

Seminar Quantitatives Risikomanagement

Prof: Hanspeter Schmidli
Betreuung: Julia Eisenberg
27.10.2009 Shen, Yi

Multivariate Modelle II

3.3 Sphärische und Elliptische Verteilung

3.3.1 Sphärische Verteilung

Definition 3.18: Ein Zufallsvektor $X=(X_1, X_2, \dots, X_d)'$ hat eine sphärische Verteilung, wenn für jede orthogonale Matrix $U \in \mathbb{R}^{d \times d}$ die Gleichung $UX \stackrel{D}{=} X$ gilt.

Satz 3.19: Die folgenden Aussagen sind äquivalent.

- (1) Der Zufallsvektor $X \in \mathbb{R}^d$ hat eine sphärische Verteilung.
- (2) Es existiert eine Funktion ψ , sodass die charakteristische Funktion von X folgendermaßen gegeben wird:

$$\phi_X(t) = E(e^{it'X}) = \psi(t't) = \psi(t_1^2 + \dots + t_d^2) \quad \text{für alle } t \in \mathbb{R}^d \quad (1)$$

$$(3) \text{ Für jeden Vektor } a \in \mathbb{R}^d \text{ gilt } a'X \stackrel{D}{=} \|a\| X_1, \quad (2)$$

wobei $\|a\|^2 = a'a = a_1^2 + \dots + a_d^2$.

Bew: (1) \Rightarrow (2) Wenn X eine sphärische Verteilung hat, dann für jeder orthogonale Matrix U gilt:

$$\phi_X(t) = \phi_{UX}(t) = E(e^{it'UX}) = \phi_X(Ut) \Rightarrow \phi_X(t) = \psi(t't), t't \geq 0$$

$$(2) \Rightarrow (3) \text{ Zuerst beobachten wir, dass } \phi_{X_1}(t) = E(e^{itX_1}) = \phi_X(te_1) = \psi(t^2),$$

wobei e_1 die Einheitsvektoren in \mathbb{R}^d .

Für jedes $a \in \mathbb{R}^d$ gilt:

$$\phi_{a'X}(t) = \phi_X(ta) = \psi(t^2 a'a) = \psi(t^2 \|a\|^2) = \phi_{X_1}(t\|a\|) = \phi_{\|a\|X_1}(t)$$

(3) \Rightarrow (1) Für jede orthogonale Matrix U gilt:

$$\phi_{UX}(t) = E\left(e^{i(Ut)'X}\right) = E\left(e^{i\|U't\|X_1}\right) = E\left(e^{i\|t\|X_1}\right) = E\left(e^{it'X}\right) = \phi_X(t)$$

Notation einer sphärischen Verteilung: $X \sim S_d(\psi)$

Bsp 3.20 (Multivariate Normalverteilung)

Ein Zufallsvektor X mit standard unkorrelierte Normalverteilung

$N_d(0, I_d)$ hat sphärische Verteilung.

Die charakteristische Funktion ist

$$\phi_X(t) = E(\exp(it'X)) = \exp\left\{it'0 - \frac{1}{2}t'I_d t\right\} = \exp\left(-\frac{1}{2}t't\right), \quad I_d \text{ Einheitsmatrix}$$

Nach Satz 3.19 (2) ist die charakteristische Funktion einer sphärischen Verteilung durch eine Funktion ψ , wie oben, beschrieben. Deshalb nennt man ψ charakteristische Generator der sphärischen Verteilung.

Bsp 3.21 (Varianz-gemischte Normalverteilung)

Ein Zufallsvektor X mit standardisiert unkorrelierte Varianz-gemischte

Normalverteilung $M_d(0, I_d, \hat{H})$ hat sphärische Verteilung.

