DIPLOMARBEIT

Faktormodelle mit idiosyncratischen Fehlern und PCA

Ausgeführt am Institut für

Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von Herrn
Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Deistler

durch

Dipl.-Ing. Alexander Juschitz
Afritschgasse 61
1220 Wien

Wien, am 1.April 2011

Danksagung

Diese Diplomarbeit entstand am Institut für Wirtschaftsmathematik an der Technischen Universität Wien unter der Leitung von Em.O.Univ.Prof.Dipl.-Ing.Dr.techn. Manfred Deistler. Ich habe schon meine erste Diplomarbeit im Zuge meines Diplomstudiums Wirtschaftsingenieurwesen-Maschinenbau unter seiner Betreuung geschrieben. Die damalige gute Betreuung veranlasste mich auch diese Diplomarbeit unter seiner Aufsicht zu schreiben. Ich möchte ihm für die intensive Betreuung danken. Durch Nachfragen und gezieltes Fragestellen wurden mir schnell Unklarheiten bewusst und somit konnten diese bis zur nächsten Sprechstunde beseitigt werden. Ihm verdanke ich auch meine Begeisterung für die Ökonometrie, die seit meiner ersten Vorlesung bei Herrn Prof. Deistler im Wintersemester 2008 anhält.

Des weiteren möchte ich mich bei meinen Freunden bedanken, die mich in meinem Leben begleitet haben und mir stets hilfreich zur Seite standen. Hier möchte ich im Speziellen MMag. Philipp Krasa für die Gespräche abseits des Studiums und M.Sc. Antoine Espinet für seine Hilfe bei den mathematischen Beweisen danken.

Bei meiner Freundin MASt Cécile Deterre möchte ich mich für ihr Verständnis bedanken, welches sie mir während der ganze Zeit entgegenbrachte. Sie schreibt derzeit an ihrer Doktorarbeit und kann nur zu gut die Probleme verstehen, die während des Abfassens einer wissenschaftlichen Arbeit entstehen können. Sie stand mir stets kritisch, offen und moralisch unterstützend während dieser Zeit zur Seite.

Ganz besonderer Dank gilt meiner Familie. Meinen Eltern verdanke ich die schönen Jahre des Studiums und zu einem großen Umfang die Tatsache, dass ich zu dem geworden bin, der ich heute bin. Bei meiner Schwester Barbara möchte ich mich für ihre Unterstützung abseits des Studiums bedanken. Sie half mir niemals weltfremd zu werden, sei es auf Grund des Studiums oder der Tatsache, dass ich mich ausschliesslich auf die Diplomarbeit konzentrieren wollte.

Alexander Juschitz Wien, März 2011

Inhaltsverzeichnis

Kapitel 1

Faktoren Analyse

1.1 Einleitung

Die Faktorenanalyse basiert auf einem Modell, in dem der beobachtete Vektor aufgeteilt wird in zwei Teile. Der erste Teil besteht aus den Linearkombinationen einer relativ kleinen Anzahl an latenten Variablen. Die latenten Variablen werden auch als Faktoren bezeichnet. Der zweite Teil besteht aus unbeobachtbaren Fehlern. Die Komponenten des Fehlervektors werden als unkorreliert und unabhängig angenommen. Die Analyse trennt die Ergebnisse der Faktoren von den Fehlern. Dieser Ansatz wird oft verglichen mit dem der Hauptkomponentenanalyse. Die Faktorenanalyse wurde ursprünglich entwickelt um die Resultate von psychologischen Tests zu analysieren. Die Methode wird auch in der Chemometrie und Ökonometrie angewandt.

Das Modell der Faktorenanalyse wird im Unterkapitel 1.2.1 diskutiert. In Abschnitt 1.2.2 befassen wir uns mit den notwendigen Bedingungen für Identifizierbarkeit der Modellparameter. Im Unterkapitel 1.3 leiten wir die Maximum-Likelihood-Schätzer der Modellparameter her.

1.2 Das Modell

1.2.1 Definition des Modells

Sei Y ein stochastischer Vektor. Der Vektor Y kann angeschrieben werden als

$$Y = m + Bu + \varepsilon. \tag{1.1}$$

Die Parameter des Modells sind m, B, Φ und Ψ. Für die Dimension der Parameter gilt

$$\begin{aligned} dim\ Y = dim\ m = dim\ \varepsilon &= p \times 1, \\ dim\ B &= p \times q, \\ dim\ u &= q \times 1, \end{aligned}$$

wobei wir annehmen, dass q fixiert ist und $q < p$ gilt. Sei ε und u Vektoren von Zufallsvariablen. Der stochastische Vektor u stellt die latenten Variablen dar. Wir bezeichnen die Elemente des Vektors u als Faktoren. Wir nehmen an, dass $\mathbb{E}u = 0$ gilt und bezeichnen die Kovarianzmatrix mit $\Phi = \mathbb{E}uu^T$. Der Vektor ε stellt den Fehlerterm dar. Für den Erwartungswert gilt $\mathbb{E}\varepsilon = 0$. Wir bezeichnen die Kovarianzmatrix mit $\Psi = \mathbb{E}\varepsilon\varepsilon^T$. Die Matrix Ψ sei diagonal. Wir nehmen an, dass ε von u unabhängig ist ($\mathbb{E}(\varepsilon u^T) = 0$ für $u, \varepsilon \sim N$). Im Allgemeinen nehmen wir an, dass u und ε normalverteilt sind. Die Matrix B wird auch als Ladungsmatrix bezeichnet. Sie hat vollen Spaltenrang. Der Spaltenvektor m sei skalar und konstant.

Wir nehmen an, dass die Beobachtung von Y identisch und unabhängig verteilt sind und von der Zeit unabhängig sind.

Die Parameter des Faktormodells sind ohne weitere Bedingungen nicht eindeutig bestimmbar. Sei $u = Cu^*$ ($u^* = C^{-1}u$) und $B^* = BC$, wobei C eine nichtsinguläre $q \times q$-Matrix ist. Dann kann Y aus Gleichung 1.1 neu angeschrieben werden als

$$Y = m + B^* u^* + \varepsilon. \tag{1.2}$$

Die Kovarianzmatrix des Zufallsvektors u^* wird zu $\mathbb{E}u^*u^{*T} = C^{-1}\Phi(C^{-1})^T = \Phi^*$. Das Modell mit den Parametern aus Gleichung 1.1 ist beobachtungs-equivalent zu dem Modell mit den Parametern aus Gleichung 1.2 unter der Voraussetzung, dass die Beobachtung von Y in beiden Fällen der Verteilung $N(m, \Omega)$ folgen.

Ein Teil der Unbestimmtheit kann beseitigt werden durch die Bedingung $\mathbb{E}uu^T = I$. Wenn wir annehmen, dass $\Phi = I$, dann gilt $\mathbb{E}u^*u^{*T} = C^{-1}(C^{-1})^T = I$ ($I = CC^T$). Die Unbestimmtheit kommt durch die Transformation durch eine orthogonale Matrix zustande. Die Annahme, dass Φ diagonal ist, bedeutet, dass die Elemente von u unabhängig sind.

Die Kovarianzmatrix von Y ist von der Form

$$\Omega = \mathbb{E}(Y - m)(Y - m)^T = \mathbb{E}(Bu + \varepsilon)(Bu + \varepsilon)^T = B\Phi B^T + \Psi. \tag{1.3}$$

Für $\mathbb{E}uu^T = I$ erhalten wir

$$\Omega = BB^T + \Psi. \tag{1.4}$$

1.2.2 Identifizierbarkeit

Wir nehmen an, dass q fixiert ist. Der Erwartungswert und die Kovarianzmatrix von Y sind eindeutig für gegebene Parameter m, B, Φ und Ψ. Wir fragen nun ob es eine injektive Abbildung von (m, Ω) auf (m, B, Φ, Ψ) existiert und welche Bedingungen dazu nötig sind. Wir untersuchen die Bedingungen, die es ermöglichen die Summanden der Gleichung 1.3 eindeutig zu bestimmen. Im weiteren Schritt untersuchen wir die Bedingungen, unter denen wir mit Hilfe von $B\Phi B^T$ die Parameter B und Φ eindeutig bestimmen können.

Sei die Kovarianzmatrix Ω und die Anzahl der Faktoren q gegeben sind. Wir untersuchen nun die Existenz und die Eindeutigkeit der Matrizen B, Φ und Ψ, die die Gleichung 1.3 erfüllen. Wir setzen die Restriktion, dass die Kovarianzmatrizen Φ positiv definit und Ψ positiv semidefinit sind und dass Ψ eine Diagonalmatrix ist. Da jedes Triplett B, Φ, Ψ, welches die Gleichung 1.3 erfüllt, in eine equivalente Form BC, $C^{-1}\Phi C^{-1}$ und Ψ transformiert werden kann, benötigen wir zusätzlich q^2 Bedingungen um die Beobachtungsequivalenz zu beseitigen . Die Anzahl der Elemente im beobachtbaren Ω und die Anzahl der Bedingungen für Eindeutigkeit bezüglich beträgt $\frac{1}{2}p(p+1) + q^2$. Die Anzahl der Elemente in B, Φ und Ψ, die wir bestimmen müssen, sind pq, $\frac{1}{2}q(q+1)$ und p.

