache Umformungen durch, gelangt man zu dem Maximierungsproblem

$$\max_{C'C/n=1} (nT)^{-1} \text{tr} [C'X'XC]. \quad (4.24)$$

Diese Maximierungsaufgabe entspricht jedoch dem klassischen PCA-problem und wird dadurch gelöst, indem man für C die ersten r Eigenvektoren der Matrix $T^{-1}X'X$ nimmt. Das heißt, für F_t ergibt sich

$$\hat{F}_t = \hat{C}'x_t/n. \quad (4.25)$$

Zu beachten ist, dass die Spalten von C (und auch die von \hat{C}) auf \sqrt{n} normiert sind, wodurch das Dividieren durch n in (4.25) Sinn macht

$$\hat{C}\hat{F}_t = (\hat{C}\hat{C}'/n)x_t.$$

Diese Überlegungen führen zu dem folgenden Satz, der die konsistente Schätzung der Faktoren (bis auf das Vorzeichen) beschreibt.

Satz 4.2.1. *Sei S_i eine Variable, die je nach Bedarf +/-1 annimmt, und sei angenommen, dass die Annahmen A1, B1', F und G zutreffen, dann gilt*

$$\text{plim}_{n,t \rightarrow \infty} S_i \hat{F}_{it} = F_{it}; \text{ für } i = 1, \dots, r.$$

Obwohl Stock und Watson in ihrem Paper [21] auch noch zeigen, dass der quadratische Fehler über die Zeit, in Wahrscheinlichkeit gegen 0 konvergiert, möchte ich mich auf die Konsistenz, aus Gründen der Übersichtlichkeit, beschränken.

Die Beweise der folgenden Hilfslemmata werde ich im Appendix nachbringen. Desweiteren sollen die Annahmen A1, B1', F und G stets erfüllt sein.

Zusätzlich werden folgende Definitionen verwendet:

$$H = \{\gamma | \gamma' \gamma / n = 1\}; \quad e = (e_1, \dots, e_T)'; \quad \hat{\Gamma}^e = T^{-1} e' e$$

$$R(\gamma) = (n^2 T)^{-1} \gamma' \sum_{t=1}^T x_t x_t' \gamma; \quad R(\gamma)^* = (n^2 T)^{-1} \gamma' \sum_{t=1}^T C F_t F_t' C' \gamma$$

Lemma 4.2.2. $\text{plim}_{n, T \rightarrow \infty} \sup_{\gamma \in H} n^{-2} \gamma' \hat{\Gamma}^e \gamma = 0$

Dieses Lemma zeigt die bereits weiter oben erwähnte Beschränktheit der Eigenwerte der geschätzten Fehler-Varianz-Kovarianzmatrix.

Nun kann folgende wichtige Eigenschaft gezeigt werden:

Lemma 4.2.3. $\text{plim}_{n, t \rightarrow \infty} \left| \sup_{\gamma \in H} R(\gamma) - \sup_{\gamma \in H} R(\gamma)^* \right| = 0$

Das heißt, der größte Eigenwert der vorhandenen Matrix $T^{-1} X' X$ konvergiert in Wahrscheinlichkeit gegen den größten Eigenwert der nicht beobachtbaren Matrix $T^{-1} C F' F C'$.

Diese Resultate führen dazu, dass das nachfolgende Lemma gültig und dadurch der Beweis von Satz(4.2.1) geebnet ist.

Lemma 4.2.4. *Sei \hat{C} die Matrix der ersten r Eigenvektoren von $T^{-1} X' X$ mit der Normierung, dass $\hat{C}' \hat{C} / n = I$ ist. Sei weiters $S = \text{diag}(\text{sign}(\hat{C}' C))$, dann gilt*

$$\text{plim}_{n, t \rightarrow \infty} S \hat{C}' C / n = I.$$

Somit hat man das nötige Werkzeug zusammen, um die konsistente Schätzung der Faktoren zu beweisen.

Beweis (Satz 4.2.1)

$$\begin{aligned} S_i \hat{F}_{it} - F_{it} &= S_i n^{-1} \hat{C}'_i x_t - F_{it} \\ &= S_i n^{-1} \hat{C}'_i (C F_t + e_t) - F_{it} = S_i n^{-1} \hat{C}'_i C F_t - F_{it} - n^{-1} S_i \hat{C}'_i e_t \end{aligned}$$

Da $\text{plim}_{t \rightarrow \infty} |F_t| = O(1)$, da $\mathbb{E}(F_t F_t') = \Gamma_0^F$ wegen Annahme F2 gilt, konvergiert $S_i n^{-1} \hat{C}_i' C F_t - F_{it}$ wegen Lemma(4.2.4) in Wahrscheinlichkeit gegen 0. Das heißt, man muss noch $\text{plim}_{n, T \rightarrow \infty} n^{-1} S_i \hat{C}_i' e_t = 0$ zeigen.

$$S_i \hat{C}_i' e_t / n = (S_i \hat{C}_i' - C_i') e_t / n + C_i' e_t / n$$

Der erste Teil konvergiert, da

$$|(S_i \hat{C}_i' - C_i') e_t / n| \leq \left((S_i \hat{C}_i' - C_i')' (S_i \hat{C}_i - C_i) / n \right)^{1/2} (e_t' e_t / n)^{1/2}$$

gilt, die hintere Klammer wegen Korollar(7.0.1) mit Wahrscheinlichkeit 1 beschränkt bleibt, und für die vordere Klammer gilt

$$(S_i \hat{C}_i' - C_i')' (S_i \hat{C}_i - C_i) / n = \hat{C}_i' \hat{C}_i / n + C_i' C_i / n - 2(S_i \hat{C}_i' C_i / n) = 2 - 2(S_i \hat{C}_i' C_i / n) \rightarrow 0$$

wegen Lemma(4.2.4).

Das Wurzelziehen ändert an den Konvergenzen nichts, da die Wurzel eine stetige Funktion ist und die Konvergenz in Wahrscheinlichkeit durch den Satz von Slutsky (Satz 2.2.3) übertragen wird.

Der zweite Teil konvergiert sogar im quadratischen Mittel gegen 0, da

$$\begin{aligned} \mathbb{E} (C_i' e_t / n)^2 &= n^{-2} \sum_{k=1}^n \sum_{j=1}^n \mathbb{E} c_{ik} c_{ij} (e_{kt}, e_{jt}) = n^{-2} \sum_{k=1}^n \sum_{j=1}^n c_{ik} c_{ij} \tau_{kj,t} \\ &\leq \bar{c}^2 n^{-2} \sum_{k=1}^n \sum_{j=1}^n |\tau_{kj,t}| \rightarrow 0. \end{aligned}$$

Die Ungleichung gilt wegen Annahme F3 und die Konvergenz folgt aus Annahme G2.

□

Mit der Gültigkeit des Satzes(4.2.1) ist der Großteil des Vorhabens erledigt. Die fehlende konsistente Schätzung des latenten Prozesses muss zwar noch gezeigt werden, ist aber vom Aufwand wesentlich geringer als die Schätzung der Faktoren.

Satz 4.2.5. Sei $\hat{\beta} = (\hat{\beta}_F, \hat{\beta}_\omega)$ der OLS-Schätzer für β aus (4.21), der durch Regression von y_{t+h} auf die geschätzten Faktoren $\hat{F}_t = \hat{C}' x_t$ bestimmt wird, dann gilt unter den Annahmen A, B, F, G und H

1. $\text{plim}_{n, T \rightarrow \infty} (S \hat{\beta}_F, \hat{\beta}_\omega) = \beta$.
2. $\text{plim}_{n, T \rightarrow \infty} \hat{\beta}' (\hat{F}_t', \omega_t')' = \beta' (F_t', \omega_t')'$.

Das heißt, sowohl die Koeffizienten β als auch der latente Prozess $\chi_{t+h} = \beta'(F'_t, \omega'_t)'$ werden konsistent geschätzt.

Beweis (Satz 4.2.5)

1.

$$\begin{aligned}
\hat{\beta} &= \left(\begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} y_{t+h} \\
&= \left(\begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} ((F, W)\beta + \varepsilon_{t+h}) \\
&= \left(\begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (F, W)\beta + \left(\begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} \varepsilon_{t+h} \\
&= \left(T^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} T^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (F, W)\beta \\
&\quad + \left(T^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} (S\hat{F}, W) \right)^{-1} T^{-1} \begin{pmatrix} S\hat{F}' \\ W' \end{pmatrix} \varepsilon_{t+h} \\
&= \left(\begin{pmatrix} T^{-1}\hat{F}'\hat{F} & T^{-1}S\hat{F}'W \\ T^{-1}W'S\hat{F} & T^{-1}W'W \end{pmatrix} \right)^{-1} \begin{pmatrix} T^{-1}S\hat{F}'F & T^{-1}S\hat{F}'W \\ T^{-1}W'F & T^{-1}W'W \end{pmatrix} \beta \\
&\quad + \left(\begin{pmatrix} T^{-1}\hat{F}'\hat{F} & T^{-1}S\hat{F}'W \\ T^{-1}W'S\hat{F} & T^{-1}W'W \end{pmatrix} \right)^{-1} \begin{pmatrix} T^{-1}S\hat{F}'\varepsilon_{t+h} \\ T^{-1}W'\varepsilon_{t+h} \end{pmatrix}
\end{aligned}$$

Mit Hilfe von Korollar(7.0.10) (1., 2., 3.) kommt man bereits auf

$$\begin{aligned}
plim_{n,T \rightarrow \infty} \hat{\beta} &= \beta \\
&\quad + plim_{n,T \rightarrow \infty} \left(\begin{pmatrix} T^{-1}\hat{F}'\hat{F} & T^{-1}S\hat{F}'W \\ T^{-1}W'S\hat{F} & T^{-1}W'W \end{pmatrix} \right)^{-1} \begin{pmatrix} T^{-1}S\hat{F}'\varepsilon_{t+h} \\ T^{-1}W'\varepsilon_{t+h} \end{pmatrix}
\end{aligned}$$

und da wegen Korollar(7.0.10) (4.) und Annahme H3

$$plim_{n,T \rightarrow \infty} \begin{pmatrix} T^{-1}S\hat{F}'\varepsilon_{t+h} \\ T^{-1}W'\varepsilon_{t+h} \end{pmatrix} = 0$$

gilt, folgt

$$plim_{n,T \rightarrow \infty} \hat{\beta} = \beta$$

da die Matrix Γ_0^z wegen Annahme H1 regulär ist, und somit der Satz von Slutsky (Satz 2.2.3) die Konvergenz überträgt. \square

2. Sei $\tilde{S} = \begin{pmatrix} S & 0 \\ 0 & I \end{pmatrix}$, so gilt

$$\begin{aligned} \hat{\beta}' \hat{z}_t - \beta' z_t &= \hat{\beta}'(\hat{F}_t', \omega_t')' - \beta' z_t = (\tilde{S}\hat{\beta})' \tilde{S}\hat{z}_t - \beta' z_t \\ &= (\tilde{S}\hat{\beta})' \tilde{S}\hat{z}_t - \beta' z_t + (\tilde{S}\hat{\beta})' z_t - (\tilde{S}\hat{\beta})' z_t = (\tilde{S}\hat{\beta} - \beta)' z_t + (\tilde{S}\hat{\beta})' (\tilde{S}\hat{z}_t - z_t). \end{aligned}$$

Da nun z_t wegen Annahme H1 mit Wahrscheinlichkeit 1 beschränkt ist und $\text{plim}_{n,T \rightarrow \infty} \tilde{S}\hat{\beta} = \beta$ wegen (1.) gilt, strebt der erste Ausdruck gegen 0. Da der zweite Ausdruck nur noch die Faktoren enthält, strebt dieser wegen Satz(4.2.1) ebenfalls gegen 0.

$$\Rightarrow \text{plim}_{n,T \rightarrow \infty} (\hat{\beta}' \hat{z}_t - \beta' z_t) = 0$$

□

4.3 Vergleich der beiden Methoden

In diesem Abschnitt möchte ich die beiden eben vorgestellten Methoden vom theoretischen Standpunkt miteinander vergleichen.

Das Modell

Beiden Methoden liegt ein dynamisches Faktormodell zu Grunde, welches Abhängigkeiten der idiosynkratischen Fehler sowohl im Querschnitt als auch über die Zeit zulässt. Ein dynamisches Modell

$$x_t = \sum_{j=-\infty}^{\infty} b_j f_{t-j}$$

lässt die unendliche Vergangenheit und Zukunft zu. Da zukünftige Werte für Prognosezwecke naturgemäß Fehl am Platz sind, wird im Allgemeinen eine kausale Struktur angenommen.

$$x_t = \sum_{j=0}^{\infty} b_j f_{t-j} \quad (4.26)$$

Ohne Bedingung der Allgemeinheit kann f_t als Weißes Rauschen angenommen werden.

Hier setzen beide Methoden ihre erste wichtige Annahme. Sie gehen davon aus, dass der beobachtbare Prozess x_t eine endliche Faktorstruktur zulässt

$$x_t = \sum_{j=0}^s b_j f_{t-j}. \quad (4.27)$$

Jedoch gelangt man auf die obige Darstellung, falls f_t beispielsweise ein autoregressiver Prozess ist, dessen Lagpolynom nur Nullstellen außerhalb des Einheitskreises besitzt.

$$A(z)f_t = u_t, A(z) \neq 0 \text{ für } \|z\| \leq 1$$

Somit kann man nämlich weiterschreiben

$$x_t = \sum_{j=0}^s b_j f_{t-j} = \sum_{j=0}^s b_j A(z)^{-1} u_{t-j} = \sum_{j=0}^{\infty} c_j u_{t-j}$$

und man hat wieder die Darstellung wie in (4.26), jedoch mit dem wesentlichen Vorteil, dass man auch die Darstellung aus (4.27) verwenden kann. Diese wiederum lässt eine

statische Formulierung zu, welche nunmal wesentlich einfacher zu handhaben ist als eine dynamische, aber dennoch die eingeschränkte Dynamik zu lässt.

$$x_t = CF_t, \text{ mit } F_t = (f'_t, f'_{t-1}, \dots, f'_{t-s})' \quad (4.28)$$

Die Annahmen

Die Annahmen sehen in der Notation komplett unterschiedlich aus, durch geeignete Überlegungen ähneln sie einander doch sehr stark. Vorallem der Bezug zu den Bedingungen von Chamberlains und Rothschilds Darstellungstheorie ist nicht wegzuleugnen. FHLR haben zwar die Bedingungen vor allem auf die Eigenwerte der spektralen Dichte angewandt, doch wird mit Bedingung D und Lemma(4.1.3) der Bezug auch auf die statische Formulierung klar ersichtlich. SW sorgen mit ihrer Annahme F1 und Lemma(4.2.2) dafür, dass Chamberlains und Rothschilds Bedingungen erkennbar sind.