Die charakteristische Funktion ist

$$\phi_X(t) = E(\exp(it'X)) = \exp(it'0) \hat{H}\left(\frac{1}{2}t'I_d t\right) = \hat{H}\left(\frac{1}{2}t't\right)$$

Nach Satz 3.19 (2) $X \sim S_d(\psi)$ mit charakterischem Generator

$$\psi(t) = \hat{H}\left(\frac{1}{2}t\right).$$

Somit $X \sim M_d(0, I_d, \hat{H}(\cdot))$ und $X \sim S_d\left(\hat{H}\left(\frac{1}{2}\cdot\right)\right)$ sind verschiedene

Darstellungen einer gemischte Verteilung.

Satz 3.22: X lässt sich als $X \stackrel{D}{=} RS$ (3) repräsentieren, wobei der Zufallsvektor $S \in \mathbb{R}^d$ gleichverteilt auf der Einheitskugel S^{d-1} , $S^{d-1} = \{s \in \mathbb{R}^d : \|s\| = 1\}$ ist, und $R \geq 0$ eine von S unabhängige ZV ist.

Bew:

1.) Beh: Wenn der Zufallsvektor S gleichmäßig verteilt auf der Einheitskugel und $R \geq 0$ ist, dann hat RS eine sphärische Verteilung.

Bew: Die charakteristische Funktion ist

$$\phi_{RS}(t) = E(e^{iRt'S}) = E(E(e^{iRt'S} | R)) = E(\phi_S(Rt)).$$

Da S selbst sphärisch verteilt ist, ist Ω_d der charakteristische Generator.

Nach Satz 3.19 (2) erhalten wir,

$$\phi_{RS}(t) = E(\Omega_d(R^2 t't)) = \int \Omega_d(r^2 t't) dF(r) \quad (4)$$

wobei F die Verteilungsfunktion von r ist.

\Rightarrow Es ex. eine Funktion von $t't$, nach Satz 3.19(2) hat RS sphärische Verteilung.

2.) Beh: Der Zufallsvektor X hat sphärische Verteilung, dann hat X die Darstellung (3).

Bew: Für jedes $s \in S^{d-1}$ muss den charakteristischen Generator ψ von X gelten $\psi(t't) = \phi_X(t) = \phi_X(\|t\|s)$.

Wir setzen den Zufallsvektor S ein.

$$\Rightarrow \psi(t't) = \int_{S^{d-1}} \phi_X(\|t\|s) dF_S(s) = \int_{S^{d-1}} E(e^{i\|t\|s'X}) dF_S(s)$$

Vertauschen der Reihenfolge der Integration.

$$\Rightarrow \psi(t't) = E\left(\Omega_d\left(\|t\|^2\|X\|^2\right)\right) = \int \Omega_d(t'tr^2) dF_{\|X\|}(r) \quad (5)$$

Durch Vergleich mit (4) können wir sehen, dass (5) die charakteristische Funktion von RS ist.

\Rightarrow Beh.

Korollar 3.23: Angenommen $X = RS \sim S_d^+(\psi)$, dann

$$\left(\|X\|, \frac{X}{\|X\|}\right)^d = (R, S) \quad (6)$$

Bew: Sei $f_1(x) = \|x\|$ und $f_2(x) = x/\|x\|$. Aus Satz 3.22 folgt

$$\left(\|X\|, \frac{X}{\|X\|}\right)^d = (f_1(x), f_2(x))^d = (f_1(RS), f_2(RS)) = (R, S).$$

Bsp 3.24: $X \sim N_d(0, I_d)$, Sei $X'X \sim x_d^2$ eine Chi-Quadrat-Verteilung mit Freiheitsgrad d, dann folgt aus Kor. 3.23 $R^2 \sim x_d^2$.

Wir können die $E(S)$ und $Cov(S)$ dadurch berechnen.

$$0 = E(X) = E(R)E(S) \Rightarrow E(S) = 0$$

$$I_d = cov(X) = E(R^2)cov(S) \Rightarrow cov(S) = I_d/d \quad (7)$$

da $E(R^2) = d$ wenn $R^2 \sim x_d^2$.