Es gilt

$$\frac{1}{2}p(p+1) + q^2 - \left[pq + \frac{1}{2}q(q+1) + p\right] = \frac{1}{2}\left[(p-q)^2 - p - q\right]. \tag{1.5}$$

Ist die rechte Seite der Gleichung 1.5 positiv, ist die Existenz einer Lösung ungewiss, jedoch eindeutig bei Existenz. Ist die rechte Seite der Gleichung 1.5 negativ, so ist die Existenz einer Lösung anzunehmen, aber nicht notwendigerweise eindeutig. Ist die Gleichung 1.5 gleich Null kann eine Lösung existieren und eindeutig sein. Die Existenz einer Lösung hängt sehr davon ab, ob eine diagonale Matrix Ψ mit nichtnegativen Diagonalelementen existiert, sodass $\Omega - \Psi$ eine positiv semidefinite Matrix mit Rang q ist. [Anderson and Rubin, 1956] bespricht einen großen Teil der bekannten Ergebnissen zu diesem Problem.

Existiert eine Lösung und ist sie eindeutig, so wird das Modell als identifiziert bezeichnet. Wie oben besprochen, müssen q^2 Bedingungen für B und Φ gefunden werden, damit eine Transformation $B^* = BC$ und $\Phi^* = C^{-1}\Phi C^{T-1}$ für $C \neq I$ ausgeschlossen werden kann. Die Bedingung $\Phi = I$ bedeutet, dass die Transformationsmatrix C orthogonal sein muss. Wir führen die Bedingung ein, dass

$$\eta = B^T \Psi^{-1} B$$

diagonal ist. Diese Restriktion erweist sich als praktisch bei der Herleitung der Maximum-Likelihood-Schätzer der Parameter. Eine weitere Restriktion ist, dass die Diagonalelemente von η geordnet und unterschiedlich ($\eta_{11} > \eta_{22} > \ldots > \eta_{qq}$) sind. Eine Alternative für den Fall $\Phi = I$ wäre zu verlangen, dass die ersten q Zeilen von B eine untere Dreiecksmatrix bilden.

andere Bedingungen

Neben einer einfachen Modellstruktur oder dem Nullsetzen bestimmter Elemente gibt es die Möglichkeit die obere quadratische Teilmatrix von B als Einheitsmatrix zu definieren. Dies ist eine Bedingung für vollen Spaltenrang der Matrix B. Sei

$B^* = (B_1^{*T}, B_2^{*T})^T$ eine beliebige $p \times q$ Matrix und B_1^* quadratisch und nichtsingulär, dann erfüllt $B = B^* B_1^{*-1} = (I_q, B_2^T)^T$ die oben angeführte Bedingung. Aus Gleichung 1.3 folgt somit

$$\Omega = \begin{pmatrix} I & \Phi B_2^T \\ B_2 \Phi & B_2 \Phi B_2^T \end{pmatrix} + \Psi.$$

1.3 Maximum-Likelihood-Schätzer für orthogonale Zufallsfaktoren

In diesem Abschnitt wird die Herleitung der Maximum-Likelihood-Schätzer der Parameter behandelt, für den Fall, dass die Beobachtungen unabhängig und identisch normalverteilt sind. Seien $y_1, \ldots, y_T$ Beobachtungen von Y und wir nehmen an, dass sie unabhängig und identisch $N(m, \Omega)$ verteilt sind. Wir nehmen an, dass $\Phi = I$ gilt und das Ψ positiv definit Diagonalmatrix ist. Wir führen die Bedingung ein, dass $\eta = B^T \Psi^{-1} B$ diagonal sei.

Die Maximum-Likelihood-Funktion der Stichprobe lautet

$$l_\theta(y|m, \Omega) = \frac{1}{(2\pi)^{pT/2} |\Omega|^{T/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{T} (y_i - m)^T \Omega^{-1} (y_i - m) \right\}.$$

Da $\mathbb{E}Y = m$ gilt, ist $\hat{m} = \overline{y} = (1/T) \sum_{i=1}^T y_i$ der Maximum-Likelihood-Schätzer von m. Die Matrix $\sum_{i=1}^T (y_i - m)^T \Omega^{-1} (y_i - m)$ hat die Dimension 1×1 und entspricht ihrer Spur. Sei

$$S_y = \frac{1}{T} \sum_{i=1}^{T} (y_i - m)(y_i - m)^T.$$

Durch zyklisches Vertauschen der Terme in der Spur, Logarithmierung und Wegfall konstanter Terme erhalten wir

$$\begin{aligned} l_\theta^*(y_1, \ldots, y_T|\Omega) &= -\frac{T}{2} tr\left(\Omega^{-1} S_y\right) - \frac{T}{2} \ln|\Omega| \\ &= -\frac{T}{2}\left[tr\left((BB^T + \Psi)^{-1} S_y\right) + \ln|BB^T + \Psi|\right]. \end{aligned}$$

Wir minimieren die Gleichung

$$L = \ln|\Omega| + tr\left(\Omega^{-1} S_y\right),$$

welche nur von B und Ψ abhängt.

Wir erhalten

$$\frac{\partial L}{\partial \Psi} = \frac{\partial \ln|\Omega|}{\partial \Psi} + \frac{\partial \left[tr\left(\Omega^{-1} S_y\right)\right]}{\partial \Psi}.$$

Durch die Ableitungsregeln für Matrizen

$$\begin{aligned} \frac{\partial \ln|X|}{\partial \theta} &= tr\left[X^{-1}\left(\frac{\partial X}{\partial \theta}\right)\right] \\ \frac{\partial tr\left[A\left(X+B\right)^{-1} C\right]}{\partial X} &= -\left[\left(X+B\right)^{-1} CA\left(X+B\right)^{-1}\right]^T \end{aligned}$$

erhalten wir

$$\begin{aligned} \frac{\partial \ln|\Omega|}{\partial \Psi_{ii}} &= tr\left[\Omega^{-1}\left(\frac{\partial \Omega}{\partial \Psi_{ii}}\right)\right] \\ &= (\Omega)^{-1}_{(i,i)} \end{aligned}$$

wobei $(\partial\Omega)/(\partial\Psi_{ii})$ eine $p \times p$ Nullmatrix bildet mit Ausnahme des (i,i)-Elements mit Wert Eins. In Matrixschreibweise erhalten wir für die Ableitung

$$\frac{\partial \ln|\Omega|}{\partial \Psi} = diag\left(\Omega\right)^{-1}.$$

Wir erhalten

$$\begin{aligned}\frac{\partial L}{\partial \Psi} &= \frac{\partial \ln|BB^T+\Psi|}{\partial \Psi} + \frac{\partial \left[tr\left((BB^T+\Psi)^{-1}S_y\right)\right]}{\partial \Psi} \\ &= diag\left(BB^T+\Psi\right)^{-1} - \left(BB^T+\Psi\right)^{-1} S_y \left(BB^T+\Psi\right)^{-1} \frac{\partial \Psi}{\partial \Psi}\end{aligned}$$

wobei $\frac{\partial \Psi}{\partial \Psi} = I$ gilt. Durch Nullsetzen erhalten wir

$$\frac{\partial L}{\partial \Psi} = diag\left(\hat{B}\hat{B}^T+\hat{\Psi}\right)^{-1} - \left(\hat{B}\hat{B}^T+\hat{\Psi}\right)^{-1} S_y \left(\hat{B}\hat{B}^T+\hat{\Psi}\right)^{-1} = 0$$

bzw.

$$diag\left[\hat{\Omega}^{-1}(\hat{\Omega}-S_y)\hat{\Omega}^{-1}\right] = 0, \tag{1.6}$$

wobei $\hat{\Omega} = \hat{B}\hat{B}^T + \hat{\Psi}$. Gleichung 1.6 ist equivalent zu

$$diag(\hat{\Omega}) = diag(S_y).$$

Als Nächstes, werden wir die abgeänderte log-Likelihoodfunktion nach B differenzieren und erhalten

$$\begin{aligned}\frac{\partial L}{\partial B} &= \frac{\partial \ln|BB^T+\Psi|}{\partial B} + \frac{\partial \left[tr\left((BB^T+\Psi)^{-1}S_y\right)\right]}{\partial B} \\ &= \frac{1}{|BB^T+\Psi|}\frac{\partial|\Omega|}{\partial B} + tr\left[S_y(BB^T+\Psi)^{-1}\frac{\partial}{\partial B}S_y(BB^T+\Psi)^{-1}\right].\end{aligned}$$

Wir betrachten die Ableitung des ersten Termes nach dem Element $B_{k\tau}$ und erhalten

$$\frac{\partial \ln|BB^T+\Psi|}{\partial B_{k\tau}} = \sum_{j=1}^{p}(\Omega^{-1})_{kj}B_{j\tau}.$$

Wir können die Ableitung des ersten Termes in Matrixschreibweise neu anschreiben als

$$\frac{\partial \ln|\Omega|}{\partial B} = \Omega^{-1}B.$$

Für den zweiten Term erhalten wir

$$tr\left[S_y(BB^T+\Psi)^{-1}\frac{\partial}{\partial B_{ij}}S_y(BB^T+\Psi)^{-1}\right] = tr\left[\Omega^{-1}S_y\Omega^{-1}\frac{\partial \Omega}{\partial B_{ij}}\right].$$

Wir können das Ergebnis in Matrixschreibweise neu anschreiben als

$$\frac{\partial}{\partial B} tr\left[S\left(BB^T+\Psi\right)^{-1}\right] = \Omega^{-1}S_y\Omega^{-1}B.$$

Wir erhalten für die Ableitung des abgeänderten negativen log-Likelihoodfunktion nach dem Parameter B

$$\frac{\partial L}{\partial B} = \Omega^{-1}B - \Omega^{-1}S_y\Omega^{-1}B.$$

Durch Nullsetzen erhält man

$$\left(\hat{\Omega} - S_y\right)\hat{\Omega}^{-1}\hat{B} = 0. \tag{1.7}$$

Die Gleichungen 1.6 und 1.7 begründen das Lawley Modell ([Lawley, 1953]).