Der Schätzer

Die grundlegende Idee ist auch hier bei beiden Methoden die selbe. Zuerst soll der durch die Faktoren aufgespannte Raum geschätzt werden und anschließend auf diese regressiert werden. Doch bei der Bestimmung des Faktorraumes kommen nun unterschiedliche Methoden zur Verwendung.

Stock und Watson verfolgen, wie auch Chamberlain und Rothschild, die jedoch nicht auf zeitliche Abhängigkeiten eingegangen sind, die Faktoren mittels „Principal Components“ zu schätzen.

Forni, Hallin, Lippi und Reichlin hingegen verwenden „generalized Principal Components“, deren Verwendbarkeit auch Chamberlain und Rothschild bereits, wieder jedoch ohne zeitliche Abhängigkeiten, gezeigt haben, um den Faktorraum zu bestimmen.

Dass Stock und Watson eine eigene Prognosegleichung formulieren, ist kein Unterschied zu Forni und Co., da diese nicht separat auf die Existenz exogener Variablen für die Prognose eingehen.

Principal Components

Wie bereits erwähnt, verwenden beide Methoden den Ansatz der Principal Components, um die Faktoren zu schätzen. Der Unterschied liegt darin, dass FHLR bereits zuvor mit Hilfe von Principal Components die Varianz-Kovarianz-Matrizen des latenten und des Fehlerprozesses schätzen, indem sie PCA im dynamischen Fall wie in

(2.4.2) verwenden. Da dies konsistente Schätzer sind, verwenden sie diese geschätzten Varianz-Kovarianz-Matrizen, um die „generalized Principal Components“ (3.8) anzuwenden.

Obwohl beide Methoden asymptotisch das Gleiche liefern, erhoffen sich FHLR einen Effizienzgewinn bei Prozessen, bei denen die Dynamik stärker ausgeprägt ist. Das heißt, dass zum Beispiel ein dynamischer Faktor mit bis zu einem Lag von 3 relevant ist. Die Effizienzsteigerung soll dadurch entstehen, dass die geschätzten Varianz-Kovarianz-Matrizen an Bedeutung gewinnen je dynamischer das Modell ist, da sie mit dem dynamischen PCA Ansatz geschätzt und in weiterer Folge bei der verallgemeinerten PCA verwendet werden.

Kapitel 5

Monte Carlo Simulation

In diesem Kapitel möchte ich die beiden Schätzmethoden mit Hilfe von simulierten Daten miteinander vergleichen. Dazu verwende ich 5 verschiedene Modelle, beginnend bei einem schon fast statischen Modell und endend mit einem hochdynamischen. Desweiteren teile ich die Berechnung der Resultate in zwei Bereiche auf. Zuerst werden die in- bzw. out-of-sample Schätzer mit den wahren statischen und dynamischen Faktoranzahlen geschätzt. Danach werden die Resultate präsentiert, denen eine Schätzung der Faktoranzahlen vorausgeht.

Jedes der 5 Modelle wird des Weiteren mit allen Kombinationen aus $T = 30, 200$ und $n = 50, 100, 150$ durchgerechnet. Das Zeitintervall 200 stellt einen repräsentativen Zeithorizont in der Finanzmathematik dar. So sind 200 tägliche Handelstage fast ein ganzes Jahr und 200 wöchentliche Handelstage können einen ganzen Konjunkturzyklus beinhalten. 30 Zeitpunkte hingegen sind in etwa das, mit dem Ökonomen, die in der Regel nur jährliche Daten zur Verfügung haben, vorlieb nehmen müssen. Die Dimension n des beobachtbaren Vektors x wurde willkürlich gewählt.

Modell 1

Das erste Modell ist ein äußerst simples. Es wird angenommen, dass ein autoregressiver Faktor mit einem Lag den latenten Prozess treibt und der Fehlerprozess White

Noise ist.

$$\begin{aligned}
 x_t &= \chi_t + \varepsilon_t = \Lambda F_t + \varepsilon_t \\
 \chi_{it} &= \sum_{j=0}^1 \lambda_{ij} f_{t-j} \\
 f_t &= 0,5 f_{t-1} + \xi_t \\
 F_t &= (f_t, f_{t-1})'.
 \end{aligned}$$

ε_{it} und λ_{ij} sind unabhängig $N(0/1)$. Das heißt, die Anzahl der Faktoren ist somit $q = 1$, $s = 1$ und somit $r = 2$.

χ_t und ε_t werden sowohl hier als auch in allen darauffolgenden Modellen auf $Var(\chi_i) = Var(\varepsilon_i) = 0,5$ genormt.

Modell 2

Das zweite Modell wird nun durch 2 MA-Prozesse generiert. Der Fehlerprozess folgt auch hier einem White Noise Prozess.

$$\begin{aligned}
 x_t &= \chi_t + \varepsilon_t = \Lambda F_t + \varepsilon_t \\
 \chi_{it} &= \sum_{j=0}^2 \lambda_{1,ij} u_{1,t-j} + \sum_{j=0}^2 \lambda_{2,ij} u_{2,t-j} \\
 F_t &= (u_{1,t}, u_{1,t-1}, u_{1,t-2}, u_{2,t}, u_{2,t-1}, u_{2,t-2})'.
 \end{aligned}$$

ε_{it} , $u_{1,t}$, $u_{2,t}$, $\lambda_{1,ij}$ und $\lambda_{2,ij}$ sind unabhängig $N(0/1)$. Dadurch ergeben sich $q = 2$, $s = 2$ und $r = 6$.

Modell 3

Modell 3 wird ebenfalls wieder durch 2 MA-Prozesse generiert. Diesmal aber bis zu Lag 3, und der Fehlerprozess besitzt nun Kreuzkovarianzen ungleich 0.

$$\begin{aligned}
 x_t &= \chi_t + \varepsilon_t = \Lambda F_t + \varepsilon_t \\
 \chi_{it} &= \sum_{j=0}^3 \lambda_{1,ij} u_{1,t-j} + \sum_{j=0}^3 \lambda_{2,ij} u_{2,t-j} \\
 F_t &= (u_{1,t}, u_{1,t-1}, u_{1,t-2}, u_{1,t-3}, u_{2,t}, u_{2,t-1}, u_{2,t-2}, u_{2,t-3})' \\
 \varepsilon_{it} &= a_{i,t} + 0,5 (a_{i-1,t} + a_{i+1,t}). \tag{5.1}
 \end{aligned}$$

a_{it} , $u_{1,t}$, $u_{2,t}$, $\lambda_{1,ij}$ und $\lambda_{2,ij}$ sind unabhängig $N(0/1)$. Dadurch ergeben sich $q = 2$, $s = 3$ und $r = 8$.

Modell 4

Modell 4 wird durch 2 AR(1)-Prozesse generiert, und der Fehlerprozess gleicht dem aus Modell 3.

$$\begin{aligned}x_t &= \chi_t + \varepsilon_t = \Lambda F_t + \varepsilon_t \\ \chi_{it} &= \sum_{j=0}^1 \lambda_{1,ij} f_{1,t-j} + \sum_{j=0}^1 \lambda_{2,ij} f_{2,t-j} \\ f_t &= 0,5 f_{t-1} + \xi_t \\ F_t &= (f_{1,t}, f_{1,t-1}, f_{2,t}, f_{2,t-1})' \\ \varepsilon_{it} &= a_{i,t} + 0,5 (a_{i-1,t} + a_{i+1,t}).\end{aligned}\tag{5.2}$$

a_{it} , $\lambda_{1,ij}$ und $\lambda_{2,ij}$ sind unabhängig $N(0/1)$. Dadurch ergeben sich $q = 2$, $s = 1$ und $r = 4$.

Modell 5

Das fünfte Modell wird durch einen AR(2)-Prozess generiert, welcher bis zu einem Lag 3 auftritt. Der Fehlerprozess bekommt nun auch leichte Abhängigkeiten über die Zeit.

$$\begin{aligned}x_t &= \chi_t + \varepsilon_t = \Lambda F_t + \varepsilon_t \\ \chi_{it} &= \sum_{j=0}^3 \lambda_{ij} f_{t-j} \\ f_t &= 0,5 f_{t-1} + 0,2 f_{t-2} + \xi_t \\ F_t &= (f_t, f_{t-1}, f_{t-2}, f_{t-3})' \\ \varepsilon_{it} &= a_{i,t} + 0,5 (a_{i-1,t} + a_{i+1,t}) + 0,2 a_{i,t-1}.\end{aligned}\tag{5.3}$$

a_{it} und λ_{ij} sind unabhängig $N(0/1)$. Dadurch ergeben sich $q = 1$, $s = 3$ und $r = 4$.

Überprüfung der Bedingungen

Ich möchte die Bedingungen von Chamberlain und Rothschild überprüfen, da sie einen guten Überblick darstellen, und die meisten technischen Annahmen der Methoden von Forni und CO. bzw Stock und Watson leicht zu überprüfen, beziehungsweise durch simple weiterfolgende Überlegungen nachzuvollziehen sind.

Die Bedingung der Divergenz der ersten r Eigenwerte

$$\sup_n \mu_{nr} = \infty$$

ist durch die Konstruktion von χ_t gewährleistet ($\Lambda_{ij} \sim N(0/1)$). Desweiteren ist

$$\inf_n \mu_{nn} > 0$$

durch die Konstruktion von ε_t erfüllt. Somit bleibt zu zeigen, dass

$$\mu_{r+1} < \infty$$

gilt.

Da $\mu_{r+1}^x = 0$ laut Konstruktion und wegen Satz(2.1.15)

$$\mu_{r+1} < \mu_1^\varepsilon$$

gilt, muss man nur noch die Beschränktheit von μ_1^ε zeigen. Das heißt, der größte Eigenwert von Γ_0^ε muss beschränkt sein. Da Γ_0^ε stets eine Bandmatrix ist (in dieser Simulation), folgt aus Korollar(2.1.16), dass ihr größter Eigenwert durch die Summe einer ihrer mittleren Zeilen beschränkt ist. \square

Kriterien

Die zu vergleichenden in- bzw out-ofsample Schätzer lauten:

$$\begin{aligned}\hat{\chi}_t^{FHLR} &= [\hat{\Gamma}_0^x \hat{Z} (\hat{Z}' \hat{\Gamma}_0 \hat{Z})^{-1}] \hat{Z}' x_t, t = 1, \dots, T, \\ \hat{\chi}_t^{SW} &= [\hat{\Gamma}_0 \hat{C} (\hat{C}' \hat{\Gamma}_0 \hat{C})^{-1}] \hat{C}' x_t, t = 1, \dots, T,\end{aligned}$$

$$\begin{aligned}\hat{\chi}_{T+h|T}^{FHLR} &= [\hat{\Gamma}_h^x \hat{Z} (\hat{Z}' \hat{\Gamma}_0 \hat{Z})^{-1}] \hat{Z}' x_T, h > 1, \\ \hat{\chi}_{T+h|T}^{SW} &= [\hat{\Gamma}_h \hat{C} (\hat{C}' \hat{\Gamma}_0 \hat{C})^{-1}] \hat{C}' x_T, h > 1.\end{aligned}$$

Zum Vergleich verwende ich folgende Kriterien:

- Den quadratischen in-sample Fehler

$$MSE_{(\dots)} = 1/(nT) \sum_{t=1}^T \sum_{i=1}^n (\chi_{it} - \hat{\chi}_{it}^{(\dots)})^2.$$

- Den quadratischen out-of-sample Fehler

$$MSE_{(\dots)} = 1/(n) \sum_{i=1}^n (\chi_{i,T+1} - \hat{\chi}_{i,T+1|T}^{(\dots)})^2.$$

- Sowie die prozentuale Trefferquote

$$TQ_{(\dots)} = 1/n \sum_{i=1}^n \mathbb{I}_{\text{sign}(\chi_{i,T+1}) = \text{sign}(\hat{\chi}_{i,T+1|T}^{(\dots)})}.$$

Die ersten zwei Kriterien zielen rein auf die Erklärung des latenten Teils ab. Mathematisch gesehen sind sie natürlich korrekt und objektiv. Die Trefferquote ist mathematisch gesehen nicht sehr spektakulär, gibt aber einen quantitativen Überblick. Die Frage, die sich aber aufwirft, ist, inwiefern der latente Teil ausreichend ist, um den tatsächlichen Prozess zu erklären. Deswegen berechne ich alle drei Kriterien auch mit x an der Stelle von χ :

- Den quadratischen in-sample Fehler

$$MSE_{(\dots)} = 1/(nT) \sum_{t=1}^T \sum_{i=1}^n (x_{it} - \hat{\chi}_{it}^{(\dots)})^2.$$

- Den quadratischen out-of-sample Fehler

$$MSE_{(\dots)} = 1/(n) \sum_{i=1}^n (x_{i,T+1} - \hat{\chi}_{i,T+1|T}^{(\dots)})^2.$$

- Sowie die prozentuale Trefferquote

$$TQ_{(\dots)} = 1/n \sum_{i=1}^n \mathbb{I}_{\text{sign}(x_{i,T+1}) = \text{sign}(\hat{\chi}_{i,T+1|T}^{(\dots)})}.$$

Der quadratische x -in-sample Fehler hat eigentlich wenig Aussagekraft und ist unfair gegenüber der Methode von Forni und CO., da Stock und Watson genau diesen Fehler als zu minimierende Funktion verwenden. Dennoch führe ich ihn an, da man durch ihn sehr schön die Veränderungen durch falsche Faktoranzahlen erkennen kann. Interessant sind auf alle Fälle der quadratische x -out-of-sample Fehler, sowie die x -Trefferquote, da sie Kennzahlen sind, die für praktische Modellvergleiche (bei denen der latente Prozess naturgemäß nicht bekannt ist) hergenommen werden.

Wie eben schon angedeutet, spielen die Faktoranzahlen eine wesentliche Rolle. Benötigt man für die Methode von Stock und Watson lediglich die Anzahl der statischen Faktoren r , so verwendet die Methode von Forni und CO. auch die Anzahl der dynamischen Faktoren.