Jetzt nehmen wir an, dass X eine sphärische Varianz-gemischte

Normalverteilung $X \sim M_d(0, I_d, \hat{H})$ hat. Wir möchten die Verteilung von

$R^2 = X'X$ in diesem Fall ermitteln.

Sei $X = \sqrt{W}Y$, wobei $Y \sim N_d(0, I_d)$ und W unabhängig von Y ist.

Es folgt $R^2 = W\tilde{R}^2$, wobei $\tilde{R}^2 \sim x_d^2$ und W unabhängig von \tilde{R} ist.

Satz 3.25: Sei Ψ_∞ die Menge der Generator-Funktionen, die für bel. d-dim. gilt. Dann gilt $X \sim S_d(\psi)$ mit $\psi \in \Psi_\infty \Leftrightarrow X = \sqrt{W}Z$, wobei $Z \sim N_d(0, I_d)$ unabhängig von $W \geq 0$ ist.

Wenn eine sphärische Verteilung die Dichte f hat, dann durch die

Verwendung der Umkehrformel
$$f(x) = \frac{1}{(2\pi)^d} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-it'x} \phi_X(t) dt_1 \dots dt_d,$$

ist es leicht von Satz 3.19 abzuleiten, dass für jede orthogonale Matrix U

$f(x) = f(Ux)$ gilt, so dass die Dichte die Form

$$f(x) = g(x'x) = g(x_1^2 + \dots + x_d^2) \quad (8) \quad \text{hat.}$$

Wobei g eine Funktion einer skalaren Variable, g wird als Dichtegenerator bezeichnet.

Offensichtlich, die gemeinsame Dichte ist konstant auf der Hypersphäre

$$\{x : x_1^2 + \dots + x_d^2 = c\} \quad \text{in } \mathbb{R}^d$$

3.3.2 Elliptische Verteilungen

Definition 3.26: Ein Zufallsvektor $X \in \mathbb{R}^d$ hat eine elliptische Verteilung

wenn $X \stackrel{D}{=} \mu + AY$, wobei $Y \sim S_k(\psi)$, $\mu \in \mathbb{R}^d$ ein konstanter Vektor ist

und $A \in \mathbb{R}^{d \times k}$ eine konstante Matrix ist.

Die charakteristische Funktion ist:

$$\phi_X(t) = E(e^{it'X}) = E(e^{it'(\mu + AY)}) = e^{it'\mu} E(e^{i(A't)'Y}) = e^{it'\mu} \psi(t'\Sigma t), \quad \text{wobei } \Sigma = AA'.$$

Notation: elliptische Verteilungen $X \sim E_d(\mu, \Sigma, \psi)$

μ heißt Positionsvektor, Σ heißt Dispersionsmatrix, ψ heißt charakteristischer Generator der elliptischen Verteilung.

Bem3.27: $X \sim E_d(\mu, \Sigma, \psi)$ ist nicht eindeutig bestimmt. μ und Σ sind bis auf eine pos. Konstante eindeutig bestimmt.

Proposition3.28: $X \sim E_d(\mu, \Sigma, \psi)$ genau dann wenn $X \stackrel{D}{=} \mu + RAS$ (9),

wobei (i) S ist ein auf der Einheitskugel $S^{k-1} = \{s \in \mathbb{R}^k : s's = 1\}$

gleichverteilter Zufallsvektor.

(ii) $R \geq 0$ ist eine von S unabhängige nicht negative Zufallsvariable.

(iii) $A \in \mathbb{R}^{d \times k}$ mit $\Sigma = AA'$.