Lemma 1.3.1. *Sei $\Omega = BB^T + \Psi$ eine Faktorzerlegung von Ω. Dann gilt*

$$\Omega^{-1} = \Psi^{-1} - \Psi^{-1}B\left(I+\eta\right)^{-1}B^T\Psi^{-1}. \tag{1.8}$$

Beweis. Durch Rechtmultiplikation der Gleichung 1.8 mit Ω erhält man

$$\begin{aligned}
\Omega^{-1}\Omega &= \left[\Psi^{-1} - \Psi^{-1}B\left(I+\eta\right)^{-1}B^T\Psi^{-1}\right]\left(BB^T+\Psi\right) \\
&= \Psi^{-1}\left(BB^T+\Psi\right) - \Psi^{-1}B\left(I+\eta\right)^{-1}B^T\Psi^{-1}\left(BB^T+\Psi\right) \\
&= \Psi^{-1}BB^T + I - \Psi^{-1}B(I+\eta)^{-1}B^T\Psi^{-1}BB^T \\
&\quad -\Psi^{-1}B(I+\eta)^{-1}B^T\Psi^{-1}\Psi \\
&= \Psi^{-1}BB^T + I - \Psi^{-1}B(I+\eta)^{-1}\eta B^T - \Psi^{-1}B(I+\eta)^{-1}B^T \\
&= \Psi^{-1}BB^T + I - \Psi^{-1}BB^T \\
&= I,
\end{aligned}$$

wobei $\eta = B^T\Psi^{-1}B$ diagonal ist. □

Durch Rechtmultiplizieren der Gleichung 1.7 mit B erhalten wir folgendes Lemma

Lemma 1.3.2. *Sei $\Omega = BB^T + \Psi$ eine Faktorzerlegung von Ω. Dann gilt*

$$\Omega^{-1}B = \Psi^{-1}B\,(I+\eta)^{-1} \tag{1.9}$$

Beweis. Durch Rechtsmultiplikation der Gleichung 1.8 mit B erhalten wir

$$\begin{aligned}
\Omega^{-1}B &= \Psi^{-1}B - \Psi^{-1}B\,(I+\eta)^{-1}\,B^T\Psi^{-1}B \\
&= \Psi^{-1}B - \Psi^{-1}B\,(I+\eta)^{-1}\,\eta.
\end{aligned}$$

Wir haben vorausgesetzt, dass η diagonal ist mit Diagonalelementen ungleich Null. Wir erhalten durch Rechtmultiplikation von $\eta^{-1}(I+\eta)$

$$\begin{aligned}
\Omega^{-1}B\eta^{-1}\,(I+\eta) &= \Psi^{-1}B\eta^{-1}\,(I+\eta) - \Psi^{-1}B \\
\Omega^{-1}B\eta^{-1} + \Omega^{-1}B &= \Psi^{-1}B\eta^{-1} + \Psi^{-1}B - \Psi^{-1}B \\
&= \Psi^{-1}B\eta^{-1}.
\end{aligned}$$

Rechtsmultiplikation von η führt zu

$$\Omega^{-1}B + \Omega^{-1}B\eta = \Psi^{-1}B$$

und in weiter Folge zu

$$\Omega^{-1}B = \Psi^{-1}B\,(I+\eta)^{-1}.$$

Durch Einsetzen der Gleichung 1.9 in Gleichung 1.7 erhalten wir

$$\left(\hat{\Omega} - S_y\right)\hat{\Psi}^{-1}\hat{B}(I+\hat{\eta})^{-1} = 0.$$

Durch Rechtsmultiplikation von $(I+\eta)$ und Ersetzen von Ω erhalten wir

$$\hat{B}\hat{B}^T\hat{\Psi}^{-1}\hat{B} + \hat{B} - S_y\hat{\Psi}^{-1}\hat{B} = 0$$

oder

$$S_y \hat{\Psi}^{-1} \hat{B} = \hat{B}(\hat{\eta} + I).$$

Durch Linksmultiplikation von $\hat{\Psi}^{-1/2}$ können wir schreiben

$$\left[\hat{\Psi}^{-1/2} S_y \hat{\Psi}^{-1/2} - (I + \hat{\eta})\right] \hat{\Psi}^{-1/2} \hat{B} = 0. \tag{1.10}$$

Sei $S^* = \hat{\Psi}^{-1/2} S_y \hat{\Psi}^{-1/2}$ eine gewichtete Stichprobenkovarianzmatrix. Die endgültige Form der Gleichungen ist

$$\left[S^* - (1 + \hat{\eta}_i)\right] \hat{\Psi}_i^{-1/2} \hat{B}_i = 0 \qquad i = 1, 2, \ldots, q,$$

wobei $\hat{\eta} + 1$ die Eigenwerte von S^* sind. □

Theorem 1.3.3. *Sei Y ein p-dimensionaler Zufallsvektor, sodass $Y \sim N(m, \Omega)$ und $\Omega = BB^T + \Psi$ gilt. Dann gilt:*

1. *Existiert eine eindeutige Matrix Ψ mit positiven Diagonalelementen, sodass die ersten q größten Eigenwerte von $\Omega^* = \Psi^{-1/2} \Omega \Psi^{-1/2}$ alle unterschiedlich und größer als Eins, und alle restlichen $p - q$ Eigenwerte gleich Eins sind, dann kann B eindeutig definiert werden, sodass gilt*

$$\left[\Omega^* - (1 + \hat{\eta}_i) I\right] \hat{\Psi}_i^{-1/2} \hat{B}_i = 0. \tag{1.11}$$

2. *Sei*

$$\Omega^* = \Psi^{-1/2} \Gamma \Psi^{-1/2} + I$$

die Varianz des skalierten Modells, wobei $\Gamma = BB^T$. Dann ist $\eta_i = B_i^T \Psi_{ii}^{-1} B_i$ mit $i = 1, 2, \ldots, q$ der i-te Eigewert von $\Psi^{-1/2} \Gamma \Psi^{-1/2}$ sodass $\eta_1 > \eta_2 > \ldots > \eta_q$ und $\eta_{q+1} = \eta_{q+2} = \ldots = \eta_p = 0$.

Beweis. 1. Die Kovarianzmatrix Ω^* ist eine nicht singuläre Matrix. Sie kann dargestellt werden als

$$(\Omega^* - \lambda_i I)\,\Pi_i = 0$$

für $\lambda_1 > \lambda_2 > \ldots > \lambda_p$. Ist Ψ eindeutig, so sind es auch B und Π und es gilt $\lambda_i = \eta_i + 1$ und $\Pi = \Psi^{-1/2}B_i$. Wir erhalten Gleichung 1.11

2. Es gilt $\Omega^* = \Psi^{-1/2}\Gamma\Psi^{-1/2} + I$. Durch Gleichung 1.10 erhalten wir

$$\begin{aligned} \left[\left(\Psi^{-1/2}\Gamma\Psi^{-1/2} + I\right) - (\eta_i + 1)\right]\Psi_{ii}^{-1/2}B_i &= 0 \qquad \text{oder} \\ \left(\Psi^{-1/2}\Gamma\Psi^{-1/2} - \eta_i I\right)\Psi Y_i^{-1/2}B_i &= 0. \end{aligned}$$

Da $\Gamma = BB^T$ eine $p \times p$-patrix und $\rho(\Gamma) = q$ und die letzten $p - q$ Eigenwerte von Γ Null sein müssen, nehmen die letzten $p - q$ Eigenwerte von Ω^* den Wert Eins an. □

Die Matrix $\Omega = BB^T + \Psi$ ist positiv definit für alle positiv semidefinite Matrizen BB^T und Ψ, wobei Ψ eine Diagonalmatrix ist. Die Matrix Ω kann ebenfalls positiv definit sein für nicht positiv semidefinite Diagonalmatrizen Ψ. Die Likelihoodfunktion kann zunehmen, wenn sich mehrere Diagonalelemente von Ψ dem Wert Null annähern. In diesem Fall, können die Ableitungen nicht Null gesetzt werden.

Die Gleichungen 1.8 und 1.9 können neu angeschrieben werden als Polynomgleichungen. Sie können jedoch nicht direkt gelöst werden. Es gibt zahlreiche iterative Verfahren um ein Parameterset zu bestimmen, wie zum Beispiel das Gradientenverfahren, das Verfahren von Newton-Raphson bzw Flechter-Powel [Lawley, 1953].