Bai und Ng haben in ihrem Paper [2] Informationskriterien vorgestellt, welche eine konsistente Schätzung der statischen Faktoranzahl r , für $(n, T) \rightarrow \infty$, ermöglichen, für die gilt

$$r = \arg \min_{1 \leq r \leq rmax} 1/(nT) \sum_{t=1}^T \sum_{i=1}^n (x_{it} - C_i^r F_t^r)^2 = \arg \min_{1 \leq r \leq rmax} LF(r).$$

Da dies ebenso die zu minimierende Funktion von Stock und Watson ist, wird die statische Faktoranzahl mit Hilfe der noch folgenden Informationskriterien geschätzt. Dies wird ebenso für die Methode von Forni und CO. angewandt, da in zahllosen von mir durchgeführten Versuchen, BIC- und AIC-Kriterien konsequent unterschätzt haben, im Gegensatz zu den Informationskriterien von Bai und Ng:

- $IC_1 = \ln(\hat{LF}(r)) + r \left(\frac{n+T}{nT}\right) \ln\left(\frac{nT}{n+T}\right)$
- $IC_2 = \ln(\hat{LF}(r)) + r \left(\frac{n+T}{nT}\right) \ln(\min(n, T))$
- $IC_3 = \ln(\hat{LF}(r)) + r \left(\frac{\ln(\min(n, T))}{\min(n, T)}\right)$

Die automatische Auswahl für r wurde folgendermaßen festgelegt:

1. $\hat{r} = \hat{r}_{IC_3}$ falls $\hat{r}_{IC_3} < rmax$
2. $\hat{r} = \lceil (\hat{r}_{IC_1} + \hat{r}_{IC_2})/2 \rceil$ falls $\hat{r}_{IC_3} = rmax$, $\lceil \cdot \rceil$ bedeutet die nächste größere natürliche Zahl.

Der Grund für diese Bestimmung von r ist folgender:

Alle Informationskriterienkurven haben einen konvex-konkaven Verlauf. Das heißt, falls $rmax$ zu hoch angesetzt ist, fallen die Kurven vor der Schranke unter das lokale Minimum, das eigentlich gesucht ist. Das dritte Kriterium lieferte in meinen Voruntersuchungen die konstanteste Trefferquote, sofern $rmax$ ($rmax = \min(n/2, T/2)$) nicht zu groß war, da ihr Kurvenverlauf der flacheste ist und somit am schnellsten fällt. Die anderen beiden streuten ein wenig um die tatsächliche Faktoranzahl. Daher wird ihr

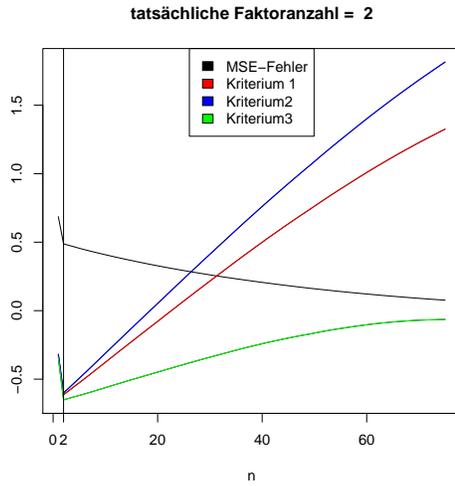


Abbildung 5.1: Modell 1

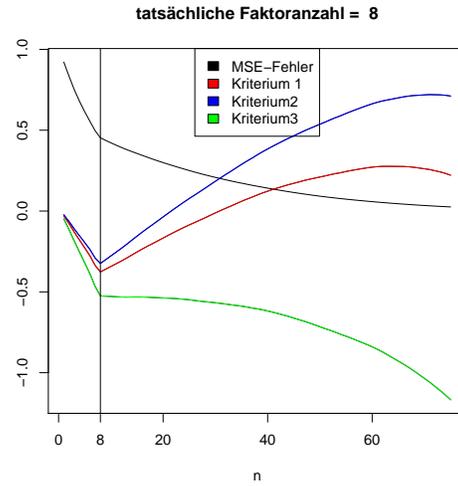


Abbildung 5.2: Modell 3

Mittelwert im Falle der Fehlschätzung des dritten Kriteriums herangezogen.

Somit fehlt noch die automatische Selektion von q für die Methode von Forni und CO. Hierfür haben Hallin und Liska in ihrem Paper [14], ähnlich zu Bai und Ng, Informationskriterien entwickelt, mit welchen es möglich ist, für $(n, T) \rightarrow \infty$, q konsistent zu schätzen. Das hier verwendete Kriterium lautet:

$$IC(q) = \ln \left(\frac{1}{n} \sum_{i=q+1}^n \frac{1}{2M_T + 1} \sum_{j=-M_T}^{M_T} \hat{\lambda}_{nT,i}(\theta_j) \right) + q \cdot p(n, T)$$

$$p(n, T) = (\min(n, M_T^2, \sqrt{T}M_T^{-1/2}))^{-1} \ln(\min(n, M_T^2, \sqrt{T}M_T^{-1/2}))$$

Das ergibt

$$\hat{q} = \arg \min_{1 \leq q \leq qmax} IC(q),$$

mit $qmax = \sqrt{n}$, $M_T = \sqrt{T}$ und θ_j wie bei der Schätzung der spektralen Dichte (siehe unten). Es tritt hier nur ein Problem auf. Alle Informationskriterien der Bauart

$$IC(q)_c = \ln \left(\frac{1}{n} \sum_{i=q+1}^n \frac{1}{2M_T + 1} \sum_{j=-M_T}^{M_T} \hat{\lambda}_{nT,i}(\theta_j) \right) + c \cdot q \cdot p(n, T),$$

mit $c \in \mathbb{R}^+$ erfüllen die Bedingungen der konsistenten Schätzung.

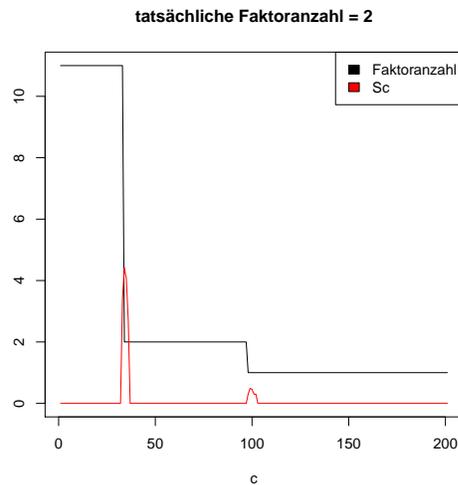
$$\hat{q}_c = \arg \min_{1 \leq q \leq qmax} IC(q)_c$$

Hallin und Liska schlagen deswegen vor, die Informationskriterien nicht nur für verschiedene (n,T) Kombinationen durchzurechnen, sondern auch verschiedene c's zu betrachten. Das heißt, zuerst definiert man Sequenzen für c, n und T. Implementiert wurde:

$$\begin{aligned}
 c &= (0, 0.1, 0.2, \dots, 2) \\
 J &= 10 \\
 n_j &= n - 20 + 2 * j, j = 1, \dots, J \\
 T_j &= T - 10 + j, j = 1, \dots, J, \text{ für } T = 30, \text{ und} \\
 T_j &= T - 50 + 5 * j, j = 1, \dots, J, \text{ für } T = 200.
 \end{aligned}$$

Und anschließend werden alle $\hat{q}_{c,(n_j,T_j)}$ berechnet. Der Sinn dahinter ist, c's zu erwischen, für die die quadratische Abweichung der $(\hat{q}_{c,(n_j,T_j)})_{j=1,\dots,J}$ gleich 0 ist.

$$S_c = \frac{1}{J} \sum_{j=1}^J \left(\hat{q}_{c,(n_j,T_j)} - \frac{1}{J} \sum_{j=1}^J \hat{q}_{c,(n_j,T_j)} \right)^2$$



Da $c \rightarrow S_c$ einige Intervalle mit $S_c = 0$ besitzt, schlagen Hallin und Liska vor, das zweite „Stabilitätsintervall“ zu nehmen. Der Grund ist, dass ein zu kleines c zu einer Überschätzung der Faktoranzahl neigt.

Für die Berechnung der Kriterien werden alle auftretenden Varianz-Kovarianzmatrizen

$\Gamma_h^x = \mathbb{E}(x_{t+h}x_t')$ mit

$$\hat{\Gamma}_h^x = \frac{1}{T} \sum_{t=1}^{T-h} (x_{t+h} - \bar{x}_T)(x_t - \bar{x}_T)^*$$

aus (2.1) geschätzt.

Die spektrale Dichte von x_t wird mit

$$\hat{\Sigma}(\theta_j) = \frac{1}{2\pi} \sum_{h=-M_T}^{M_T} w(M_t^{-1}h) \hat{\Gamma}_h^x e^{-i\theta_j h}$$

aus (2.3) mit $M_T = \sqrt{T}$ und $w(x) = 1 - |x| = 1 - \frac{|h|}{M_T}$ geschätzt. Die verwendeten Frequenzen werden folgendermaßen bestimmt:

$$\theta_j = \frac{j\pi}{M_T + 0.5}, j = -M_T, \dots, M_T.$$

Dadurch sehen die Schätzer der Varianz-Kovarianz-Matrizen von Γ_h^x und Γ_h^ε für die FHLR-Methode so aus:

$$\hat{\Gamma}_h^x = \frac{2\pi}{2M_T + 1} \sum_{j=-M_T}^{M_T} \hat{\Sigma}^x(\theta_j) e^{i\theta_j h},$$

$$\hat{\Gamma}_h^\varepsilon = \frac{2\pi}{2M_T + 1} \sum_{j=-M_T}^{M_T} \hat{\Sigma}^\varepsilon(\theta_j) e^{i\theta_j h}.$$

Ich befolge aber den Rat von FHLR, und setze die Nicht-diagonal-Elemente von $\hat{\Gamma}_0^\varepsilon$ gleich 0. FHLR haben in [12] festgestellt, dass vorallem im Fall von $n > T$ das 0 setzen der Nicht-diagonal-elemente eine signifikante Verbesserung der Schätzung darstellt. Zudem erlaubt dies die Methode auch, da Γ_0^ε durch eine beliebige, symmetrische, positiv definite Matrix ersetzt werden kann, die beschränkte Eigenwerte hat.

Modell 1

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0527	0.0561	0.3093	0.3122	0.7483	0.7464
n=100	0.0433	0.0448	0.2939	0.2970	0.7445	0.7429
n=150	0.0404	0.0411	0.2814	0.2839	0.7427	0.7425

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4606	0.4480	0.7980	0.8004	0.6527	0.6517
n=100	0.4683	0.4576	0.7903	0.7931	0.6582	0.6574
n=150	0.4710	0.4608	0.7873	0.7903	0.6519	0.6523

Schätzungen der Faktoranzahlen

T=30	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	0	100	0	0	99	1
n=100	0	100	0	0	100	0	0	100	0
n=150	0	88	12	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	0	100	0	-	-	-
n=100	-	-	-	0	100	0	-	-	-
n=150	-	-	-	0	100	0	0	100	0

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0527	0.0567	0.3093	0.3120	0.7483	0.7472
n=100	0.0433	0.0448	0.2939	0.2970	0.7446	0.7429
n=150	0.0406	0.0411	0.2815	0.2839	0.7427	0.7425

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4606	0.4473	0.7980	0.8001	0.6527	0.6520
n=100	0.4682	0.4576	0.7902	0.7931	0.6582	0.6574
n=150	0.4698	0.4608	0.7872	0.7903	0.6524	0.6523

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0242	0.0254	0.2306	0.2296	0.7608	0.7624
n=100	0.0150	0.0152	0.2336	0.2329	0.7700	0.7686
n=150	0.0118	0.0118	0.2177	0.2184	0.7728	0.7725

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4772	0.4750	0.7226	0.7216	0.6742	0.6752
n=100	0.4866	0.4851	0.7313	0.7303	0.6713	0.6715
n=150	0.4897	0.4884	0.7216	0.7221	0.6679	0.6675

Schätzungen der Faktoranzahlen

T=200	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	0	100	0	0	100	0
n=100	0	100	0	0	100	0	0	100	0
n=150	0	100	0	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	0	100	0	-	-	-
n=100	-	-	-	0	100	0	-	-	-
n=150	-	-	-	0	100	0	-	-	-

Die Statistiken sind recht ausgeglichen. Für T = 30 liefert die FHLR-Methode die leicht besseren Ergebnisse, für T = 200 kann man kaum einen Unterschied bemerken. Die Ergebnisse der Statistiken mit den geschätzten Faktoranzahlen für T = 200 sind überflüssig, da alle Anzahlen perfekt geschätzt wurden. Für T = 30 ist es erfreulich, dass die Überschätzung von q keine Auswirkung auf die richtige Schätzung von r hat.

Modell 2

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1300	0.1793	0.3170	0.3415	0.7127	0.7040
n=100	0.1088	0.1397	0.3094	0.3277	0.7308	0.7253
n=150	0.1016	0.1263	0.2963	0.3149	0.7395	0.7313

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.3845	0.3426	0.8202	0.8457	0.6405	0.6330
n=100	0.4060	0.3720	0.8058	0.8229	0.6513	0.6473
n=150	0.4126	0.3817	0.8049	0.8237	0.6515	0.6479

Schätzungen der Faktoranzahlen

T=30	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	42	58	0	68	32	0	91	9	0
n=100	26	74	0	52	48	0	6	94	0
n=150	26	74	0	51	49	0	6	94	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	100	0	0	44	56	0	-	-	-
n=100	100	0	0	35	65	0	-	-	-
n=150	100	0	0	34	66	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1780	0.2068	0.3519	0.3668	0.6866	0.6817
n=100	0.1441	0.1397	0.3399	0.3275	0.7101	0.7255
n=150	0.1385	0.1267	0.3266	0.3156	0.7179	0.7309

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4893	0.4537	0.8558	0.8729	0.6250	0.6199
n=100	0.4738	0.3746	0.8387	0.8224	0.6373	0.6476
n=150	0.4795	0.3844	0.8345	0.8247	0.6392	0.6477

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0654	0.0780	0.2148	0.2173	0.7729	0.7712
n=100	0.0419	0.0466	0.2039	0.2058	0.7822	0.7807
n=150	0.0335	0.0362	0.1950	0.1961	0.7862	0.7853

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4364	0.4253	0.7227	0.7250	0.6749	0.6742
n=100	0.4614	0.4552	0.7011	0.7030	0.6783	0.6780
n=150	0.4700	0.4650	0.6942	0.6952	0.6796	0.6797

Schätzungen der Faktoranzahlen

T=200	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	1	99	0	0	100	0
n=100	0	100	0	0	100	0	0	100	0
n=150	0	100	0	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	1	99	0	-	-	-
n=100	-	-	-	0	100	0	-	-	-
n=150	-	-	-	0	100	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0655	0.0781	0.2149	0.2172	0.7728	0.7712
n=100	0.0419	0.0466	0.2039	0.2058	0.7822	0.7807
n=150	0.0335	0.0362	0.1950	0.1961	0.7862	0.7853

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4366	0.4254	0.7230	0.7249	0.6746	0.6744
n=100	0.4614	0.4552	0.7011	0.7030	0.6783	0.6780
n=150	0.4700	0.4650	0.6942	0.6952	0.6796	0.6797

Für $T = 200$ sind die out-of-sample Statistiken beider Methoden so gut wie gleich auf. Die in-sample Statistiken sprechen jedoch eher für die FHLR-Methode. Auch die Faktoranzahlen werden so gut geschätzt, dass keine gravierenden Veränderungen der Statistiken durch die Schätzung der Faktoranzahlen erfolgt.