Falls Σ positiv definiert, dann gilt folgende Relation zwischen elliptischen und sphärischen Verteilungen:

$$X \sim E_d(\mu, \Sigma, \psi) \Leftrightarrow \Sigma^{-1/2}(X - \mu) \sim S_d(\psi) \quad (10)$$

In diesem Fall, wenn der sphärischen Vektor Y den Dichtegenerator g hat, dann hat $X = \mu + \Sigma^{1/2}Y$ die Dichte

$$f(x) = \frac{1}{|\Sigma|^{1/2}} g\left((x - \mu)' \Sigma^{-1}(x - \mu)\right).$$

Die gemeinsame Dichte ist immer Konstant auf den Mengen der Form

$$\left\{x : (x - \mu)' \Sigma^{-1}(x - \mu) = c\right\}, \text{ welche die Ellipsoiden in } \mathbb{R}^d \text{ beschreiben.}$$

Offensichtlich, ist die ganze Familie von multivariate Varianz-gemischte Normalverteilungen mit Parametern μ und Σ elliptisch. Aus (6) und

(10) folgt, dass für eine nicht singuläre elliptische Variate $X \sim E_d(\mu, \Sigma, \psi)$

ohne $P(X > \mu) = 0$, gilt:

$$\left(\sqrt{(x-\mu)' \Sigma^{-1} (x-\mu)}, \frac{\Sigma^{-1/2} (x-\mu)}{\sqrt{(x-\mu)' \Sigma^{-1} (x-\mu)}} \right)^d = (R, S) \quad (11)$$

wobei S gleichverteilt auf der Einheitskugel S^{d-1} ist und R eine unabhängige skalare Zufallsvariable ist.

Proposition 3.29: Sei $X \sim E_d(\mu, \Sigma, \psi)$, Σ positiv definit und $\text{Cov}(X)$ endlich. Für alle $c \geq 0$, sodass $P\left((X-\mu)' \Sigma^{-1} (X-\mu) \geq c\right) > 0$ erhalten wir

$$\rho\left(X \mid (X-\mu)' \Sigma^{-1} (X-\mu) \geq c\right) = \rho(X) \quad (12)$$

Bew: Aus (11) folgt, dass

$$X \mid (X-\mu)' \Sigma^{-1} (X-\mu) \geq c = \mu + R \Sigma^{1/2} S \mid R^2 \geq c,$$

wobei $R = \sqrt{(X-\mu)' \Sigma^{-1} (X-\mu)}$ und S unabhängig von R und gleichverteilt

auf S^{d-1} . So erhalten wir $X \mid (X-\mu)' \Sigma^{-1} (X-\mu) \geq c = \mu + \tilde{R} \Sigma^{1/2} S, \tilde{R} = R \mid R^2 \geq c.$

^{prop.3.28}
 \Rightarrow Beh.

3.3.3 Eigenschaften der elliptischen Verteilungen

Lineare Kombinationen: Sei $X \sim E_d(\mu, \Sigma, \psi)$, für $B \in \mathbb{R}^{k \times d}$ und

$b \in \mathbb{R}^d$, dann gilt: $BX + b \sim E_k(B\mu + b, B\Sigma B', \psi)$

(13)

Falls $a \in \mathbb{R}^d$ dann $a'X \sim E_1(a'\mu, a'\Sigma a, \psi)$ (14)

Randverteilungen:

Setze $X = (X_1, X_2)'$ und $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ sowie $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Es gilt dann

$X_1 \sim E_k(\mu_1, \Sigma_{11}, \psi)$ und $X_2 \sim E_{d-k}(\mu_2, \Sigma_{22}, \psi)$.

Bedingte Verteilungen: Die bedingte Verteilung von X_2 gegeben X_1 ist elliptisch verteilt, obwohl die i.a einen anderen charakteristischen Generator $\tilde{\psi}$ hat.

Nur für die Spezialfälle von der multivariaten Normalverteilung bleibt der Generator unverändert.

Quadratische Formen: Wenn $X \sim E_d(\mu, \Sigma, \psi)$ mit Σ regulär ist, dann gilt:

$$Q := (X - \mu)' \Sigma^{-1} (X - \mu) \stackrel{D}{=} R^2 \quad (15)$$

wobei R die nicht negative Zufallsvariable aus der stochastischen Darstellung in (3).