Da kein lokales Maximum für $\Psi_{ii} > 0$ für $i = 1, \ldots, q$ existieren muss, können auf $\hat{\Psi}$ mit $\hat{\psi}_{ii} < 0$ im iterativen Verfahren vorkommen. Dies widerspricht der Interpretation von Ψ als Kovarianzmatrix der Fehler. Für die Bedingung, dass $\psi_{ii} \geq 0$ für $i = 1, \ldots, p$ kann es sich um ein Randlösung handeln und manche Ableitungen werden nicht Null gesetzt.

Kapitel 2

Faktoren und Hauptkomponenten

2.1 Einleitung

[Velicer and Jackson, 1990] hat nachgewiesen, dass die Maximum-Likelihood-Schätzer der Parameter für die Faktoranalyse und jenen für die Hauptkomponentenanalyse approximativ zu ähnlichen Resultat führen können. [Velicer et al., 1982] bespricht auf Simulationen beruhende Studien, die aufzeigen, dass die Ladungsmatrizen beider Methoden sich ähneln und die durch die Faktoranalyse berechneten Faktoren durch Hauptkomponenten approximiert werden können.

Die Ergebnisse der beiden Verfahren können aber auch sehr unterschiedlich ausfallen. Es stellt sich die Frage nach den Bedingungen für die Faktoranalyse und die Hauptkomponentenanalyse, sodass die Ergebnisse ähnlich sind.

Teilweise wurde eine Antwort in [Bekker and Kano, 1990], [Schneeweiss, 1990] und

[Schneeweiss and Mathes, 1995] gegeben. Es wurde gezeigt, dass die Ergebnisse beider Analysen sich annähern, wenn die Fehlervarianzen ψ_i sehr klein sind. Hierbei ist "klein" ein relativer Begriff. Die größte Fehlervarianz ψ_{max} sollte klein sein im Vergleich zu d_q^2, dem kleinsten quadrierten Eigenwert von BB^T. In vielen Fällen wird diese Bedingung erfüllt. Dies erklärt, weshalb man zu ähnlichen Ergebnissen in der Faktoranalyse und in der Hauptkomponentenanalyse kommt ([Schneeweiss, 1990]).

Es gibt jedoch auch andere Gründe, weshalb es zu einer starken Ähnlichkeit der beiden Verfahren kommen kann. Sei δ die Differenz zwischen der größten Fehlervarianz ψ_{max} und der kleinsten Fehlervarianz ψ_{min}. Wir werden zeigen, dass für den Fall, dass die Differenz δ im Vergleich zu dem kleinsten quadrierten Eigenwert der Ladungsmatrix relativ klein ist, es zu einer hohen Ähnlichkeit kommt.

Für den Fall, dass $q > 1$, hat [Schneeweiss and Mathes, 1995] eine Antwort, wie die Ähnlichkeit der Ladungsmatrizen gemessen werden kann. [Schneeweiss and Mathes, 1995] führt eine Methode an, die invariant bezüglich nichtsingulären Transformationen ist.

Ein interessanter Fall tritt ein, wenn die p sehr groß wird im Vergleich zu q. Diese Situation tritt oft in der Psychometrie ein und ist typisch für den Kapitalmarkt, siehe [Chamberlain and Rothschild, 1983] und [Trzeinka, 1986].

In diesem Kapitel, werden wir kurz das Faktormodell besprechen und auf die Hauptkomponentenanalyse eingehen. Wir werden die nötigen Modellgleichungen und Eigenschaften angeben, die wir im Verlauf benötigen. Danach diskutieren wir die Methode, mit deren Hilfe wir die Ähnlichkeit zweier Ladungsmatrizen messen und führen die

Bedingungen an, für die die Resultate aus der Faktoranalyse gegen die der Hauptkomponentenanalyse konvergieren.

2.2 Das Faktormodell

Wir haben das Faktormodell in Abschnitt 1.2.1 besprochen. Durch Zentierung mittels $y = Y - m$, erhalten wir

$$y = Bu + \varepsilon. \tag{2.1}$$

Sei $\mathbb{E}(yy^T) = \Omega$ und $\mathbb{E}(\varepsilon\varepsilon^T) = \Psi$. Wir nehmen an, dass $\mathbb{E}(uu^T) = \Phi = I$ gilt und erhalten

$$\Omega = B\Phi B^T + \Psi = BB^T + \Psi. \tag{2.2}$$

Wir setzen voraus, dass q so klein gewählt wurde, dass es möglich ist Ω hinsichtlich der Räume $B\Phi B^T$ und Ψ eindeutig zerlegen zu können

Die Matrix B ist bei gegebenen Ψ bis auf eine orthogonale Transformation eindeutig bestimmt. Wir verwenden diese Eigenschaft und können B so wählen, dass $B^T B$ eine Diagonalmatrix ist mit Diagonalelementen in absteigender Reihenfolge. Wir verwenden diese Eigenschaft in Theorem 2.4.2.

2.3 Das Hauptkomponentenmodell

In der Hauptkomponentenanalyse wird versucht die Kovarianzmatrix Ω eines stochastischen Vektors y mit Dimension p durch $\hat{\Omega}$ von Rang $q(< p)$ zu approximieren, sodass $spur(\Omega - \hat{\Omega})$ minimiert wird.

Sei Ω eine positiv definite $p \times p$ Matrix. Sei Λ_0 eine $p \times p$ Diagonalmatrix aus den Eigenwerten von Ω. Die Spur von Λ_0 ist gleich der Spur von Ω. Wir setzen voraus, dass alle Eigenwerte von Vielfachheit Eins sind und dass die Diagonalelemente von Λ_0 in absteigender Reihenfolge geordnet sind. Sei $\tilde{B}_0$ eine $p \times p$ Matrix bestehend aus den entsprechenden Eigenvektoren von Ω. Es gilt $\tilde{B}_0 \tilde{B}_0^T = I$. Die Spalten von $\tilde{B}_0$ sind bis auf einen zufälligen Faktor ± 1 bestimmt. Die Kovarianzmatrix hat die Eigenwertdarstellung

$$\Omega = \tilde{B}_0 \Lambda_0 \tilde{B}_0^{-1}. \tag{2.3}$$

Sei q $(< p)$ gegeben. Wir stellen die Diagonalmatrix Λ_0 als Summe zweier Diagonalmatrizen Λ_1 und Λ_2 dar. Die ersten q Diagonalelemente der Matrix Λ_1 entsprechen den ersten q Diagonalelementen von Λ_0 in gleicher Reihenfolge. Alle andere Elemente der Matrix Λ_1 haben den Wert Null. Die letzten $p - q$ Diagonalelemente der Matrix Λ_2 entsprechen den letzten $p - q$ Diagonalelementen von Λ_0 in gleicher Reihenfolge. Alle andere Elemente der Matrix Λ_2 haben den Wert Null. Wir können die Gleichung 2.3 neu anschreiben als

$$\begin{aligned}
\Omega &= \tilde{B}_0 \left(\Lambda_1 + \Lambda_2\right) \tilde{B}_0^{-1} \\
&= \tilde{B}_0 \left[\begin{pmatrix} \lambda_1 & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \lambda_q & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} + \begin{pmatrix} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \lambda_{q+1} & & \\ \vdots & \ddots & \vdots & & \ddots & \\ 0 & \dots & 0 & & & \lambda_p \end{pmatrix} \right] \tilde{B}_0^{-1} \\
&= \tilde{B}_0 \Lambda_1 \tilde{B}_0^{-1} + \tilde{B}_0 \Lambda_2 \tilde{B}_0^{-1}
\end{aligned}$$

Sei Λ die $q \times q$ obere linke Teilmatrix von Λ_1. Sie enthält die q größten Eigenwerte von Ω in absteigender Reihenfolge. Sei $\tilde{B}$ eine $p \times q$ Matrix bestehend aus den ersten q Spalten von $\tilde{B}_0$. Die Matrix $\tilde{B}$ wird auch als Ladungsmatrix in der Hauptkomponentenanalyse bezeichnet.

Wir approximieren nun die Kovarianz Ω durch $\tilde{B}\Lambda\tilde{B}^T$. Die Matrix $\tilde{B}\Lambda\tilde{B}^T$ ist von Rang q. Die Spur von $\Omega - \tilde{B}\Lambda\tilde{B}^T$ wird minimiert, da Λ die q größten Eigenwerten von Ω enthält. Wir betrachten nun die Lineartransformation

$$\tilde{u} = \tilde{B}^T y. \tag{2.4}$$

Wir bezeichnen $\tilde{u}$ als den Vektor der Hauptkomponenten. Die Spalten von $\tilde{B}$ sind bis auf einen zufälligen Faktor ± 1 bestimmt, ähnliches gilt für $\tilde{u}$.