Bei $T = 30$ scheint die Methode von FHLR der von SW leicht überlegen zu sein, sofern die Faktoranzahlen bekannt sind. Die Schätzung eben dieser funktioniert hier nämlich so ganz und gar nicht. So wird q sehr häufig unterschätzt, und das führt in weiterer Folge zu einer systematischen Unterschätzung von r . Bei der Methode von SW funktioniert die Faktoranzahlschätzung für $n = 100, 150$ weitaus besser, was zur Folge hat, dass diese Statistiken mit geschätzten Faktoranzahlen klar für die Methode von SW sprechen. Für $n = 50$ liefert die Schätzung der Faktoranzahl jedoch bei SW eine konsequente Unterschätzung, wodurch FHLR dort wieder die Nase vorne hat.

Modell 3

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1779	0.3075	0.3535	0.3952	0.6974	0.6830
n=100	0.1463	0.2248	0.2938	0.3215	0.7248	0.7121
n=150	0.1354	0.1936	0.3294	0.3450	0.7282	0.7243

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.3593	0.2671	0.8485	0.8915	0.6344	0.6243
n=100	0.3888	0.3184	0.7974	0.8245	0.6456	0.6402
n=150	0.3981	0.3351	0.8278	0.8452	0.6520	0.6491

Schätzungen der Faktoranzahlen

T=30	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	91	9	0	100	0	0	19	57	23
n=100	71	29	0	100	0	0	95	4	1
n=150	67	33	0	100	0	0	36	60	5

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	100	0	0	100	0	0	-	-	-
n=100	100	0	0	100	0	0	-	-	-
n=150	100	0	0	100	0	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.2901	0.3149	0.4308	0.4034	0.6366	0.6734
n=100	0.2676	0.2486	0.3778	0.3528	0.6533	0.6850
n=150	0.2681	0.1918	0.4272	0.3468	0.6546	0.7244

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.6454	0.3020	0.9262	0.9024	0.5941	0.6169
n=100	0.6379	0.4902	0.8808	0.8584	0.5966	0.6195
n=150	0.6524	0.3531	0.9253	0.8469	0.6035	0.6491

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0910	0.1527	0.2120	0.2301	0.7788	0.7669
n=100	0.0581	0.0767	0.1802	0.1853	0.7951	0.7929
n=150	0.0462	0.0557	0.1750	0.1776	0.8060	0.8042

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4175	0.3859	0.7214	0.7383	0.6816	0.6750
n=100	0.4507	0.4358	0.6781	0.6832	0.6889	0.6868
n=150	0.4622	0.4512	0.6705	0.6729	0.6915	0.6908

Schätzungen der Faktoranzahlen

T=200	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	38	62	0	0	0	100
n=100	0	100	0	0	100	0	0	100	0
n=150	0	100	0	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	38	62	0	-	-	-
n=100	-	-	-	0	100	0	-	-	-
n=150	-	-	-	0	100	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0979	0.4392	0.2239	0.2472	0.7722	0.7579
n=100	0.0581	0.0767	0.1802	0.1853	0.7951	0.7929
n=150	0.0462	0.0557	0.1750	0.1776	0.8060	0.8042

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4299	0.0679	0.7336	0.7603	0.6765	0.6661
n=100	0.4507	0.4358	0.6781	0.6832	0.6889	0.6868
n=150	0.4622	0.4512	0.6705	0.6729	0.6915	0.6908

Bei bekannten Faktoranzahlen liefert die FHLR-Methode klar bessere Ergebnisse.

Leider sieht die Faktoranzahlschätzung für $T = 30$ bei beiden Methoden traurig aus, wobei SW noch besser schätzt und somit bei den Statistiken mit geschätzten Faktoranzahlen die Oberhand hat. Hier sieht man sehr schön, wie dramatisch die x-MSE-in-sample Fehler bei der FHLR-Methode in die Höhe schnellen, während die chi-MSE-in-sample Fehler noch annähernd mit denen von SW mithalten können. Wie gesagt, diese Statistik stellt keine wirkliche Beurteilung dar, aber man sieht hier am deutlichsten die Fehlschätzung der Faktoranzahlen.

Für $T = 200$ funktioniert die Faktoranzahlschätzung sehr gut, außer bei SW für $n = 50$. Hier ist auch der Grund offensichtlich. Wie man sieht, ist die wenig aussagekräftige x-MSE-in-sample-Fehler Statistik bei SW durch die Hinzunahme von vielen unnötigen Faktoren sogar kleiner als die chi-MSE-in-sample-Fehler Statistik bei SW, mit wahren Faktoranzahlen. Das heißt, die korrekte Schätzung von r war nicht möglich, da der x-MSE-in-sample-Fehler am Anfang zu langsam gesunken ist, und die Informationskriterien die Hinzunahme von Faktoren nicht stark genug bestrafen konnten. Es sei aber erwähnt, dass diese Überschätzung durch die Automatisierung verursacht wird. Würde man jedes Modell einzeln betrachten, würde man das lokale Minimum sehen. Leider ist r_{\max} hierfür zu groß, ein kleineres r_{\max} wäre jedoch bei der Automatisierung unverantwortlich, da ein gewisser Spielraum zur wahren Faktoranzahl gewährleistet sein muss, um die Objektivität zu bewahren. $r_{\max} \approx r+2$ würde sicherlich zum richtigen Ergebnis führen, kann aber nicht Sinn und Zweck sein.

Modell 4

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1096	0.1411	0.3656	0.3713	0.7195	0.7185
n=100	0.0847	0.1019	0.3341	0.3430	0.7329	0.7292
n=150	0.0759	0.0893	0.3140	0.3189	0.7418	0.7407

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4090	0.3868	0.8649	0.8710	0.6515	0.6525
n=100	0.4275	0.4113	0.8302	0.8383	0.6540	0.6531
n=150	0.4337	0.4193	0.8093	0.8133	0.6606	0.6608

Schätzungen der Faktoranzahlen

T=30	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	33	67	0	35	65	0	2	20	77
n=100	14	86	0	14	86	0	2	91	7
n=150	10	90	0	12	88	0	0	99	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	98	2	0	4	96	0	-	-	-
n=100	91	9	0	2	98	0	-	-	-
n=150	100	0	0	2	98	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1397	0.2950	0.3964	0.4458	0.6992	0.6828
n=100	0.0985	0.1066	0.3473	0.3456	0.7233	0.7267
n=150	0.0870	0.0895	0.3258	0.3192	0.7357	0.7405

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4682	0.2215	0.8948	0.9452	0.6370	0.6247
n=100	0.4512	0.4075	0.8428	0.8407	0.6483	0.6510
n=150	0.4514	0.4192	0.8210	0.8135	0.6571	0.6609

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0525	0.0645	0.2500	0.2501	0.7465	0.7470
n=100	0.0309	0.0348	0.2340	0.2347	0.7623	0.7614
n=150	0.0238	0.0260	0.2299	0.2304	0.7587	0.7580

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4523	0.4459	0.7585	0.7590	0.6591	0.6579
n=100	0.4718	0.4688	0.7247	0.7252	0.6668	0.6663
n=150	0.4782	0.4760	0.7375	0.7379	0.6589	0.6587

Schätzungen der Faktoranzahlen

T=200	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	0	100	0	0	0	100
n=100	0	100	0	0	100	0	0	100	0
n=150	0	100	0	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	0	100	0	-	-	-
n=100	-	-	-	0	100	0	-	-	-
n=150	-	-	-	0	100	0	-	-	-

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0525	0.4524	0.2500	0.3075	0.7465	0.7154
n=100	0.0309	0.0348	0.2340	0.2347	0.7623	0.7614
n=150	0.0238	0.0260	0.2299	0.2304	0.7587	0.7580

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4523	0.0490	0.7585	0.8177	0.6591	0.6386
n=100	0.4718	0.4688	0.7247	0.7252	0.6668	0.6663
n=150	0.4782	0.4760	0.7375	0.7379	0.6589	0.6587

Für T = 200 sind die Ergebnisse annähernd gleich, vielleicht mit einem minimalen

Vorteil für FHLR. Erneut hat SW bei der Faktorschätzung mit $n = 50$ dasselbe Problem wie bei Modell 3.

Für $T = 30$ ist der Vorteil der FHLR-Methode mit bekannten Faktoranzahlen jedoch größer. Interessant ist hier die Faktoranzahlschätzung. So zeigt sich hier, dass die richtige/ falsche Schätzung von q maßgeblich für die von r verantwortlich ist. Obwohl hier FHLR für $n = 100, 150$ größere Probleme hat als SW, kann sie dennoch mit den Statistiken von SW mithalten. Für $n = 50$ hat SW wieder das bereits bekannte Problem.

Modell 5

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1130	0.1341	0.2635	0.2463	0.7997	0.7945
n=100	0.0992	0.0992	0.2495	0.2315	0.8042	0.8048
n=150	0.0942	0.0878	0.2294	0.2033	0.8222	0.8219

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4654	0.3894	0.7626	0.7444	0.6958	0.6964
n=100	0.4776	0.4121	0.7510	0.7335	0.6989	0.6992
n=150	0.4819	0.4196	0.7349	0.7103	0.7096	0.7081

Schätzungen der Faktoranzahlen

T=30	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	90	10	41	59	0	25	70	5
n=100	0	90	10	34	66	0	3	93	3
n=150	0	91	9	27	73	0	3	97	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	44	56	0	10	90	0
n=100	-	-	-	37	63	0	6	94	0
n=150	-	-	-	30	70	0	0	100	0

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.1226	0.1402	0.2732	0.2539	0.7906	0.7902
n=100	0.1064	0.1008	0.2543	0.2327	0.8003	0.8049
n=150	0.1009	0.0878	0.2354	0.2035	0.8147	0.8221

Statistiken für den wahren Prozess						
T=30	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4790	0.4014	0.7718	0.7518	0.6906	0.6934
n=100	0.4886	0.4119	0.7559	0.7347	0.6971	0.6986
n=150	0.4921	0.4209	0.7412	0.7105	0.7052	0.7085

Schätzungen mit den wahren Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0498	0.0621	0.1487	0.1500	0.8308	0.8264
n=100	0.0325	0.0340	0.1331	0.1290	0.8414	0.8419
n=150	0.0266	0.0255	0.1265	0.1233	0.8517	0.8521

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4632	0.4464	0.6511	0.6513	0.7078	0.7066
n=100	0.4797	0.4690	0.6368	0.6329	0.7097	0.7109
n=150	0.4854	0.4762	0.6208	0.6177	0.7194	0.7182

Schätzungen der Faktoranzahlen

T=200	$\hat{q} < q$	$\hat{q} = q$	$\hat{q} > q$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_{SW} < r$	$\hat{r}_{SW} = r$	$\hat{r}_{SW} > r$
n=50	0	100	0	1	99	0	1	99	0
n=100	0	69	31	0	100	0	0	100	0
n=150	0	48	52	0	100	0	0	100	0

	wenn $\hat{q} < q$, dann			wenn $\hat{q} = q$, dann			wenn $\hat{q} > q$, dann		
	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$	$\hat{r}_F < r$	$\hat{r}_F = r$	$\hat{r}_F > r$
n=50	-	-	-	1	99	0	-	-	-
n=100	-	-	-	0	100	0	0	100	0
n=150	-	-	-	0	100	0	0	100	0

Schätzungen mit den geschätzten Faktoranzahlen

Statistiken für den latenten Teil						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.0500	0.0622	0.1485	0.1499	0.8314	0.8268
n=100	0.0316	0.0340	0.1317	0.1290	0.8420	0.8419
n=150	0.0249	0.0255	0.1260	0.1233	0.8517	0.8521

Statistiken für den wahren Prozess						
T=200	in-sample		out-of-sample			
	MSE_{FHLR}	MSE_{SW}	MSE_{FHLR}	MSE_{SW}	TQ_{FHLR}	TQ_{SW}
n=50	0.4636	0.4468	0.6512	0.6515	0.7084	0.7069
n=100	0.4775	0.4690	0.6354	0.6329	0.7103	0.7109
n=150	0.4821	0.4762	0.6204	0.6177	0.7190	0.7182

Bei bekannten Faktoranzahlen sind die in-sample Statistiken von FHLR ein wenig besser als die von SW. Dafür sind die out-of-sample Statistiken für $T = 30$ von SW

denen von FHLR überlegen.

Die Faktoranzahlen werden bei beiden Methoden für $T = 30$ nicht perfekt, dafür für $T = 200$ sehr gut geschätzt, wodurch sich das selbe Bild, für die Statistiken mit geschätzten Faktoranzahlen, wie für die Statistiken mit bekannten Faktoranzahlen, ergibt.

Zusammenfassung

- Die FHLR-Methode liefert bei bekannten Faktoranzahlen die konstant besseren Werte. Bei Modell 1 sind die in-sample Ergebnisse besser, bei Modell 5 ebenso, die out-of-sample Ergebnisse hinken da denen der SW-Methode jedoch hinterher. Bei den Modellen 2,3 und 4 ist FHLR die konstant bessere.

- Die Schätzung der Faktoren für $T = 30$ könnte besser sein. Es ist aber zu beachten, dass $T = 30$ wirklich klein ist und erst durch $n = 100$ halbwegs kompensiert werden kann. Dennoch zeichnet sich ein Trend ab, dass eine Unterschätzung von q ebenso zu einer Unterschätzung von r führt. Wird q jedoch richtig geschätzt, ist es äußerst wahrscheinlich, auch r richtig zu schätzen. Dies wird vorallem bei $T = 200$ untermauert, wo sogar Fehlschätzungen meist zu einer richtigen Schätzung von r führen.