Faltung: Seien $X \sim E_d(\mu, \Sigma, \psi)$ und $Y \sim E_d(\tilde{\mu}, \Sigma, \tilde{\psi})$ zwei unabhängige Zufallsvektoren. Es gilt dann $X + Y \sim E_d(\mu + \tilde{\mu}, \Sigma, \bar{\psi})$ (16)

wobei $\bar{\psi}(u) = \psi(u)\tilde{\psi}(u)$.

Σ muss i.a dieselbe für X und Y sein.

Wenn $aX \neq Y$ (a ist konstant und X, Y sind die Dispersionsmatrizen), dann bleibt die Faltung nicht unbedingt elliptisch, obwohl die zwei Generatoren identisch sind.

3.3.4 Schätzung der Dispersion und Korrelation

Nehmen wir an, dass die Risiko-Faktor Return Daten X_1, \dots, X_n elliptisch verteilt sind.

M-Schätzer: M-Schätzer ist eine relativ alte Idee. Der Vorteil dieser Schätzer ist, dass sie sehr einfach zu implementieren sind. Seien $\hat{\mu}$ und $\hat{\Sigma}$ die Schätzer de Erwartungsvektors und der Dispersionsmatrix. Für

jede Beobachtung X_i berechnen wir

$D_i^2 = (X_i - \hat{\mu})' \hat{\Sigma}^{-1} (X_i - \hat{\mu})$. Wenn man verbesserte Schätzer ermitteln möchte, insbesondere für heavy-tailed Daten, reduziert den Einfluss der Betrachtungen mit grossen D_i durch Einsetzen von Gewichtsfunktion $w_j : \mathbb{R}^+ \rightarrow \mathbb{R}^+$. Die Schätzung der Dispersionsmatrix ist i.a eine verzerrte Schätzung von der echten Varianzmatrix.

Korrelationsschätzung via Kendall's tau:

Eine sehr einfach durchzuführende Methode für die Korrelationsschätzung basiert auf Kendall's tau.

Sei $(X_1, X_2) \sim E_2(\mu, \Sigma, \psi)$, dann $\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho)$. (17)

wobei $\rho = \sigma_{12} / (\sigma_{11}\sigma_{22})^{1/2}$ der pseudo-Korrelationskoeffizient von der elliptischen Verteilung. (siehe Kap.5)

Diese Methode schätzt die Korrelation durch Ausnutzung der Geometrie einer elliptischen Verteilung, die Varianz- und Kovarianzschätzung brauchen wir nicht.

Bsp3.30 (Effiziente Korrelationsschätzung für Heavy-tailed Daten)

Angenommen, wir berechnen die Korrelationen der Risiko-Faktor Returns auf Basis von 90 Tagen; Es scheint, dass genug Daten haben, um eine akkurate Schätzung der realen Korrelation zu ermitteln, unter der Annahme der identisch verteilten Daten im Betrachtungszeitraum.

Die Abbildung 3.5 zeigt die Ergebnisse eines simulationsexperiments

Wir haben 3000 bivariate Stichproben von iid Daten aus t-Verteilung mit

3 Freiheitsgraden und Korrelation $\rho = 0.5$ generiert; Dies ist eine heavy-tailed elliptische Verteilung. Die Verteilung von den Werten der Standard-Korrelationskoeffizient ist nicht besonders eng um den echten Wert konzentriert und hat ein paar sehr schlechte Schätzer für einige Stichproben. Auf der anderen Seite, die Schätzungen von dem Kendall's tau-Transform-Methode kommen viel näher an den echten Wert.