Es gilt $\mathbb{E}(\tilde{u}\tilde{u}^T) = \Lambda$. Auf Grund der Eigenwertzerlegung gilt

$$\Omega\tilde{B} = \tilde{B}\Lambda. \tag{2.5}$$

Sei $\tilde{\varepsilon}$ ein p-dimensionaler stochastischer Vektor. Der Vektor $\tilde{\varepsilon}$ gibt den Fehler an, der durch die Approximation entsteht. Der Vektor y kann angeschrieben werden als

$$y = \tilde{B}\tilde{u} + \tilde{\varepsilon}, \tag{2.6}$$

wobei $\mathbb{E}(\tilde{u}\tilde{\varepsilon}^T) = 0$ gilt. Auf Grund der Ähnlichkeit der Gleichungen 2.1 und 2.6 stellen wir die Frage, wie sehr die Vektoren u und $\tilde{u}$ und die Matrizen B und $\tilde{B}$ und zusammenhängen.

2.4 Korrelation zwischen den Faktoren und Hauptkomponenten

[Schneeweiss and Mathes, 1995] hat folgende Methoden zur Messung der Ähnlichkeit zweier beliebiger q-dimensionalen Zufallsvektoren ξ_1 und ξ_2 mit $\mathbb{E}\xi_i = 0$ und nichtsingulären Kovarianzmatrix $\mathbb{E}\xi_i\xi_i^T$ für $i = 1, 2$ beschrieben. Es wurde auch ein Verfahren diskutiert, mit dessen Hilfe man die Ähnlichkeit zweier $p \times q$ Matrizen B_1 und B_2 mit vollem Spaltenrang misst, wobei $q \leq p$

$$
\begin{aligned}
\rho^2 = \rho^2(\xi_1, \xi_2) &= \frac{1}{q} tr \left[\Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \right] = \frac{1}{q} tr P\left(\xi_1, \xi_2\right) && (2.7) \\
r^2 = r^2\left(B_1, B_2\right) &= \frac{1}{q} tr \left[\left(B_1^T B_1\right)^{-1/2} B_1^T B_2 \left(B_2^T B_2\right)^{-1} B_2^T B_1 \left(B_1^T B_1\right)^{-1/2} \right] \\
&= \frac{1}{q} tr R\left(B_1, B_2\right), && (2.8)
\end{aligned}
$$

wobei $\Sigma_{ij} = \mathbb{E}(\xi_i \xi_j^T)$.

Das Verfahren beruht auf dem Prinzip der kanonischen Korrelation. Der Faktor $1/q$ dient der Normalisierung. Für die $q \times q$ Matrizen in den eckigen Klammern, $P = P(\xi_1, \xi_2)$ und $R = R(B_1, B_2)$, gilt $0 \leq P \leq I$ und $0 \leq R \leq I$. Daraus folgt $0 \leq \rho^2 \leq 1$ und $0 \leq r^2 \leq 1$. Die Gleichung $\rho = 1$ bedeutet, dass $P = I$ und $\xi_1 = A\xi_2$ gilt mit Wahrscheinlichkeit 1. $r^2 = 1$ bedeutet, dass $R = I$ und $B_1 = B_2 A$. In beiden Fällen ist A eine $q \times q$ nichtsinguläre Transformationsmatrix. Für $\xi_i^* = A_i \xi_i$ und $B_i^* = B_i A_i$ wobei A_i nichtsinguläre ist, gilt $\rho^2(\xi_1^*, \xi_2^*) = \rho^2(\xi_1, \xi_2)$ und $r^2(B_1^*, B_2^*) = r^2(B_1, B_2)$.

Wir werden nun die Ähnlichkeit zwischen den Vektoren u und $\tilde{u}$ und zwischen den

Matrizen B und $\tilde{B}$ messen. Sei

$$\rho^2(u,\tilde{u}) = \frac{1}{q} tr \left\{ \mathbb{E}\left(u\tilde{u}^T\right) \left[\mathbb{E}\left(\tilde{u}\tilde{u}^T\right)\right]^{-1} \mathbb{E}\left(\tilde{u}u^T\right) \right\} = \frac{1}{q} trP \tag{2.9}$$

$$r^2\left(B,\tilde{B}\right) = \frac{1}{q} tr \left\{ \tilde{B}^T B \left(B^T B\right)^{-1} B^T \tilde{B} \right\} = \frac{1}{q} trR, \tag{2.10}$$

wobei nun P und R die Matrizen in den geschwungenen Klammern aus den Gleichungen 2.9 und 2.10 bezeichnen.

Mit Hilfe der Gleichungen 2.1 bis 2.5 erhalten wir

$$P = B^T \tilde{B} \Lambda^{-1} \tilde{B}^T B. \tag{2.11}$$

Beweis. Es gilt mit Hilfe von Gleichung 2.4

$$\mathbb{E}\left(\tilde{u}\tilde{u}^T\right) = \mathbb{E}\left(\tilde{B}^T y y^T \tilde{B}\right) = \tilde{B}^T \Omega \tilde{B}.$$

Durch Verwendung von Gleichung 2.5 erhalten wir

$$\mathbb{E}\left(\tilde{u}\tilde{u}^T\right) = \tilde{B}^T \Omega \tilde{B} = \tilde{B}^T \tilde{B} \Lambda = \Lambda \tag{2.12}$$

auf Grund von $\tilde{B}^T \tilde{B} = I$. Durch Transponierung der Gleichung 2.4 und anschließender Linksmultiplikation durch u erhalten wir

$$u\tilde{u}^T = u y^T \tilde{B}.$$

Durch Einsetzen der Gleichung 2.1 und Berechnung des Erwartungswertes kommen wir auf die Gleichung

$$\mathbb{E}\left(u\tilde{u}^T\right) = \mathbb{E}\left[u\left(u^T B^T + \varepsilon^T\right)\tilde{B}\right] = \mathbb{E}\left(uu^T\right) B^T \tilde{B} + \mathbb{E}\left(u\varepsilon^T\right)\tilde{B}.$$

Wir haben angenommen, dass die Faktoren u und die Fehler ε unkorreliert sind. Wir erhalten

$$\mathbb{E}\left(u\tilde{u}^T\right) = B^T\tilde{B}. \tag{2.13}$$

Mit Hilfe der Gleichungen 2.12 und 2.13 erhalten wir

$$\begin{aligned} P &= \mathbb{E}\left(u\tilde{u}^T\right)\left[\mathbb{E}\left(\tilde{u}\tilde{u}^T\right)\right]^{-1}\mathbb{E}\left(\tilde{u}u^T\right) \\ &= B^T\tilde{B}\Lambda^{-1}\tilde{B}^T B. \end{aligned}$$

□

Wir möchten die Bedingungen identifizieren, für die die Korrelation den maximalen Wert 1 annimmt.

Es wird angenommen, dass eine Folge von p_t dimensionalen Zufallsvektoren y_t mit $t = 1, 2, \ldots$ gegeben sei mit dem Faktormodell $y_t = B_t u_t + \varepsilon_t$ und q fixiert ist. Alle anderen Parameter dürfen sich mit t verändern. Die Dimension p des beobachtbaren Vektors y kann sich verändern und gegen Unendlich streben. Wir werden für die bisher erarbeiteten Aussagen den Index t weglassen, um die Lesbarkeit zu erhöhen. Dies betrifft alle Parameter außer q (und die Konstante c aus Annahme A, die wir später einführen werden).

Sei $\psi_{max} = \lambda_{max}(\Psi)$ der größte Eigenwert von Ψ und $\psi_{min} = \lambda_{min}(\Psi)$ der kleinste Eigenwert. Wir bezeichnen die Differenz zwischen dem größten und dem kleinsten Eigenwert mit $\delta = \psi_{max} - \psi_{min}$. Sei $D = diag(d_1, d_2, \ldots, d_q)$, wobei $d_1 \geq d_2 \geq \ldots \geq d_q \geq 0$ die Eigenwerte von $B^T B$ sind. Der kleinste Eigenwert d_q ist größer Null, da der Spaltenrang von B voll ist. Wir werden nun zwei Bedingungen betrachten. Die erste Bedingungen ist, dass $\lim_{t\to\infty}(\delta D^{-1}) = 0$,die zweite Bedingung $\lim_{t\to\infty}(\psi_{max} D^{-1}) = 0$ gilt. Die zweite Bedingung wurde schon von [Schneeweiss and Mathes, 1995] besprochen. Die zweite Bedingung beinhaltet die erste und ist deshalb strenger als die erste

$$\lim_{t\to\infty}\left(\psi_{max} D^{-1}\right) = 0 \quad \Rightarrow \quad \lim_{t\to\infty}\left(\delta D^{-1}\right) = 0.$$

Theorem 2.4.1. *Sei eine Folge von Faktormodellen, wie wir sie vorher beschrieben haben, gegeben und q fixiert.*

1. *Für* $\lim_{t\to\infty}(\delta D^{-1}) = 0$ *gilt* $\lim_{t\to\infty} r^2(B\tilde{B}) = 1$

2. *Für* $\lim_{t\to\infty}(\psi_{max} D^{-1}) = 0$ *gilt* $\lim_{t\to\infty} \rho^2(u, \tilde{u}) = 1$.

Da die zweite Bedingung stärker ist als die erste, folgt aus $\lim_{t\to\infty}(\psi_{max} D^{-1}) = 0$, dass $\lim_{t\to\infty} r^2(B\tilde{B}) = 1$ gilt. Aus der schwächeren Bedingung $\lim_{t\to\infty}(\delta D^{-1}) = 0$ folgt nicht $\lim_{t\to\infty} \rho^2(u, \tilde{u}) = 1$. Der Beweis lässt sich im Anhang A finden.