Man darf jedoch nicht vergessen, dass trotz einer richtigen Schätzung von r , bei einer vorherigen Fehlschätzung von q , die Statistiken der MSE-Fehler nicht unbedingt sehr gut sein müssen, da immerhin Γ_0^x bereits falsch in die Berechnung der FHLR-Methode einfließt.

Die Informationskriterien von Bai und Ng erfüllen ihren Dienst sehr gut, sieht man von dem Problem der Automatisierung ab, wodurch mitunter das lokale Minimum nicht erkannt wird.

- Die Statistiken mit den geschätzten Faktoranzahlen hinken denen mit den wahren Faktoranzahlen, nur in den Extremfällen der Fehlschätzungen der Faktoranzahlen, wirklich hinterher. Vorallem die Trefferquoten des wahren Prozesses x sind konstant über 60%, was wirklich beachtlich ist.

Ansonsten ist der Vergleich zwischen FHLR und SW so zu betrachten wie bei bekannten Faktoranzahlen, außer, dass durch die Probleme bei der Faktoranzahlschätzung der Vorteil von FHLR bei den Modellen 2 und 3 für $T = 30$ zu SW wandert.

- Durch die Steigerung der Dynamik und Abhängigkeiten der Fehler von Modell 1 zu Modell 5, lässt die stetig besser werdende Performance der FHLR-Methode

gegenüber der SW-Methode, mit bekannten Faktoranzahlen, bei den Modellen 2, 3 und 4, einen Vorteil dieser Methode gegenüber der SW-Methode erahnen. Modell 5 relativiert jedoch diesen etwaigen Vorteil.

Die Simulation hat gezeigt, dass keine der beiden Methoden die andere klar übertrifft. Die Frage, ob durch die Verwendung der konsistent geschätzten Varianz-Kovarianz-Matrix der Fehler Γ_0^ε eine signifikante Verbesserung der Schätzung der latenten Variablen eintritt, bleibt offen.

Die Theorie der dynamischen verallgemeinerten Faktormodelle ist jedoch noch jung, und es lässt sich erhoffen, diese durch weitere Forschungsarbeiten noch besser erschließen zu können.

Kapitel 6

Appendix A

Beweise für die Methode von FHLR

Beweis (*Lemma 4.1.2*) Seien $(v_{nj})_{j=1,\dots,(k-1)}$ die verallgemeinerten Eigenvektoren von (A_n, B_n) und $(w_{nj})_{j=1,\dots,k}$ die Eigenvektoren von A_n (wie immer seien sie der Größe ihrer zugehörigen Eigenwerte nach geordnet und orthonormal).

Weiters sei $\alpha = (\alpha_{n1}, \dots, \alpha_{nk}) \in \mathbb{R}^{1 \times k}$ eine nicht-triviale Lösung des linearen Gleichungssystems mit $(k-1)$ Gleichung und k Unbekannten in x_i

$$(x_1 w'_{n1} + \dots + x_k w'_{nk}) B_n v_{nj} = 0, \quad j = 1, \dots, k-1$$

$$\Leftrightarrow V'_n B_n W_n (x_1, \dots, x_k)' = 0, \quad \text{mit } W_n = (w_{n1} \cdots w_{nk}) \text{ und } V_n = (v_{n1} \cdots v_{n(k-1)}).$$

Da die w_{ni} orthonormal sind, ist $q'_n = (\alpha_{n1} w'_{n1} + \dots + \alpha_{nk} w'_{nk}) \neq 0$. Deswegen ist auch $q'_n B_n q_n > 0$ (B_n ist positiv definit). Da man α mit einer Konstanten multiplizieren kann, ist es möglich,

$$q'_n B_n q_n = 1 \quad \text{mit} \quad (6.1)$$

$$q'_n B_n v_{nj} = 0, \quad j = 1, \dots, k-1 \quad (6.2)$$

zu bekommen. Für $k=1$ tritt (6.2) nicht auf und man setzt $\alpha_{n1} = \frac{1}{\sqrt{w'_{n1} B_n w_{n1}}} = \frac{1}{\sqrt{M}}$.

Für $k > 1$ gilt $\alpha_{n1}^2 + \dots + \alpha_{nk}^2 \geq \frac{1}{M}$, da

$$\begin{aligned} 1 = q'_n B_n q_n &= \alpha_{n1} w'_{n1} B_n w_{n1} \alpha_{n1} + \dots + \alpha_{nk} w'_{nk} B_n w_{nk} \alpha_{nk} \\ &\leq M(\alpha_{n1}^2 + \dots + \alpha_{nk}^2). \end{aligned}$$

Die Ungleichung gilt, da w_{ni} Norm 1 haben und der maximale Wert von $w'_{ni} B_n w_{ni}$ dadurch der größte Eigenwert von B_n ist. Dadurch gilt weiters,

$$q'_n A_n q_n = \alpha_{n1}^2 \mu_{n1} + \dots + \alpha_{nk}^2 \mu_{nk} \geq \mu_{nk} \frac{1}{M} \quad (6.3)$$

(q_n ist eine Linearkombination der Eigenvektoren von A_n). Beachtet man noch, dass

$$\mu_{nk} = \max_{a_n, a'_n B_n v_{nj}=0, j < k} a'_n A_n a_n \text{ mit } a'_n B_n a_n = 1$$

und dadurch

$$\lambda_{nk} \geq q'_n A_n q_n \geq \mu_{nk}/M \rightarrow \infty \text{ f\u00fcr } n \rightarrow \infty$$

gilt, folgt somit die Behauptung. Die erste Ungleichung gilt wegen (6.1) und (6.2) und die zweite wegen (6.3). \square

Beweis (*Lemma 4.1.3*) Sei a ein normierter Spaltenvektor, dann gilt

$$\begin{aligned} a' \check{\Gamma}_0^\varepsilon a &= a' \int_{-\pi}^{\pi} \check{\Sigma}^\varepsilon(\theta) d\theta a = \int_{-\pi}^{\pi} a' \check{\Sigma}^\varepsilon a(\theta) d\theta \leq \int_{-\pi}^{\pi} \check{\lambda}_1^\varepsilon(\theta) d\theta \\ &= \int_{-\pi}^{\pi} \lambda_{q+1}(\theta) d\theta \leq \int_{-\pi}^{\pi} \lambda_1^\varepsilon(\theta) d\theta = \alpha \leq 2\pi M, \end{aligned} \quad (6.4)$$

da $\check{\lambda}_1^\varepsilon = \lambda_{q+1}$ laut Konstruktion ist, die erste Ungleichung in der zweiten Zeile laut (Satz 2.1.15): $\lambda_{q+1} \leq \lambda_1^\varepsilon$ f\u00fcr $H_1 = \Sigma$ und $H_2 = \Sigma^\varepsilon$ gilt, da $H_1 - H_2 = \Sigma - \Sigma^\varepsilon = \Sigma^\chi$ und $rk(\Sigma^\chi) = q$ laut Definition ist, und die zweite Ungleichung unmittelbar aus der Annahme C2 folgt.

Da obige Ungleichungen f\u00fcr alle normierten Spaltenvektoren gelten, trifft sie auch f\u00fcr den zum gr\u00f6\u00dften Eigenwert geh\u00f6rigen Eigenvektor zu, wodurch (ii) folgt.

F\u00fcr die erste Behauptung des Lemmas ben\u00f6tigt man folgende Beobachtung:

$$\begin{aligned} A := \Gamma_0^\varepsilon - \check{\Gamma}_0^\varepsilon &= \int_{-\pi}^{\pi} \Sigma^\varepsilon(\theta) - \check{\Sigma}^\varepsilon(\theta) d\theta = \int_{-\pi}^{\pi} \Sigma(\theta) - \check{\Sigma}(\theta) + \check{\Sigma}^\chi(\theta) - \Sigma^\chi(\theta) d\theta \\ &= \int_{-\pi}^{\pi} \check{\Sigma}^\chi(\theta) - \Sigma^\chi(\theta) d\theta = \check{\Gamma}_0^\chi - \Gamma_0^\chi, \end{aligned} \quad (6.5)$$

wobei

$$|a' A a| = |a' (\Gamma_0^\varepsilon - \check{\Gamma}_0^\varepsilon) a| = |a' \Gamma_0^\varepsilon a - a' \check{\Gamma}_0^\varepsilon a| \leq 2\pi M$$

wegen Annahme C2 und (6.4) gilt, und dadurch die Matrix $(2\pi M I_n + A)$ positiv semi-definit ist.

Durch Umformen von (6.5) und Addieren der Matrix $2\pi M I_n$ erh\u00e4lt man

$$\check{\Gamma}_0^\chi + 2\pi M I_n = \Gamma_0^\chi + 2\pi M I_n + A.$$

Aus (Satz 2.1.15) mit $H_1 = \check{\Gamma}_0^\chi + 2\pi M I_n$ und $H_2 = \Gamma_0^\chi$ folgt, dass $\check{\mu}_r^\chi + 2\pi M \geq \mu_r^\chi$ ist, und somit die Behauptung (i). \square

Beweis (Lemma 4.1.4)

$$\text{Var}(a_n \varepsilon_t) = a_n \Gamma_0^\varepsilon a_n' \quad (6.6)$$

$$= a_n \int_{-\pi}^{\pi} \Sigma_n^\varepsilon(\theta) d\theta a_n' = \int_{-\pi}^{\pi} a_n \Sigma_n^\varepsilon(\theta) a_n' d\theta \quad (6.7)$$

$$= \int_{-\pi}^{\pi} \|a_n\| \left(\frac{a_n}{\|a_n\|} \right) \Sigma_n^\varepsilon(\theta) \left(\frac{a_n}{\|a_n\|} \right)' \|a_n\| d\theta \quad (6.8)$$

$$\leq \int_{-\pi}^{\pi} \|a_n\|^2 \lambda_1^\varepsilon(\theta) d\theta \quad (6.9)$$

$$\leq \|a_n\|^2 2\pi M = 2\pi M \sum_{i=1}^n a_{ni}^2 \quad (6.10)$$

(6.6) \Rightarrow (6.7): da $\Gamma_0^\varepsilon = \int_{-\pi}^{\pi} \Sigma_n^\varepsilon(\theta) d\theta$,

(6.8) \Rightarrow (6.9): weil $\lambda_1^\varepsilon(\theta) = \sup_{a, \|a\|=1}$ und

(6.9) \Rightarrow (6.10): wegen Annahme C2 □

Beweis (Lemma 4.1.5) o.B.d.A. sei \mathbf{V} orthonormal¹. Da K ein linearer Teilraum von H ist, kann man \mathbb{V}_n eindeutig als

$$\mathbb{V}_n = w_n \mathbf{V} + W_n \quad (6.11)$$

darstellen ($w_n \in \mathbb{R}^{k \times k}$). Da $w_n \mathbf{V} = (\text{Proj}(v_{n1}|K), \dots, \text{Proj}(v_{nk}|K))'$ ist, konvergiert W_n laut Annahme im quadratischen Mittel gegen 0 für $n \rightarrow \infty$. Weiters gilt, da

$$v_{nj} \dots \text{orthonormal} \Leftrightarrow \text{Var}(v_{nj}) = 1$$

und dadurch aus

$$\begin{aligned} \text{Var}(v_{nj}) &= \text{Var}(w_{nj} \mathbf{V} + W_{nj}) = \text{Var}(w_{nj} \mathbf{V}) + \text{Var}(W_{nj}) \\ &\geq \text{Var}(w_{nj} \mathbf{V}) = w_{nj} \mathbb{E}(\mathbf{V}\mathbf{V}') w_{nj}' = \|w_{nj}\|^2, \end{aligned}$$

wobei w_{nj} die j -te Zeile von w_n und W_{nj} das j -te Element von W_n darstellen. Die zweite Gleichheit gilt, da W_n das orthogonale Komplement von $w_{nj} \mathbf{V}$ ist und die letzte Gleichheit stimmt, da $\mathbb{E}(\mathbf{V}\mathbf{V}') = I_k$ gilt.

¹Durch das Schmidtsche Orthogonalisierungsverfahren bekommt man jederzeit ein orthogonales k -Tupel, Orthonormalität erreicht man im Anschluss jederzeit.

Daraus folgt, dass w_n beschränkt ist.

Das Analogon von (6.11) für \mathbf{V} ist

$$\mathbf{V} = s_n \mathbb{V}_n + S_n = w'_n \mathbb{V}_n + S_n. \quad (6.12)$$

Die Gleichheit von s_n und w'_n überlegt man sich folgendermaßen:

$$\begin{aligned} w_n &= \mathbb{E}(\mathbb{V}_n \mathbf{V}) (\mathbb{E}(\mathbf{V} \mathbf{V})^{-1}) = \mathbb{E}(\mathbb{V}_n \mathbf{V}) \\ s_n &= \mathbb{E}(\mathbf{V} \mathbb{V}_n) (\mathbb{E}(\mathbb{V}_n \mathbb{V}_n)^{-1}) = \mathbb{E}(\mathbf{V} \mathbb{V}_n) \\ \mathbb{E}(\mathbf{V} \mathbf{V}) &= \mathbb{E}(\mathbb{V}_n \mathbb{V}_n) = I_k. \end{aligned}$$

Indem man auf (6.11) und (6.12) den Varianz-Operator anwendet, kommt man durch die selben Überlegungen wie ein paar Zeilen weiter oben, auf die Gleichungen $I_k = w_n w'_n + \text{Var}(W_n)$ und $I_k = w'_n w_n + \text{Var}(S_n)$, woraus, durch Berechnung der Spuren der obigen Matrizen, die Gleichheit von $\text{tr}(\text{Var}(W_n))$ und $\text{tr}(\text{Var}(S_n))$ folgt.

Da weiters die Konvergenz gegen 0 im quadratischen Mittel von W_n bedeutet, dass $\text{Var}(W_n)$ gegen 0 strebt und dies äquivalent dazu ist, dass die Spur dieser Varianz-Kovarianz-Matrix gegen 0 konvergiert folgt daraus, dass auch S_n gegen 0 im quadratischen Mittel konvergiert.

In der bereits bekannten Weise spaltet man $v \in H$ auf:

$$v = b_n \mathbb{V}_n + t_n \quad (6.13)$$

$$\begin{aligned} &= b_n (w_n \mathbf{V} + W_n) + t_n = b_n w_n \mathbf{V} + b_n W_n + t_n \\ &= b \mathbf{V} + t, \end{aligned} \quad (6.14)$$

wobei b_n und b die $(1 \times k)$ Projektionsmatrizen auf die Unterräume K_n bzw. K darstellen.