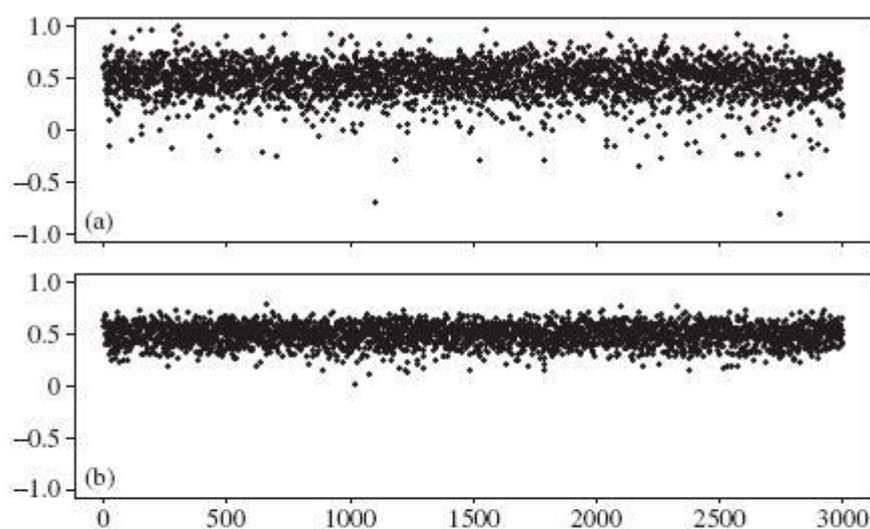


Figure 3.5. For 3000 independent samples of size 90 from a bivariate t distribution with three degrees of freedom and linear correlation 0.5: (a) the standard (Pearson) estimator of correlation; (b) the Kendall's tau transform estimator. See Example 3.31 for commentary.

3.3.5 Testen für elliptische Symmetrie

Das Problem dieses Abschnitts ist zu prüfen, ob eine Stichprobe von den identisch verteilten Datenvektoren X_1, \dots, X_n bei einigen μ , Σ und Generator ψ eine elliptische Verteilung hat. In der Realität können wir generell nicht davon ausgehen, dass die Beobachtungen iid Zufallsvektoren sind, aber wir können annehmen, dass sie zumindest die gleiche Verteilung haben.

Auch wenn wir annehmen würden, dass X_1, \dots, X_n unabh. sind, die Tatsache, dass wird Abhängigkeiten in die unten beschriebene Methoden einbringen.

Stabile Korrelationenschätzung : Eine explorative Methode.

Wir können versuchen, $\rho(X|h(X) \geq c)$, $h(x) = (x - \hat{\mu})' \hat{\Sigma}^{-1} (x - \hat{\mu})$ für verschiedene Werte von $c \geq 0$ zuschätzen. Wir erwarten, dass für elliptisch verteilte Daten die Schätzungen grob stabil über eine Reihe von unterschiedlichen c bleiben.

In Abbildung 3.6 geben wir ein Beispiel, die beiden mit simulierten und realen Daten. Bei beiden zeigen keine starke Abweichung von der Annahme von stabilen Korrelationen. Die Methode ist explorativ, und erlaubt uns nicht zu einem form. Ergebnis zu kommen.