Die zweite, stärkere Bedingung ist oft approximativ erfüllt für großes p, der Anzahl der beobachtbaren Variablen [Schneeweiss and Mathes, 1995]. Die erste, schwächere Bedingung kann erfüllt werden, wenn die Faktoren approximativ die gleiche Varianz haben.

Auf Grund der Unbestimmtheit bezüglich Rotationen im Faktormodell, können wir B so wählen, dass $B^T B$ eine Diagonalmatrix mit absteigenden Diagonalelementen ist. Wir nehmen an, dass so eine Rotation vorgenommen wurde. Da die Eigenwerte von $B^T B$ sich bezüglich einer Rotation nicht ändern, entspricht die Diagonalmatrix $B^T B$ der zuvor eingeführten Matrix D. In diesem Fall und unter der zusätzlichen Annahme A können wir zeigen, dass nicht nur die Spaltenräume von B und $\tilde{B}$ gegeneinander konvergieren, sondern auch $\tilde{B}$ und $\bar{B} = BD^{-1/2}$ gegeneinander konvergieren bezüglich jedes Elementes. Eine ähnliche Aussage kann bezüglich u und $\tilde{u}$ getroffen werden, wenn $\tilde{u}$ durch den standardisierten Vektor $\breve{u} = \Lambda^{-1/2}\tilde{u}$ ersetzt wird. Der Vektor $\breve{u}$ hat die Kovarianzmatrix I.

Annahme A. *Es existiert eine Konstante $c > 0$ (unabhängig von t) und eine Zahl t_0 sodass für $t > t_0$ gilt*

$$d_j > (1+c)d_{j+1} \qquad j = 1, \ldots, q-1.$$

Diese Annahme sichert, dass sich die Eigenwerte von $B^T B$ von Vielfachheit Eins sind für genügend großes t. Sie sichert damit auch, dass B bis auf das Vorzeichen der Spalten eindeutig bestimmt ist bei gegebenen Ψ.

Theorem 2.4.2. *Sei eine Folge von Faktormodellen, wie wir in Theorem 2.4.1 beschrieben haben, gegeben. Wir nehmen an, dass $B^T B = D$, wobei D die Annahme A erfüllt. Sei $\bar{B} = BD^{-1/2}$ und $\breve{u} = \Lambda^{-1/2}\tilde{u}$.*

1. *Für* $\lim_{t\to\infty}(\delta D^{-1}) = 0$ *gilt* $\lim_{t\to\infty}\|\bar{B} - \tilde{B}S\| = 0$

2. *Für* $\lim_{t\to\infty}(\psi_{max} D^{-1}) = 0$ *gilt* $\lim_{t\to\infty} \mathbb{E}\|u - S\breve{u}\|^2 = 0,$

wobei $S = diag(s_1, \ldots, s_q), s_i = \pm 1$ eine passend gewählte Vorzeichenmatrix in beiden Fälle ist und die Norm $\|A\|$ der Matrix A definiert ist durch $\|A\|^2 = tr(A^T A)$.

Anhang A

Beweise

Wir geben zuerst eine Reihe hilfreicher Lemma an.

Lemma A.1.3. *Die folgenden Ungleichungen gelten:*

$$\psi_{min}I \leq \quad \Psi \quad \leq \psi_{max}I \tag{A.1}$$

$$D + \psi_{min}I \leq \quad \Lambda \quad \leq D + \psi_{max}I. \tag{A.2}$$

Beweis. Die Gleichung A.1 folgt aus der Definition von ψ_{max} und ψ_{min}. Für Gleichung A.2 folgt aus Gleichung A.1 und $\Omega = BB^T + \Psi$

$$BB^T + \psi_{min}I \leq \Omega \leq BB^T + \psi_{max}I.$$

Aus einem Theorem über das monotone Verhalten von Eigenwerten aus [Magnus and Neudecker, 1988], erhalten wir für die Eigenwerte der drei Matrizen die Ungleichung

$$d_i + \psi_{min}I \leq \lambda_i \leq d_i + \psi_{max}I.$$

□

Lemma A.1.4. *Wenn* $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$, *dann folgt* $\lim_{t\to\infty}(\Lambda^{-1}D) = I$.

Beweis. Ungleichung A.2 kann neu angeschrieben werden als

$$I + \psi_{min}D^{-1} \leq \Lambda D^{-1} \leq I + \psi_{max}D^{-1}.$$

Mit $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$ erhalten wir $\lim_{t\to\infty}(\psi_{min}D^{-1}) = 0$ und weiters $\lim_{t\to\infty}(\Lambda^{-1}D) = I$. □

Beweis von Theorem 2.4.1. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $B^TB = D$ gilt. Wir sind vorher darauf eingegangen, dass die Korrelation invariant bezüglich Rotationen ist.

1. Wir nehmen an, dass $\lim_{t\to\infty}(\delta D^{-1}) = 0$ gilt. Wir beweisen nun, dass daraus $\lim_{t\to\infty} r^2(B, \tilde{B}) = 1$ folgt.

   Wir müssen zeigen, dass $\lim_{t\to\infty} R = I$ mit $R = \tilde{B}^TBD^{-1}B^T\tilde{B}$ gilt. Wie wir zuvor angemerkt haben, gilt $0 \leq R \leq I$. Wir berechnen

   $$\det R = \det\left(\tilde{B}^TBD^{-1}B^T\tilde{B}\right) / \det D. \tag{A.3}$$

   Aus $\Omega = BB^T + \Psi$ und der Gleichung 2.5 erhalten wir

   $$\tilde{B}^TBD^{-1}B^T\tilde{B} = \Lambda - \tilde{B}^T\Psi\tilde{B}.$$

   Durch Gleichung A.1 und A.2 kommen wir zu

   $$\Lambda - \tilde{B}^T\Psi\tilde{B} \geq D - \delta I.$$

   Für kleines δ gilt $\det R \geq \det(D - \delta I)/\det D = \det(I - \delta D^{-1})$ und damit weiters $\det R = 1$, was wiederum equivalent ist zu $\lim_{t\to\infty} R = I$ und zu $\lim_{t\to\infty} r^2(B, \tilde{B}) = 1$.

2. Wir nehmen an, dass $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$ gilt. Wir möchten zeigen, dass $\lim_{t\to\infty} P = I$. Durch Gleichung 2.10 und 2.11 gilt

$$\det P = \det R \det \left(D\Lambda^{-1}\right).$$

Aus $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$ folgt $\lim_{t\to\infty}(\delta D^{-1}) = 0$ und damit $\lim_{t\to\infty} \det R = 1$. Auf Grund von Lemma A.1.4 gilt $\lim_{t\to\infty} \det(D\Lambda^{-1}) = 1$. Es gilt daher $\lim_{t\to\infty} \det P = 1$ und damit $\lim_{t\to\infty} P = I$.

□

Beweis von Theorem 2.4.2. 1. Wir nehmen an, dass $\lim_{t\to\infty}(\delta D^{-1}) = 0$ gilt. Zu zeigen ist, dass $\lim_{t\to\infty}\|\bar{B} - \tilde{B}S\| = 0$.

Sei $Q = \bar{B}^T\tilde{B}$. Wir möchten zuerst zeigen, dass $\lim_{t\to\infty}(QS) = I$, wobei S eine Vorzeichenmatrix ist, die so gewählt wurde, dass die Diagonalelemente von QS alle nichtnegativ sind. Aus den Gleichungen 2.2 und 2.5 erhalten wir

$$\left(\bar{B}D\bar{B}^T + \Psi\right)\tilde{B} = \tilde{B}\Lambda.$$

Durch Linksmultiplikation mit $\bar{B}^T$ und Rechtsmultiplikation mit D^{-1} erhalten wir

$$\begin{aligned}
\bar{B}^T\left(\bar{B}D\bar{B}^T + \Psi\right)\tilde{B}D^{-1} &= \bar{B}^T\tilde{B}\Lambda D^{-1} \\
\bar{B}^T\bar{B}D\bar{B}^T\tilde{B}D^{-1} + \bar{B}^T\Psi\tilde{B}D^{-1} &= \bar{B}^T\tilde{B}\Lambda D^{-1} \\
DQD^{-1} + \bar{B}^T\Psi\tilde{B}D^{-1} &= Q\Lambda D^{-1}, \qquad \text{(A.4)}
\end{aligned}$$

da gilt

$$\bar{B}^T\bar{B} = \left(BD^{-1/2}\right)^T BD^{-1/2} = D^{-1/2}B^TBD^{-1/2} = D^{-1/2}DD^{-1/2} = I.$$