Das heißt, die Differenz der beiden Projektionen läßt sich folgendermaßen darstellen:

$$\text{Proj}(v|K_n) - \text{Proj}(v|K) = b_n \mathbb{V}_n - b \mathbf{V} = (b_n w_n - b) \mathbf{V} + b_n W_n = t - t_n. \quad (6.15)$$

Beachtet man, dass v quadratisch integrierbar ist, was gleichbedeutend dazu ist, dass die $\text{Var}(v) \leq \infty$ ist, und dass die Varianzen der $v_{nj} = 1$ sind, erhält man durch Anwendung des Varianz-Operators auf (6.13), dass b_n beschränkt ist. Dadurch konvergiert auch $(t - t_n) - b_n W_n = (b_n w_n - b) \mathbf{V} \rightarrow 0$ im quadratischen Mittel für $n \rightarrow \infty$.
Formal:

$$\begin{aligned} &\text{Var}((t - t_n) - b_n W_n = (b_n w_n - b) \mathbf{V}) = \\ &\text{Var}((b_n w_n - b) \mathbf{V}) + \text{Var}(t - t_n) - 2\text{Cov}((b_n w_n - b) \mathbf{V}, (t - t_n)) \rightarrow 0 \end{aligned} \quad (6.16)$$

Bleibt noch zu zeigen, dass

$$\text{Cov}((b_n w_n - b)\mathbf{V}, (t - t_n)) \rightarrow 0, \text{ für } n \rightarrow \infty. \quad (6.17)$$

Da

$$\begin{aligned} \text{Cov}(\mathbf{V}, t_n) &= \text{Cov}(w'_n \mathbf{V}_n, t_n) + \text{Cov}(S_n, t_n) = \text{Cov}(S_n, t_n) \leq \text{Var}(S_n) \text{Var}(t_n) \\ \text{Var}(t_n) &\leq \text{Var}(v) < \infty \\ \text{Var}(S_n) &\rightarrow 0 \text{ für } n \rightarrow \infty \end{aligned}$$

gelten, folgt $\text{Cov}(\mathbf{V}, t_n) \rightarrow 0$.

Weiters folgt, dass

$\text{Cov}((b_n w_n - b)\mathbf{V}, t_n) = \text{Cov}(b_n w_n \mathbf{V}, t_n) - \text{Cov}(b\mathbf{V}, t_n) = \text{Cov}(b_n w_n \mathbf{V}, t_n)$ wegen der Beschränktheit von b_n und w_n gegen 0 konvergiert.

Da $\text{Cov}(\mathbf{V}, t) = 0$ wegen (6.14) gilt, ist auch (6.17) erfüllt. Somit reduziert sich die linke Seite von (6.16) auf eine Summe zweier positiver Summanden, wodurch die quadratische Konvergenz von $(t - t_n) \rightarrow 0$ und dadurch die von $\text{Proj}(v|K_n) \rightarrow \text{Proj}(v|K)$ durch (6.15) gezeigt ist. \square

Kapitel 7

Appendix B

Beweise für die Methode von SW

Unter den Annahmen A1, B1', F, und G aus dem Kapitel 4.2 gelten folgende Korollare.

Korollar 7.0.1.

$$plim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n e_{it}^2 = O(1).$$

D.h., die Summe $n^{-1} \sum_{i=1}^n e_{it}^2$ bleibt mit Wahrscheinlichkeit 1 beschränkt.

Beweis (Korollar 7.0.1)

$$n^{-1} \sum_{i=1}^n e_{it}^2 = n^{-1} \sum_{i=1}^n \tau_{ii,t} + n^{-1} \sum_{i=1}^n (e_{it}^2 - \tau_{ii,t})$$

Da $n^{-1} \sum_{i=1}^n \tau_{ii,t} \leq \sum_{i=1}^n \sum_{j=1}^n |\tau_{ij,t}|/n$ und daher nach Annahme G2 beschränkt ist (d.h., $O(1)$ ist), genügt es zu zeigen, dass der zweite Teil in Wahrscheinlichkeit gegen 0 konvergiert. Dies ist erfüllt, da sogar die Konvergenz im quadratischen Mittel folgendermaßen zu zeigen ist:

$$\begin{aligned} \mathbb{E} \left[n^{-1} \sum_{i=1}^n (e_{it}^2 - \tau_{ii,t}) \right]^2 &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n Cov(e_{it}^2, e_{jt}^2) \\ &\leq n^{-2} \sum_{i=1}^n \sum_{j=1}^n |Cov(e_{it}^2, e_{jt}^2)| \leq n^{-2} \sup_{t,s} \sum_{i=1}^n \sum_{j=1}^n |cov(e_{is}e_{it}, e_{js}e_{jt})| \rightarrow 0 \end{aligned}$$

wegen Annahme G3. □

Beweis (Lemma 4.2.2)

$$\begin{aligned}
n^{-2}\gamma'\hat{\Gamma}^e\gamma &= (n^2T)^{-1}\gamma'e'e\gamma \\
&= (n^2T)^{-1}\sum_{t=1}^T\sum_{i=1}^n\sum_{j=1}^n\gamma_ie_{it}e_{jt}\gamma_j = n^{-2}\sum_{i=1}^n\sum_{j=1}^n\gamma_i\gamma_j T^{-1}\sum_{t=1}^Te_{it}e_{jt} \\
&\leq \left(n^{-2}\sum_{i=1}^n\sum_{j=1}^n\gamma_i^2\gamma_j^2\right)^{1/2} \left(n^{-2}\sum_{i=1}^n\sum_{j=1}^n\left(T^{-1}\sum_{t=1}^Te_{it}e_{jt}\right)^2\right)^{1/2} =
\end{aligned}$$

(diese Ungleichung gilt, da $\sum_i a_i b_i \leq (\sum_i a_i^2)^{1/2}(\sum_i b_i^2)^{1/2}$ mit $a_i = \gamma_i\gamma_j$ und $b_i = T^{-1}\sum_{t=1}^Te_{it}e_{jt}$)

$$\begin{aligned}
&= (n^{-2}(\gamma'\gamma)^2)^{1/2} \left(n^{-2}\sum_{i=1}^n\sum_{j=1}^n\left(T^{-1}\sum_{t=1}^Te_{it}e_{jt}\right)^2\right)^{1/2} \\
&= \left(n^{-2}\sum_{i=1}^n\sum_{j=1}^n\left(T^{-1}\sum_{t=1}^Te_{it}e_{jt}\right)^2\right)^{1/2} = \left((nT)^{-2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{s=1}^Te_{it}e_{is}e_{jt}e_{js}\right)^{1/2}
\end{aligned}$$

Da im Folgenden die Konvergenz des letzten Ausdrucks im quadratischen Mittel gegen 0 gezeigt wird, folgt das Gewünschte.

$$\begin{aligned}
&\mathbb{E}\left[\left((nT)^{-2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{s=1}^Te_{it}e_{is}e_{jt}e_{js}\right)^{1/2}\right]^2 = \mathbb{E}\left[(nT)^{-2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{s=1}^Te_{it}e_{is}e_{jt}e_{js}\right] \\
&= (nT)^{-2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{s=1}^T\gamma_{i,t,s}\gamma_{j,t,s} \\
&+ (nT)^{-2}\sum_{i=1}^n\sum_{j=1}^n\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}[(e_{it}e_{is} - \gamma_{i,t,s})(e_{jt}e_{js} - \gamma_{j,t,s})]
\end{aligned}$$

mit $\gamma_{i,t,s} = \mathbb{E}(e_{it}e_{is})$.

Der erste Teil konvergiert gegen 0, da

$$\begin{aligned}
(nT)^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T \gamma_{i,t,s} \gamma_{j,t,s} &= T^{-2} \sum_{t=1}^T \sum_{s=1}^T \left(n^{-1} \sum_{i=1}^n \gamma_{i,t,s} \right) \left(n^{-1} \sum_{j=1}^n \gamma_{j,t,s} \right) \\
&= T^{-2} \sum_{t=1}^T \sum_{s=1}^T \left(n^{-1} \sum_{i=1}^n \mathbb{E}(e_{it}e_{is}) \right) \left(n^{-1} \sum_{j=1}^n \mathbb{E}(e_{jt}e_{js}) \right) = T^{-2} \sum_{t=1}^T \sum_{s=1}^T \gamma_{n,t}(s) \gamma_{n,t}(s) \\
&= T^{-2} \sum_{t=1}^T \sum_{s=1}^T \gamma_{n,t}(s)^2 \leq T^{-2} \left(T \sup_t \sum_{s=1}^T \gamma_{n,t}(s) \right) = T^{-1} \sup_t \sum_{s=1}^T \gamma_{n,t}(s) \rightarrow 0 \text{ wegen G1,}
\end{aligned}$$

und der zweite Teil wegen

$$\begin{aligned}
&(nT)^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}[(e_{it}e_{is} - \gamma_{i,t,s})(e_{jt}e_{js} - \gamma_{j,t,s})] \\
&= (nT)^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \sum_{s=1}^T \text{Cov}(e_{it}e_{is}, e_{jt}e_{js}) \\
&\leq n^{-2} \sup_{t,s} \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(e_{it}e_{is}, e_{jt}e_{js})| \rightarrow 0 \text{ wegen G3.}
\end{aligned}$$

□

Korollar 7.0.2. Sei v_t eine Folge von Zufallsvariablen für die gilt

$$\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T v_t = O(1),$$

dann gilt

$$\text{plim}_{n,t \rightarrow \infty} \sup_{\gamma \in H} |T^{-1} \sum_{t=1}^T v_t (n^{-1} \sum_{i=1}^n \gamma_i e_{it})| = 0.$$

Beweis (Korollar 7.0.2)

$$\sup_{\gamma \in H} |T^{-1} \sum_{t=1}^T v_t (n^{-1} \sum_{i=1}^n \gamma_i e_{it})| \leq (T^{-1} \sum_{t=1}^T v_t^2)^{1/2} (\sup_{\gamma \in H} T^{-1} \sum_{t=1}^T (n^{-1} \sum_{i=1}^n \gamma_i e_{it})^2)^{1/2}$$

Da der erste Teil laut Annahme beschränkt ist, und der zweite Teil gegen 0 konvergiert, folgt die Behauptung:

$$\begin{aligned} \sup_{\gamma \in H} T^{-1} \sum_{t=1}^T (n^{-1} \sum_{i=1}^n \gamma_i e_{it})^2 &= \sup_{\gamma \in H} (n^2 T)^{-1} \sum_{t=1}^T \sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j e_{it} e_{jt} \\ &= (n^2 T)^{-1} \gamma' e' e \gamma = n^{-2} \gamma' \hat{\Gamma}^e \gamma \rightarrow 0 \end{aligned}$$

wegen Lemma(4.2.2). \square

Korollar 7.0.3. $\text{plim}_{n,T \rightarrow \infty} \sup_{\gamma \in H} (n^2 T)^{-1} |\gamma' C F' e \gamma| = 0$

Beweis (Korollar 7.0.3)

$$\begin{aligned} \sup_{\gamma \in H} |(n^2 T)^{-1} \gamma' C F' e \gamma| &= \sup_{\gamma \in H} |(n^2 T)^{-1} \gamma' C \sum_{t=1}^T F_t e'_t \gamma| \\ &= \sup_{\gamma \in H} |(n^2 T)^{-1} \gamma' C \sum_{t=1}^T F_t \sum_{i=1}^n e_{it} \gamma_i| = \sup_{\gamma \in H} |(n^2 T)^{-1} \gamma' \sum_{j=1}^n C_j \sum_{t=1}^T F_{jt} \sum_{i=1}^n e_{it} \gamma_i| = \end{aligned}$$

(wobei C_j die j-te Spalte von C bedeutet)

$$\begin{aligned} &= \sup_{\gamma \in H} \left| \sum_{j=1}^n (n^{-1} \gamma' C_j) T^{-1} \sum_{t=1}^T F_{jt} (n^{-1} \sum_{i=1}^n e_{it} \gamma_i) \right| \\ &\leq \sup_{\gamma \in H} \left| \sum_{j=1}^n (n^{-1} \gamma' C_j) \right| \left| T^{-1} \sum_{t=1}^T F_{jt} (n^{-1} \sum_{i=1}^n e_{it} \gamma_i) \right| \\ &\leq \left(\max_j \sup_{\gamma \in H} |(n^{-1} \gamma' C_j)| \right) \left(\sum_{j=1}^n \sup_{\gamma \in H} |T^{-1} \sum_{t=1}^T F_{jt} (n^{-1} \sum_{i=1}^n e_{it} \gamma_i)| \right) \\ &\leq \left(\max_j (n^{-1} C'_j C_j) \right)^{1/2} \left(\sup_{\gamma \in H} (n^{-1} \gamma' \gamma) \right)^{1/2} \left(\sum_{j=1}^n \sup_{\gamma \in H} |T^{-1} \sum_{t=1}^T F_{jt} (n^{-1} \sum_{i=1}^n e_{it} \gamma_i)| \right) \\ &= \left(\sum_{j=1}^n \sup_{\gamma \in H} |T^{-1} \sum_{t=1}^T F_{jt} (n^{-1} \sum_{i=1}^n e_{it} \gamma_i)| \right) \rightarrow 0 \text{ in Wahrscheinlichkeit} \end{aligned}$$

Die letzte Gleichheit gilt, da die beiden vorderen Klammern beide gleich 1 sind. Und die Konvergenz der letzten Klammer folgt, da $T^{-1} \sum_{t=1}^T F_{jt}^2 \rightarrow \mu_j = \text{Var}(F_j)$ wegen Annahme F4 gilt, und daher F_{jt} die Bedingung für v_t aus Korollar(7.0.2) erfüllt. \square

Korollar 7.0.4.

$$plim_{n,t \rightarrow \infty} \sup_{\gamma \in H} |R(\gamma) - R(\gamma)^*| = 0$$

Beweis (Korollar 7.0.4)

$$\begin{aligned} \gamma' X' X \gamma &= \gamma' (CF' + e')(FC' + e)\gamma = \gamma' (CF' FC' + e' FC' + CF' e + e' e)\gamma \\ \Rightarrow \gamma' X' X \gamma - \gamma' CF' FC' \gamma &= \gamma' e' e \gamma + 2\gamma' CF' e \gamma \\ \Rightarrow R(\gamma) - R(\gamma)^* &= (n^2 T)^{-1} \gamma' e' e \gamma + (n^2 T)^{-1} 2\gamma' CF' e \gamma \\ \Rightarrow \sup_{\gamma \in H} |R(\gamma) - R(\gamma)^*| &= \sup_{\gamma \in H} |(n^2 T)^{-1} \gamma' e' e \gamma| + \sup_{\gamma \in H} |(n^2 T)^{-1} 2\gamma' CF' e \gamma| \rightarrow 0 \end{aligned}$$