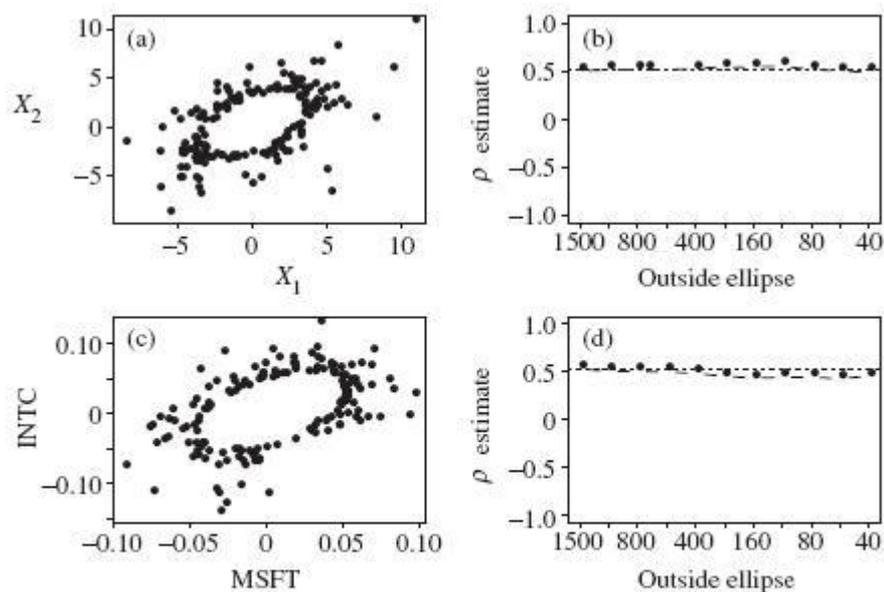


Figure 3.6. Correlations are estimated using the Kendall's tau method for points lying outside ellipses of progressively larger size (as shown in (a) and (c)). (a), (b) Two thousand t -distributed data with four degrees of freedom and $\rho = 0.5$. (c), (d) Two thousand daily log-returns on Microsoft and Intel. Dashed lines and points show estimates for an ellipse that is allowed to grow until there are only 40 points outside; dotted lines show estimates of correlation for all data.

QQplots: Wenn μ und Σ bekannt waren, dann können wir die elliptische Symmetrie testen, durch Testen der Daten $\{\Sigma^{-1/2}(X_i - \mu) : i = 1, \dots, n\}$ für sphärische Symmetrie. Ersetzen dieser Parameter durch die Schätzer $\{Y_i = \hat{\Sigma}^{-1/2}(X_i - \hat{\mu}) : i = 1, \dots, n\}$ (18), wir prüfen, ob die Daten konsistent mit einer sphärischen Verteilung sind. Dabei ignorieren wir die Wirkung des Schätzfehlers.

Lemma3.31: Sei $T(Y)$ eine Statistik, so dass

$$T(aY) = T(Y) \quad (19) \quad \text{fast sicher } \forall a > 0.$$

Dann hat $T(Y)$ die gleiche Verteilung für jeder sphärischen Vektor

$$Y \sim S_d^+(\psi).$$

Bew: Nach Satz 3.22 und (19) erhalten wir $T(Y) \stackrel{D}{=} T(RS)$ und

$T(RS) \stackrel{a.s.}{=} T(S)$. Da die Verteilung von $T(Y)$ nur von S abhängt.

\Rightarrow Beh.

Numerische Tests: Wenn wir den Fehler in der Schätzung von Position und Dispersion ignorieren, reicht es die Daten $\{Y_i = \hat{\Sigma}^{-1/2}(X_i - \hat{\mu}) : i = 1, \dots, n\}$ auf sphärische Symmetrie zu testen.

Für $i = 1, \dots, n$, seien $R_i = \|Y_i\|$ und $S_i = Y_i / \|Y_i\|$, dann sollten die

Daten S_i unter die Null-Hypothese gleichverteilt auf der Einheitskugel S^{d-1}

sein, und die Daten (R_i, S_i) sollten die Realisierungen von unabhängigen

Paaren bilden.

3.4 Dimensionreduzierungsverfahren

3.4.1 Faktormodelle

Die Grundidee von Faktormodellen ist, die zufälligkeit einer d-dim Zufallsvariable durch eine Zufallsvariable von kleinerer Dimension zu repräsentieren.

Definition 3.32 (Lineares Faktorenmodell)

Der Zufallsvektor X ist p-Faktormodell, wenn es gilt:

$$X = a + BF + \varepsilon \tag{20}$$

wobei (i) $F = (F_1, \dots, F_p)'$ ist ein Zufallsvektor von Faktoren mit $p < d$ und die Kovarianzmatrix ist pos. definit.