Nach Neuordnung der Terme aus Gleichung A.4 erhalten wir

$$\begin{aligned} DQD^{-1} &= Q\Lambda D^{-1} - \bar{B}^T \Psi \tilde{B} D^{-1} \\ DQD^{-1} - Q &= Q\left(\Lambda D^{-1} - I - \psi_{min} D^{-1}\right) - \bar{B}^T \left(\Psi - \psi_{min} I\right) \tilde{B} D^{-1} \\ &= G - H, \end{aligned}$$

wobei G den ersten und H den zweiten Term auf der rechten Seite angibt. Wir zeigen, dass H und G gegen Null gehen. Auf Grund der Ungleichung A.1 gilt

$$0 \leq \Psi - \psi_{min} I \leq \delta I.$$

Es sei angemerkt, dass $\bar{B}\bar{B}^T$ idempotent ist. Es gilt die Folge von Ungleichungen:

$$\begin{aligned} H^T H &= D^{-1} \tilde{B}^T \left(\Psi - \psi_{min} I\right) \bar{B}\bar{B}^T \left(\Psi - \psi_{min} I\right) \tilde{B} D^{-1} \\ &\leq D^{-1} \tilde{B}^T \left(\Psi - \psi_{min} I\right)^2 \tilde{B} D^{-1} \\ &\leq \delta^2 D^{-2}, \end{aligned}$$

wobei gilt $\tilde{B}^T \tilde{B} = I$. Damit gilt $\lim_{t\to\infty} H = 0$. Bezüglich des Termes G gilt auf Grund der Ungleichung A.2

$$0 \leq \Lambda D^{-1} - I - \psi_{min} D^{-1} \leq \delta D^{-1}. \tag{A.5}$$

Wir merken an, dass $Q^T Q = R \leq I$. Damit gilt

$$\begin{aligned} G^T G &= \left(\Lambda D^{-1} - I - \psi_{min} D^{-1}\right) Q^T Q \left(\Lambda D^{-1} - I - \psi_{min} D^{-1}\right) \\ &\leq \delta^2 D^{-2}, \end{aligned}$$

und somit $\lim_{t\to\infty} G = 0$. Es folgt, dass

$$\lim_{t\to\infty} \left(DQD^{-1} - Q\right) = 0. \tag{A.6}$$

Für ein beliebiges Element q_{ij} von Q gilt

$$\lim_{t\to\infty}\left[\left(d_i d_j^{-1}-1\right)q_{ij}\right]=0.$$

Aus der Annahme A folgt

$$\lim_{t\to\infty} q_{ij}=0 \qquad \forall i\neq j. \tag{A.7}$$

Auf Grund von $\lim_{t\to\infty}(Q^TQ)=\lim_{t\to\infty}R=I$ (siehe Beweis von Theorem 2.4.1) und $\lim_{t\to\infty}\sum_{j=1}^{q}q_{ij}^2=1$ folgern wir

$$\lim_{t\to\infty} q_{ii}^2=1$$

und weiters

$$\lim_{t\to\infty}(s_i q_{ii})=1. \tag{A.8}$$

Aus den Gleichung A.7 und A.8 erhalten wir

$$\lim_{t\to\infty}(QS)=I. \tag{A.9}$$

Es gilt

$$\begin{aligned}\|\bar{B}-\tilde{B}S\|^2 &= tr\left(\bar{B}^T\bar{B}+S\tilde{B}^T\tilde{B}S-\bar{B}^T\tilde{B}S-S\tilde{B}^T\bar{B}\right)\\ &= 2tr\left(I-QS\right).\end{aligned}$$

Auf Grund von Gleichung A.9 konvergiert $\|\bar{B}-\tilde{B}S\|^2$ gegen Null.

2. Der Beweis ist in [Schneeweiss, 1995] angeführt. Der Vollständigkeit halber werden wir den Beweis kurz skizzieren.

   Zuerst berechnen wir $\mathbb{E}\|u-S\breve{u}\|^2$ und erhalten die Gleichungsfolge

$$\begin{aligned}\mathbb{E}\|u-S\breve{u}\|^2 &= \mathbb{E}\left\{tr\left[(u-S\breve{u})^T(u-S\breve{u})\right]\right\}\\ &= \mathbb{E}\left[tr\left(u^Tu\right)+tr\left(\breve{u}^TS^TS\breve{u}\right)-tr\left(CS\right)-tr\left(SC^T\right)\right].\end{aligned}$$

Es sei angemerkt, dass $\mathbb{E}(uu^T) = I$ gilt. Durch Gleichung 2.12, der Definition von $\breve{u}$ und zyklisches Vertauschen erhalten wir

$$\begin{aligned}
\mathbb{E}\left[tr\left(\breve{u}^T S^T S \breve{u}\right)\right] &= \mathbb{E}\left[tr\left(\Lambda^{-1/2}\tilde{u}\tilde{u}^T\Lambda^{-1/2}\right)\right] \\
&= tr\left[\mathbb{E}\left(\Lambda^{-1/2}\tilde{u}\tilde{u}^T\Lambda^{-1/2}\right)\right] \\
&= tr\left[\mathbb{E}\left(\Lambda^{-1/2}\Lambda\Lambda^{-1/2}\right)\right] \\
&= tr\left(I\right) \\
&= r.
\end{aligned}$$

Wir können nun vereinfachen und erhalten

$$\mathbb{E}\|u - S\breve{u}\|^2 = 2r - tr(CS + SC^T), \tag{A.10}$$

wobei $C = \mathbb{E}(u\breve{u}^T)$. Für C gilt:

$$\begin{aligned}
C &= \mathbb{E}(u\breve{u}^T) \\
&= \mathbb{E}(u\tilde{u}^T\Lambda^{-1/2}) \\
&= B^T\tilde{B}\Lambda^{-1/2} \\
&= D^{1/2}\bar{B}^T\tilde{B}\Lambda^{-1/2} \\
&= D^{1/2}Q\Lambda^{-1/2}.
\end{aligned}$$

Durch die Annahme, dass $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$ gilt, wissen wir das auf Grund von Lemma A.1.4 $\lim_{t\to\infty}(\Lambda^{-1}D) = I$ gilt. Es gilt $\lim_{t\to\infty}(QS) = I$ (Ungleichung A.9). Da Q eine Diagonalmatrix können wir die Diagonalmatrizen Q und

$\Lambda^{-1/2}$ vertauschen und erhalten

$$\begin{aligned}
\mathbb{E}\left[tr\left(CS\right)\right] &= \mathbb{E}\left[tr\left(D^{1/2}Q\Lambda^{-1/2}S\right)\right] \\
&= \mathbb{E}\left[tr\left(D^{1/2}\Lambda^{-1/2}QS\right)\right] \\
&= tr\left[\mathbb{E}\left(D^{1/2}\Lambda^{-1/2}QS\right)\right] \\
&= tr\left(I\right) \\
&= r.
\end{aligned}$$

Unter der Annahme, dass $\lim_{t\to\infty}(\psi_{max}D^{-1}) = 0$ gilt, erhalten wir schlussendlich

$$\mathbb{E}\|u - S\check{u}\|^2 = 0.$$

$\square$

Literaturverzeichnis

[Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19(6):716–723.

[Anderson, 1963] Anderson, T. (1963). Asymptotic theory for principal component analysis. *Ann. J. Math. Stat.*, 34:122–148.

[Anderson, 1984] Anderson, T. (1984). Estimating linear statistical relationships. *Ann. Statist.*, 12(1):1–45.

[Anderson, 2003] Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 3 edition.

[Anderson and Rubin, 1956] Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, V:111–150.

[Bai and Silverstein, 2004] Bai, Z. and Silverstein, J. (2004). Clt for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Prob.*, 32(1A):553–605.

[Bai and Yin, 1993] Bai, Z. and Yin, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21:1275–1294.

[Bai, 1999] Bai, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statististica Sinica*, 9:611–677.

[Baik and Silverstein, 2006] Baik, J. and Silverstein, J. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivar. Anal.*, 97:1382–1408.

[Beckmann and Smith, 2004] Beckmann, C. and Smith, S. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 23(2):137–152.

[Bekker and Kano, 1990] Bekker, P. A. and Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, (25):67–74.

[Chamberlain and Rothschild, 1983] Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, (51):1281–1304.

[Cox, 2008] Cox, R. (2008). Afni national institute of mental health (nimh), bethesda md.

[Donoho and Johnstone, 1994] Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

[Donoho and Johnstone, 1995] Donoho, D. and Johnstone, I. (1995). Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

[Eastment and Krzanowski, 1982] Eastment, H. and Krzanowski, W. (1982). Cross-validatory choice of the number of components from a principal component analysis. *Technometrics*, 24(1):73–77.

[Everson and Roberts, 2000] Everson, R. and Roberts, S. (2000). Inferring the eigenvalues of covariance matrices from limited, noisy data. *IEEE Trans. Signal Process.*, 48(7):2083–2091.

[Friston et al., 1993] Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. (1993). Functional connectivity: The principal component analysis of large data sets. *J. Cereb. Blood Flow Metab.*, 13:5–14.

[Geman, 1980] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.*, 8:256–261.

[Hansen and Larsen, 1996] Hansen, L. and Larsen, J. (1996). Unsupervised learning and generalization. *Proc. IEEE Int. Conf. Neural Networks (ICNN)*, 1:25–30.

[Hansen et al., 1999] Hansen, L., Larsen, J., Nielsen, F., Strother, S., Rostrup, E., Savoy, R., Lange, N., Sidtis, J., C.Svarer, and Paulson, O. (1999). Generalizable Patterns in Neuroimaging: How Many Principal Components? *NeuroImage*, 9(5):534–544.

[Hotelling, 1935] Hotelling, H. (1935). The most predictable criterion. *J. Educ. Psychol.*, 26:139–142.