Da die beiden Ausdrücke wegen Lemma(4.2.2) und Korollar(7.0.3) in Wahrscheinlichkeit gegen 0 konvergieren. \square

Beweis (Lemma 4.2.3) $|\sup_{\gamma \in H} R(\gamma) - \sup_{\gamma \in H} R(\gamma)^*| \leq \sup_{\gamma \in H} |R(\gamma) - R(\gamma)^*| \rightarrow 0$ wegen obigem Korollar. \square

Korollar 7.0.5.

$$plim_{n,t \rightarrow \infty} \sup_{\gamma \in H} R(\gamma)^* = \mu_1 = Var(F_1)$$

Beweis (Korollar 7.0.5) Der Beweis besteht hauptsächlich darin, $R(\gamma)^*$ derart umzuschreiben, dass die nicht überraschende Konvergenz mit einem Blick ersichtlich ist. Dazu beachte man, dass die Aufspaltungen

$$C' C / n = (C' C / n)^{1/2} ((C' C / n)^{1/2})' \text{ und } \gamma = C ((C' C / n)^{-1/2})' \delta + O$$

mit $O' C = 0$ stets möglich sind. Da

$$\begin{aligned} \gamma' \gamma / n &= (\delta' (C' C / n)^{-1/2} C' C ((C' C / n)^{-1/2})' \delta) / n + O' O / n \\ &= \delta' (C' C / n)^{-1/2} (C' C / n)^{1/2} ((C' C / n)^{1/2})' ((C' C / n)^{-1/2})' \delta + O' O / n \\ &= \delta' \delta + O' O / n \end{aligned}$$

gilt, folgt, dass $\delta' \delta \leq 1$ gilt. Deswegen kann man

$$\begin{aligned} \sup_{\gamma \in H} R(\gamma)^* &= \sup_{\delta | \delta' \delta \leq 1} (n^2 T)^{-1} \delta' (C' C / n)^{-1/2} C' C F' F C' C ((C' C / n)^{-1/2})' \delta \\ &= \sup_{\delta | \delta' \delta \leq 1} T^{-1} \delta' (C' C / n)^{-1/2} (C' C / n) F' F (C' C / n) ((C' C / n)^{-1/2})' \delta \\ &= \sup_{\delta | \delta' \delta \leq 1} \delta' ((C' C / n)^{1/2})' (T^{-1} F' F) (C' C / n)^{1/2} \delta \end{aligned}$$

anschreiben und sieht die Konvergenz, da $C' C / n \rightarrow I$ und $T^{-1} F' F \rightarrow \Gamma_0^F$ wegen der Annahmen F1 beziehungsweise F4 gelten, und daher der Ausdruck wegen der Stetigkeit der Eigenwerte gegen den größten Eigenwert von Γ_0^F konvergiert. \square

Korollar 7.0.6.

$$plim_{n,t \rightarrow \infty} \sup_{\gamma \in H} R(\gamma) = \mu_1 = Var(F_1)$$

Beweis (Korollar 7.0.6) Da bereits $plim_{n,t \rightarrow \infty} \sup_{\gamma \in H} R^*(\gamma) = \mu_1$ gilt, folgt durch Lemma(4.2.3) die Behauptung. \square

Korollar 7.0.7.

$$plim_{n,t \rightarrow \infty} R(\hat{C}_j)^* = \mu_1 = Var(F_1)$$

Beweis (Korollar 7.0.7) Da $\hat{C}_j = arg \sup_{\gamma \in H} R(\gamma)$ gilt, folgt aus Korollar(7.0.4) die Konvergenz, da ja dort sogar das Supremum des Abstandes gegen 0 konvergiert. \square

Korollar 7.0.8. Sei \hat{C}_1 die erste Spalte von \hat{C} und sei $S_1 = sign(\hat{C}'_1 C_1)$, dann gilt

$$plim_{n,t \rightarrow \infty} S_1 \hat{C}'_1 C/n = (1, 0, 0, \dots)'$$

Beweis (Korollar 7.0.8) Man betrachte erneut die Darstellung $\hat{C}_1 = C((C'C/n)^{-1/2})'\delta + O$ mit $O'C = 0$ und $\delta'\delta \leq 1$ und definiere $K_{n,T} = ((C'C/n)^{1/2})'(T^{-1}F'F)(C'C/n)^{1/2}$, sodass $R(\hat{C}_1)^* = \delta'K_{n,T}\delta$ gilt. Weiters gilt

$$\begin{aligned} R(\hat{C}_1)^* - \mu_1 &= \delta'(K_{n,T} - \Gamma_0^F)\delta + \delta'\Gamma_0^F\delta - \mu_1 \\ &= \delta'(K_{n,T} - \Gamma_0^F)\delta + (\delta_1^2 - 1)\mu_1 + \sum_{i=2}^r \delta_i^2 \mu_i \rightarrow 0 \end{aligned}$$

wegen Korollar(7.0.7). Da auch $plim_{n,T \rightarrow \infty} K_{n,T} = \Gamma_0^F$ gilt, und alle $\mu_i > 0$ sind, folgt, dass

$$plim_{n,T \rightarrow \infty} \delta_1^2 = 1; \quad plim_{n,T \rightarrow \infty} \delta_i^2 = 0, \text{ für } i > 2.$$

Da aus $\hat{C}_1 = C((C'C/n)^{-1/2})'\delta + O$ folgt, dass

$$\begin{aligned} plim_{n,T \rightarrow \infty} S_1 \hat{C}'_1 C/n &= plim_{n,T \rightarrow \infty} S_1 \delta'(C'C/n)^{-1/2} C/n \\ &= S_1 (+/- 1, 0, 0, \dots)I = (1, 0, 0, \dots) \end{aligned}$$

gilt, ist der Beweis erbracht. \square

Beweis (Lemma 4.2.4) Der Beweis für die erste Spalte von \hat{C} wurde im obigen Korollar gezeigt. Für die restlichen Spalten gelten jedoch alle vorangegangenen Korollare und Lemmata, indem man iterativ nur die Orthogonalräume der bereits verwendeten Eigenvektoren betrachtet. \square

Korollar 7.0.9. Sei v_t Folge von Zufallsvariablen, welche

$$plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T v_t^2 = a, \text{ und } plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T F_t v_t = b,$$

dann gilt das selbe auch für den Schätzer von F_t

$$plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T S \hat{F}_t v_t = b.$$

Beweis (Korollar 7.0.9)

$$1/T \sum_{t=1}^T S \hat{F}_t v_t = (nT)^{-1} \sum_{t=1}^T S \hat{C}' x_t v_t = 1/T \sum_{t=1}^T S(n^{-1} \hat{C}' C) F_t v_t + (nT)^{-1} \sum_{t=1}^T S \hat{C}' e_t v_t$$

Da aus Lemma(4.2.4) die Konvergenz $plim_{T \rightarrow \infty} S(n^{-1} \hat{C}' C) = I$ folgt, konvergiert der erste Ausdruck wegen der ersten Annahme gegen b . Das heißt, es bleibt zu zeigen, dass der zweite Teil gegen 0 strebt. Dieser zweite Teil ist ein stochastischer Vektor der Dimension r , für dessen j -tes Element Folgendes gilt

$$\begin{aligned} |(nT)^{-1} \sum_{t=1}^T S_j \hat{C}'_j e_t v_t| &= |1/T \sum_{t=1}^T v_t 1/n \sum_{k=1}^n \hat{c}_{jk} e_{tk}| \\ &\leq \sup_{\gamma \in H} |1/T \sum_{t=1}^T v_t 1/n \sum_{k=1}^n \gamma_{jk} e_{tk}| \rightarrow 0. \end{aligned}$$

Die Ungleichung gilt, da \hat{C}_j in H enthalten ist und die Konvergenz gegen 0 folgt aus Korollar(7.0.2). \square

Korollar 7.0.10.

1. $plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T S \hat{F}_t F_t' = \Gamma_0^F$
2. $plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T S \hat{F}_t \hat{F}_t' = \Gamma_0^F$
3. $plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T S \hat{F}_t \omega_t' = \Gamma_0^{F\omega}$
4. $plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T S \hat{F}_t \varepsilon_{t+h} = 0$

Beweis (Korollar 7.0.10)

1. Da F_{jt} wegen Annahme F4 beide Bedingungen von Korollar(7.0.9) für v_t erfüllt, folgt die Konvergenz.
2. $plim_{T \rightarrow \infty} 1/T \sum_{t=1}^T \hat{F}_{jt}^2 = \mu_j$ gilt, da $1/T \sum_{t=1}^T \hat{F}_{jt}^2 = R(\hat{C}_j)^*$ per Definition ist, und $plim_{T \rightarrow \infty} R(\hat{C}_j)^* = \mu_j$ wegen Korollar(7.0.6) gilt. Dadurch ist die erste Bedingung von Korollar(7.0.9) mit $v_t = S_j \hat{F}_{jt}$ erfüllt. Die zweite Bedingung wird in (1.) gezeigt. Dadurch folgt die Konvergenz.
3. Da ω_{jt} wegen Annahme H2 die Bedingungen von Korollar(7.0.9) für v_t erfüllt, folgt die Konvergenz.
4. ε_{t+h} erfüllt wegen Annahme H4 die erste und wegen H3 die zweite Bedingung für v_t aus Korollar(7.0.9). Damit folgt das Gewünschte. \square

Kapitel 8

R Code

```
backtest<-function(data,tzurueck,tfaktorenbestimmen,q,s){

# tzurueck heißt wieoft soll der forecast mit den tatsächlichen verglichen werden (h = 1)
# tfaktorenbestimmen ist das Zeitintervall auf dem die Faktoren berechnet werden

# data hat die Dimension (tzurueck + tfaktorenbestimmen) x n,
# wobei der letzte Zeitpunkt oben steht!

# data:      (T,1) ..... (T,n)
#            (T-1,1) .....(T-1,n)
#            .           .
#            .           .
#            .           .
#            (T-tzurueck+1,1) ..... hierfür wird der 1. forecast berechnet
#                                     (steht dann in der forecastmatrix ganz unten)
#            (T-tzurueck,1) .... -
#            .           .           |
#            .           .           | diese Zeitpunkte werden verwendet
#                                     | - um die Faktoren zu bestimmen,
#                                     | und der Zeitpunkt (T-tzurueck,)
#                                     | wird für den 1. forecast verwendet
#            (T-tzurueck-tfaktorenbestimmen+1,1) -
#
#
#
r=q*(s+1)

rfhlrvec<-qvec<-rswvec<-forecastfhlmatrix<-forecastswmatrix<-NULL
trueforecastfhlmatrix<-trueforecastswmatrix<-NULL

for(tht in tzurueck:1){

  Y=data[c((tht+1):(tht+tfaktorenbestimmen)),]

  h=1      # forecasthorizont
  wieoft=10 # wieviele Paare (n,t) verwendet werden sollen
            # (für das Informationskriterium für die Anzahl der dyn. Faktoren)
  freqn=2  # in welchem Abstand die n's verwendet werden
```

```

if(tfaktorenbestimmen == 200) freqt=5 else freqt=1 # in welchem Abstand die t's
# verwendet werden

allnj=seq(1,wieoft*freqn,by=freqn)
alltj=seq(1,wieoft*freqt,by=freqt)
nj=allnj[length(allnj)] # das größte nj
tj=alltj[length(alltj)]

counter=0

# dichte, comdichte und wndichte sind alle folgendermaßen geordnet:
# frequenzen von -pi bis pi

n=ncol(Y)
t=nrow(Y)

# qmax ... die maximale Anzahl dynamischer Faktoren

qmax=sqrt(n)

# für welche c's das durchgeführt werden soll

c=c(seq(0,2,by=.01))
tc=length(c)

result=matrix(0,nrow=tc,ncol=length(allnj))

# für verschiedene nj's alles durchrechnen
for(u in 1:wieoft){
  counter=counter+1

  t=nrow(Y)
  n=ncol(Y)

  X=Y[c(1:(t-tj+alltj[u])),c(1:(n-nj+allnj[u]))]

  n=ncol(X)
  t=nrow(X)
  mt=floor(sqrt(t))
  # covmat sind 2 arrays die zusammen die gesamte Kovarianzfkkt
  # einschließlich gamma(mt) beeinhalteten. Beachte: covmat2 hat
  # ein leeres sheet (um anschließend die dichte besser berechnen zu können)

  covmat=array(0,c(n,n,(mt+1)))
  covmat2=array(0,c(n,n,(mt+1)))

  covmat[,1]=t(X)%*%X/t
  for(k in 1:mt){
    covmat[,,(k+1)]=t(X[-c((t+1-k):t),])%*%X[-c(1:k),]/t
    covmat2[,,(k+1)]=t(X[-c(1:k),])%*%X[-c((t+1-k):t),]/t
  }

  l=c(seq(0:0,length=(2*mt+1)))
  for(k in 1:mt){
    l[mt+1+k]=pi/(mt+0.5)*k
    l[k]=pi/(mt+0.5)*(-(mt+1-k))
  }
  l[mt+1]=0
}

```

```

# w(x) wird hier mit 1-|x| berechnet = (1-(j/mt))

dichte=array(0,c(n,n,(2*mt+1)))
for(k in 1:(2*mt+1)){

# nicht vergessen: f(lambda)=f(-lambda)'
# d.h. die 2. Hälfte bräuchte man nicht ausrechnen

dichte[,k]=covmat[,1]
i=1i
for(j in 1:mt){
dichte[,k]=dichte[,k]+(covmat[, (j+1)]*exp(-i*j*1[k])+
covmat2[, (j+1)]*exp(i*j*1[k]))*(1-(j/mt))
}
}
dichte=dichte/(2*pi)

# spaltenweise stehen die Eigenwerte der Dichtefunktion für jede Frequenz
# für -lambda und +lambda stehen hier die selben Eigenwerte

lambda=matrix(0,nrow=n,ncol=(2*mt+1))
for(j in 1:(2*mt+1)){
lambda[,j]=eigen(dichte[,j],only.values=TRUE)$values
}

# Penaltyfunktion
pnt=(min(n,mt^2,sqrt(t)*mt^(-1/2)))^(-1)*log(min(n,mt^2,sqrt(t)*mt^(-1/2)))

# Informationskriterienmatrix

IC=matrix(0,ncol=qmax,nrow=tc)

for(k in 1:tc){
for(j in 1:qmax){
IC[k,j]=log(sum(lambda[-c(1:j),])/ (n*(2*mt+1)))+c[k]*j*pnt
}
}

# die relevanten Ergebnisse werden hier gesammelt

for(k in 1:tc){
result[k,counter]=which.min(IC[k,])
}
} # u

Sc=matrix(0,nrow=tc,ncol=1)
for(k in 1:tc){
Sc[k,1]=sqrt(sum((result[k,]-mean(result[k,]))^2)/counter)
}

# automatische Bestimmung von q indem man das 2.Intervall von 0er von Sc
# als "das" stabile Intervall betrachtet

```

```

Sc = Sc
for(i in 1:tc){
  if(Sc[i,1]==0) Sc[i,1]=1
  else break
}
Scmin=which.min(Sc)
qfhlr=result[Scmin,ncol(result)]

# Berechnung der Dichten für die Common-schocks und die Idiosyncratic-schocks

comdichte=array(0,c(n,n,(2*mt+1)))
wndichte=array(0,c(n,n,(2*mt+1)))

if(qfhlr>1){
  for(k in 1:(2*mt+1)){
    comdichte[,k]=eigen(dichte[,k])$vectors[,c(1:qfhlr)]%*%
    diag(eigen(dichte[,k])$values[c(1:qfhlr)])%*%
    Conj(t(eigen(dichte[,k])$vectors[,c(1:qfhlr)]))

    wndichte[,k]=dichte[,k]-comdichte[,k]
  }
} else {
  for(k in 1:(2*mt+1)){
    comdichte[,k]=eigen(dichte[,k])$values[c(1:qfhlr)]*
    eigen(dichte[,k])$vectors[,c(1:qfhlr),drop=F]%*%
    Conj(t(eigen(dichte[,k])$vectors[,c(1:qfhlr),drop=F]))

    wndichte[,k]=dichte[,k]-comdichte[,k]
  }
}

# gamma ist die Kovarianzmatrix(matrizen) der Common-schocks

i=li
gamma=array(0,c(n,n,(h+1)))
for(j in 1:(h+1)){
  gamma[,j]=0
  for(k in 1:(2*mt+1)){
    gamma[,j]=comdichte[,k]*exp(i+1[k]*(j-1))+gamma[,j]
  }
}
gamma=gamma/(2*mt+1)*2*pi
gamma=Re(gamma)

# gammawn ist die Kovarianzmatrix der Idio-schocks
# es wird ja nur die gammawn(0) benötigt, und die Nicht-diagonal-Elemente
# werden auf 0 gesetzt

gammawn=matrix(0,ncol=n,nrow=n)
for(k in 1:(2*mt+1)){
  gammawn=wndichte[,k]+gammawn
}
gammawn=diag(diag(gammawn))/(2*mt+1)*2*pi
gammawn=Re(gammawn)

```

```

#####
# FHLR
#####

# verallgemeinerte Eigenvektoren

E=Re(eigen(gammawn)$vectors)%*%diag((eigen(gammawn)$values)^(-1/2))
gammaneu=t(E)%*%gamma[,1]%*%E
C=eigen(gammaneu)$vectors

# das sind sie jetzt

F=E%*%C

maxk=min(round(n/2),round(t/2)) # maximale Anzahl an statischen Faktoren
m=matrix(0,ncol=maxk,nrow=(4))

for(k in 1:maxk){
  chidach=gamma[,2]%*%F[,c(1:k)]%*%solve(t(F[,c(1:k)])%*%covmat[,1]%*%
  F[,c(1:k)])%*%t(F[,c(1:k)])%*%t(Y[-1,])

  chidach=t(chidach)
  res1=sum((Y[-t,]-chidach)^2)

  # die 3 Informationskriterien wie bei Stock + Watson

  chidach2=gamma[,1]%*%F[,c(1:k)]%*%solve(t(F[,c(1:k)])%*%covmat[,1]%*%
  F[,c(1:k)])%*%t(F[,c(1:k)])%*%t(Y)

  chidach2=t(chidach2)
  res12=sum((Y-chidach2)^2)
  m[1,k]=res1/(n*t)
  m[2,k]=log(res12/(n*t))+k*((n+t)/(n*t))*log(n*t/(n+t))
  m[3,k]=log(res12/(n*t))+k*((n+t)/(n*t))*log(min(n,t))
  m[4,k]=log(res12/(n*t))+k*log(min(n,t))/min(n,t)
}
vergleich1=0
for(i in 1:nrow(m)){
  vergleich1[i]=which.min(m[i,])
}

# automatische Bestimmung der Anzahl der statischen Faktoren

if(vergleich1[4]!=maxk){
  rfh1r=vergleich1[4]
} else {
  rfh1r=ceiling(mean(c(vergleich1[2],vergleich1[3])))
}

forecastfh1r=gamma[,2]%*%F[,c(1:rfh1r)]%*%solve(t(F[,c(1:rfh1r)])%*%covmat[,1]%*%
F[,c(1:rfh1r)])%*%t(F[,c(1:rfh1r)])%*%Y[1,]

insamplefh1r=gamma[,1]%*%F[,c(1:rfh1r)]%*%solve(t(F[,c(1:rfh1r)])%*%covmat[,1]%*%
F[,c(1:rfh1r)])%*%t(F[,c(1:rfh1r)])%*%t(Y)

```

```

#####
# stock + watson
#####

# die Faktoren

S=eigen(covmat[, , 1])$vectors

mstock=matrix(0, ncol=maxk, nrow=(4))

for (i in 1:maxk){
  D=S[, c(1:i)]
  Fstock=t(D)%*%t(Y)
  Ydach=D%*%Fstock

  mstock[1, i]=sum((Y-t(Ydach))^2)/(n*t)
}

# 3 vorgeschlagene "Bestrafungsfunktionen"
gnt1=((n+t)/(n*t))*log(n*t/(n+t))
gnt2=((n+t)/(n*t))*log(min(n,t))
gnt3=log(min(n,t))/min(n,t)

# 3 vorgeschlagene Kriterien
for(i in 1:maxk){
  mstock[2, i]=log(mstock[1, i])+i*(gnt1)
  mstock[3, i]=log(mstock[1, i])+i*(gnt2)
  mstock[4, i]=log(mstock[1, i])+i*(gnt3)
}

vergleich2=0
for(i in 1:nrow(mstock)){
  vergleich2[i]=which.min(mstock[i, ])
}

# automatische Bestimmung der Anzahl der statischen Faktoren

if(vergleich2[4]!=maxk){
  rsw=vergleich2[4]
} else {
  rsw=ceiling(mean(c(vergleich2[2], vergleich2[3])))
}

forecastsw=covmat[, , 2]%*%S[, c(1:rsw)]%*%solve(t(S[, c(1:rsw)])%*%covmat[, , 1]%*%
S[, c(1:rsw)])%*%t(S[, c(1:rsw)])%*%(Y[1, ])

insamplesw=covmat[, , 1]%*%S[, c(1:rsw)]%*%solve(t(S[, c(1:rsw)])%*%covmat[, , 1]%*%
S[, c(1:rsw)])%*%t(S[, c(1:rsw)])%*%t(Y)

# sammeln der Ergebnisse
qvec<-c(qvec, qfhlr)
rfhlrvec<-c(rfhlrvec, rfhlr)
rswvec<-c(rswvec, rsw)

```

```

forecastfhlrmatrix<-cbind(forecastfhlr,forecastfhlrmatrix)
forecastswmatrix<-cbind(forecastsw,forecastswmatrix)

#####
#
# mit den tatsächlichen q,r
#
#####

# Berechnung der Dichten für die common-schocks und die idiosyncratic-schocks

truecomdichte=array(0,c(n,n,(2*mt+1)))
truewndichte=array(0,c(n,n,(2*mt+1)))

if(q>1){
  for(k in 1:(2*mt+1)){
    truecomdichte[,k]=eigen(dichte[,k])$vectors[,c(1:q)]%*%
    diag(eigen(dichte[,k])$values[c(1:q)])%*%
    Conj(t(eigen(dichte[,k])$vectors[,c(1:q)]))

    truewndichte[,k]=dichte[,k]-truecomdichte[,k]
  }
} else {
  for(k in 1:(2*mt+1)){
    truecomdichte[,k]=eigen(dichte[,k])$values[c(1:q)]*
    eigen(dichte[,k])$vectors[,c(1:q),drop=F]%*%
    Conj(t(eigen(dichte[,k])$vectors[,c(1:q),drop=F]))

    truewndichte[,k]=dichte[,k]-truecomdichte[,k]
  }
}

# truegamma ist die Kovarianzmatrix(matrizen) der Common-schocks

i=1i
truegamma=array(0,c(n,n,(h+1)))
for(j in 1:(h+1)){
  truegamma[,j]=0
  for(k in 1:(2*mt+1)){
    truegamma[,j]=truecomdichte[,k]*exp(i*1[k]*(j-1))+truegamma[,j]
  }
}
truegamma=truegamma/(2*mt+1)*2*pi
truegamma=Re(truegamma)

# gammawn ist die Kovarianzmatrix der Idio-schocks
# es wird ja nur die gammawn(0) benötigt, und die Nicht-diagonal-Elemente
# werden auf 0 gesetzt

truegammawn=matrix(0,ncol=n,nrow=n)
for(k in 1:(2*mt+1)){
  truegammawn=truewndichte[,k]+truegammawn
}
truegammawn=diag(diag(truegammawn))/(2*mt+1)*2*pi
truegammawn=Re(truegammawn)

```

```

#####
# FHLR
#####

# verallgemeinerte Eigenvektoren

trueE=Re(eigen(truegammawn)$vectors)%*%diag((eigen(truegammawn)$values)^(-1/2))
truegammaneu=t(trueE)%*%truegamma[,1]%*%trueE
trueC=eigen(truegammaneu)$vectors

# das sind sie jetzt

trueF=trueE%*%trueC

trueforecastfhlr=truegamma[,2]%*%trueF[,c(1:r)]%*%solve(t(trueF[,c(1:r)])%*%
covmat[,1]%*%trueF[,c(1:r)])%*%t(trueF[,c(1:r)])%*%Y[1,]

trueinsamplefhlr=truegamma[,1]%*%trueF[,c(1:r)]%*%solve(t(trueF[,c(1:r)])%*%
covmat[,1]%*%trueF[,c(1:r)])%*%t(trueF[,c(1:r)])%*%t(Y)

#####
# SW
#####

trueforecastsw=covmat[,2]%*%S[,c(1:r)]%*%solve(t(S[,c(1:r)])%*%covmat[,1]%*%
S[,c(1:r)])%*%t(S[,c(1:r)])%*%Y[1,])

trueinsamplesw=covmat[,1]%*%S[,c(1:r)]%*%solve(t(S[,c(1:r)])%*%covmat[,1]%*%
S[,c(1:r)])%*%t(S[,c(1:r)])%*%t(Y)

# sammeln

trueforecastfhlrmatrix<-cbind(trueforecastfhlr,trueforecastfhlrmatrix)
trueforecastswmatrix<-cbind(trueforecastsw,trueforecastswmatrix)

} #for tht

# oben steht der "neueste" forecast
forecastfhlrmatrix<-t(forecastfhlrmatrix)
forecastswmatrix<-t(forecastswmatrix)

insamplefhlr<-t(insamplefhlr)
insamplesw<-t(insamplesw)

# Ergebnisse mit den tatsächlichen Faktoranzahlen

trueforecastfhlrmatrix<-t(trueforecastfhlrmatrix)
trueforecastswmatrix<-t(trueforecastswmatrix)

trueinsamplefhlr<-t(trueinsamplefhlr)
trueinsamplesw<-t(trueinsamplesw)

result<-list(qvec,rfhlrvec,rswvec,forecastfhlrmatrix,forecastswmatrix,insamplefhlr,
insamplesw,trueforecastfhlrmatrix,trueforecastswmatrix,trueinsamplefhlr,trueinsamplesw)

} # function

```

Literaturverzeichnis

- [1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York
- [2] Bai, J. and S. Ng, (2001). *Determining the Number of Factors in Approximate Factor Models*, *Econometrica*, Vol. 70, 191-221
- [3] Boivin, J. and S. Ng, (2005). *Understanding and Comparing Factor-Based Forecasts*, *International Journal of Central Banking*
- [4] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco
- [5] Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York
- [6] Chamberlain, G. (1983). *Funds, Factors, and Diversification in Arbitrage Pricing Models*, *Econometrica*, Vol. 51, No. 5
- [7] Chamberlain, G. and M. Rothschild (1983). *Arbitrage, Factor Structure, and Mean-Variance Analysis on large Assetmarkets*, *Econometrica*, Vol. 51, No. 5
- [8] Deistler, M. and C. Zinner (2007). *Modelling High Dimensional Time Series by Generalized Linear Factor Models: An Introductory Survey*, *Communications in Information and Systems*, Vol 7, 153-166
- [9] Forni, M. and M. Lippi (1999). *The generalized dynamic factor model: representation theory*, Modena, Rom
- [10] Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2006). *Opening the Black Box: Structural Factor Models with Large Cross-Sections*
- [11] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). *The generalized factor model: identification and estimation*, *The Review of Economics and Statistics* 82, 540-554

- [12] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2002). *The generalized dynamic factor model: one-sided estimation and forecasting*, CEPR Discussion Paper No. 3432
- [13] Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*, Publication No. 5, University of Oslo, Economic Institute
- [14] Hallin, M. and R. Liška (2006). *The generalized dynamic factor model: determining the number of factors*
- [15] Lancaster, P. and M. Tismenetsky (1985). *The Theory of Matrices* Academic Press: Orlando
- [16] Rao, C. R. (1965). *Linear Statistical Inference and Its Applications* John Wiley & Sons, New York
- [17] Ribarits, T. (1999). *Dynamic Principal Component Models*, Diplomarbeit, TU Wien
- [18] Scherrer, W. (1991). *Strukturtheorie von linearen dynamischen Fehler-in-den-Variablen Modellen mit diagonalem Fehler-Spektrum*, Doktorarbeit, TU Wien
- [19] Scherrer, W. and M. Deistler (1998). *A Structure Theory for Linear Dynamic Errors in Variables Models*, SIAM Journal on Control and Optimization, Vol. 36, No. 6, 2148-2175
- [20] Stock, J. H. and M. W. Watson (1999). *Diffusion indexes*
- [21] Stock, J. H. and M. W. Watson (2002a). *Forecasting using principal components from a large number of predictors*, Journal of the American Statistical Association 97, 1167-79