[Hudson, 1998] Hudson, M. (1998). Maximum likelihood restoration and choice of smoothing parameter in deconvolution of image data subject to poisson noise. *Comput. Stat. Data Anal.*, 26(4):393–410.

[Hurvich and Tsai, 1993] Hurvich, C. and Tsai, C. (1993). A corrected akaike information criterion for vector autoregressive model selection. *J. Time Series Anal*, 14:276–279.

[Jackson, 1991] Jackson, J. (1991). *A Users Guide to Principal Components.* New York: Wiley.

[Johnstone, 2001] Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Stat.*, 29(2):295–327.

[Jolliffe, 2002] Jolliffe, I. (2002). *Principal Component Analysis.* New York:Springer, 2 edition.

[Krim and Viberg, 1996] Krim, H. and Viberg, M. (1996). Two decades of array signal processing research: The parametric approach. *IEEE Signal Process. Mag.*, 13(4):67–94.

[Kshirsagar, 1972] Kshirsagar, A. (1972). *Multivariate Analysis.* New York: Marcel Dekker.

[Lawley, 1953] Lawley, D. (1953). A modified method of estimation in factor analysis and some large sample results. *Analysis, Proc. Uppsala Symp. Psychological Factor Factor Analysis*, pages 35–42.

[Linhart and Zucchini, 1986] Linhart, H. and Zucchini, W. (1986). *Model Selection.* New York: Wiley.

[Loader, 1999] Loader, C. (1999). *Local Regression and Likelihood.* New York: Springer.

[Magnus and Neudecker, 1999] Magnus, J. and Neudecker, H. (1999). *Matrix Differential Calculus.* New York: Wiley.

[Magnus and Neudecker, 1988] Magnus, J. R. and Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons, New York.

[Mallows, 1973] Mallows, C. (1973). Some comments on cp. *Technometrics*, 15(4):661–675.

[Marcenko and Pastur, 1967] Marcenko, V. and Pastur, L. (1967). Distribution of eigenvalues of some sets of random matrices. *Math. USSR-Sb.*, 1:Math. USSR–Sb.

[Marcus, 1956] Marcus, M. (1956). An eigenvalue inequality for the product of normal matrices. *Amer. Math. Month.*, 63(3):173–174.

[Mardia et al., 1979] Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis.* London, U.K.: Academic.

[Minka, 2000] Minka, T. (2000). Automatic choice of dimensionality for pca. *Proc. Ad- vances Neural Information Processing Systems (NIPSâ00)*, pages 598–604.

[Ng and Solo, 1997] Ng, L. and Solo, V. (1997). A data-driven method for choosing smoothing parameters in optical flow problem. *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 3:360–363.

[Paul, 2004] Paul, D. (2004). Asymptotics of the leading sample eigenvalues for a spiked covariance model. Technical report, Statistics Dept., Stanford Univ.,, Stanford, CA,.

[Pearson, 1901] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 6(2):559–572.

[Preisendorfer and Mobley, 1988] Preisendorfer, R. and Mobley, C. (1988). *Principal component analysis in meteorology and oceanography.* Developments in atmospheric science. Elsevier.

[Ramsay and Silverman, 1997] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. New York: Springer, 1 edition.

[Rissanen, 1978] Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471.

[Schneeweiss, 1990] Schneeweiss, H. (1990). *Modelle mit latenten variablen. LISREL versus PLS.*, volume Neuere Entwicklungen in der angewandten Ökonometrie. Heidelberg: Physica.

[Schneeweiss, 1995] Schneeweiss, H. (1995). *Factors and principal components: Their approach to each other when the factor loadings are orthogonal. In H. Rinne. B. Rüger, and H.Strecker*. Grundlagen der Statistik und ihre Anwendungen. Festschrift für Kurtz Weichselberger. Heidelberg: Physica.

[Schneeweiss, 1997] Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research*, 32(4):375–401.

[Schneeweiss and Mathes, 1995] Schneeweiss, H. and Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, (55):105–124.

[Scholkopf et al., 1998] Scholkopf, B., Smola, A., and Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319.

[Schwartz, 1978] Schwartz, G. (1978). Estimating the dimension of a model. *Ann. Stat.*, 6(2):461–464.

[Seber, 1984] Seber, G. (1984). *Multivariate Observations*. New York: Wiley.

[Seghouane and Cichocki, 2007] Seghouane, A. and Cichocki, A. (2007). Bayesian estimation of the number of principal components. *Signal Process.*, 87:562–568.

[Shi and Solo, 2004] Shi, M. and Solo, V. (2004). Empirical choice of smoothing parameters in robust optical flow estimation. *Proc. IEEE Int. Conf. Signal Processing (ICASSP)*, (3):349–352.

[Silverstein, 1986] Silverstein, J. (1986). Eigenvalues and eigenvectors of large dimensional sample covariance matrices. *Contemp. Math.*, 50:153–159.

[Snook and Gorsuch, 1989] Snook, S. C. and Gorsuch, R. L. (1989). Principal component analysis versus common factor analysis: A monte carlo study. *Psychological Bulletin*, (106):148–154.

[Solo, 1996] Solo, V. (1996). A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems. *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 3:89–92.

[Solo, 1999] Solo, V. (1999). Selection of regularization parameters for total variation denoising. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 3:1653–1655.

[Solo, 2001] Solo, V. (2001). Automatic stopping criterion for anisotropic diffusion. *roc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 6:3441–3444.

[Solo, 2005] Solo, V. (2005). Selection of tuning parameter for support vector machines. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (5):235–240.

[Stein, 1981] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Stat.*, 9(6):1135–1151.

[Stoica and Selen, 2004] Stoica, P. and Selen, Y. (2004). Model-order selection. a review of information criterion rules. *IEEE Signal Process. Mag.*, 21(4):36–47.

[Theobald, 1975] Theobald, C. (1975). An inequality with application to multivariate analysis. *Biometrika*, 62(2):Biometrika.

[Thomson, 1934] Thomson, G. H. (1934). Hotelling's method modified to give spearman's "g". *Journal of Educational Psychology*, (25):366–374.

[Tipping and Bishop, 1999] Tipping, M. and Bishop, C. (1999). Probabilistic principal component analysis. *J. Royal Stat. Soc.*, Series B, 61(3):611–622.

[Trzeinka, 1986] Trzeinka, C. (1986). On the number of factors in the arbitrage pricing model. *Journal of Finance*, (41):347–368.

[Ulfarsson and Solo, 2006a] Ulfarsson, M. and Solo, V. (2006a). Smooth principal component analysis with application to functional magnetic resonance imaging. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2:II–993–II–996.

[Ulfarsson and Solo, 2006b] Ulfarsson, M. and Solo, V. (2006b). Spatially local and temporally smooth pca for fmri. *Proc. IEEE Int. Conf. Image Processing (ICIPâ06)*, pages 2853–2856.

[Ulfarsson and Solo, 2008a] Ulfarsson, M. O. and Solo, V. (2008a). Dimension estimation in noisy pca with sure and random matrix theory. *IEEE Transactions on Signal Processing*, 56(12):5804–5816.

[Ulfarsson and Solo, 2008b] Ulfarsson, M. O. and Solo, V. (2008b). Rank selection in noist pca with sure and random matrix theory. In *ICASSP*, pages 3317–3320.

[Velicer and Jackson, 1990] Velicer, W. F. and Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, (25):1–28.

[Velicer et al., 1982] Velicer, W. F., Peakock, A. C., and Jackson, D. N. (1982). A comparison of component and factor patterns: A monte carlo approach. *Multivariate Behavioral Research*, (17):371–388.

[Wax and Kailath, 1985] Wax, M. and Kailath, T. (1985). Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Process.Proc.*, ASSP-33(2):387–392.

[Whittle, 1953] Whittle, P. (1953). On principal components and least square methods of faktor analysis. *Skand. Aktuarietidskr.*, (35):369–372.

[Wold, 1978] Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components model. *Technometrics*, 20(4):397–405.

Lebenslauf

Ausbildung

Name	Alexander Juschitz
Adresse	Afritschgasse 61, 1220 Wien
Geboren	28.04.1984 in Wien, Österreich

Ausbildung

2004–2011	Studium an der **Technischen Universität Wien**, in **Technischer Mathematik** mit Spezialisierung in **Wirtschaftsmathematik**
2003–2010	Studium an der **Technischen Universität Wien**, in **Wirtschaftsingenieurwesen-Maschinenbau**, mit Spezialisierung in **Wettbewerb und Unternehmensführung** Abschluss mit Auszeichnung
2006–2008	Studium an der **Ecole Centrale Paris**, einer führenden Ingenieurshochschule Frankreichs, im Rahmen des Austauschprogramms „TI-ME“(„Top Industrial Managers for Europe“)
2002–2003	Absolvierung des Grundwehrdienstes
2002	Matura am Gymnasium **Theresianische Akademie**

Beruflicher Werdegang und Praktika

2009	**Cornell University** Visiting Non-Degree Student an der School of Civil and Environmental Engineering, Stipendium von der Marshall Plan Stiftung	Ithaca, NY USA
2007	**Bauman Universität** Lehrwerkstätte in Verbindung mit einem Sprachkurs	Moskau, Russland