(ii) $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)'$ ist ein Zufallsvektor von idiosynkratischen Fehlertermen, die unkorreliert sind mit und Erwartungswerten gleich null.

(iii) $B \in \mathbb{R}^{d \times p}$ ist eine Matrix von konstanten Faktoren und $a \in \mathbb{R}^d$ ist konstant.

$$(iv) \text{cov}(F, \varepsilon) = E((F - E(F))\varepsilon') = 0$$

Die Annahme der Unkorreliertheit ist ein wesentlicher (wichtiger) Teil der Definition.

Wenn wir die Kovarianz-Matrix von F durch Ω und die der ε durch die Diagonalmatrix Υ bezeichnen, dann gilt:

$$\Sigma = \text{cov}(X) = B\Omega B' + \Upsilon \tag{21}$$

Falls das Faktormodell gilt, dann können die gemeinsamen Faktoren immer so transformiert werden, dass die Erwartungswerte gleich null und orthogonal sind. Wir setzen $F^* = \Omega^{-1/2}(F - E(F))$ und $B^* = B\Omega^{1/2}$, dann gilt $X = \mu + B^*F^* + \varepsilon$, wobei $E(X) = \mu$ und $\Sigma = B^*(B^*)' + \Upsilon$.

Umgekehrt kann man zeigen, dass wenn der Zufallsvektor X eine

Kovarianzmatrix hat, die erfüllt $\Sigma = BB' + \Upsilon$, (22)

wobei $B \in \mathbb{R}^{d \times p}$ mit $\text{Rang}(B) = p < d$ und Υ ist Diagonalmatrix, dann X eine Faktormodell Repräsentation hat.

3.4.2 Statistische Kalibrierungsstrategie

In einem *beobachtbaren Faktorenmodell* nehmen wir an, dass die geeigneten Faktoren für die fraglichen Return-Serien im Voraus erkannt wurden, und die Daten über diese Faktoren gesammelt wurden.

In einem *latenten Faktorenmodell*, die geeigneten Faktoren sind selbst von Daten X_1, \dots, X_n geschätzt.

In der *klassischen statistischen Faktorenanalyse*, wird es angenommen, dass die Daten identisch verteilt mit einer Verteilung, deren Kovarianzmatrix die Faktorstruktur $\Sigma = BB' + \Upsilon$ hat.

In Risikomanagement, das Hauptziel aller Ansätze zu den Faktormodellen ist es, die Faktordaten F_t und die Ladungsmatrizen B zu erhalten.

In Abhängigkeit davon, ob die gemeinsamen Faktoren wahrnehmbar sind oder nicht, stehen numerische Methoden zur Verfügung.

Ein Beispiel dafür ist das Ein-Faktor Modell, wobei X sich durch einen einzigen Marktindex darstellen lässt.

Bsp 3.33

Angenommen, der Zufallsvektor X beschreibt in diesem Beispiel Returns

von d verschiedenen Unternehmen, so dass die Zufallsvariable

$$Z_{(d)} = (1/d) \sum_{j=1}^d X_j$$

als Return des Portfolios für die gleiche Investition in jedem dieser Unternehmen angesehen werden kann.

$$\text{Wir berechnen, dass } Z_{(d)} = \frac{1}{d} 1'BF + \frac{1}{d} 1'\varepsilon = \sqrt{\rho}F + \frac{1}{d} \sum_{j=1}^d \varepsilon_j .$$

Das Risiko im ersten Term wird nicht von der steigende Größe d des Portfolios beeinträchtigt, wohingegen das Risiko im zweiten Term reduziert werden kann.

Angenommen, wir messen das Risiko einfach durch Berechnung von Varianzen; erhalten wir

$$\text{var}(Z_{(d)}) = \rho + \frac{1-\rho}{d} \rightarrow \rho, \quad d \rightarrow \infty,$$

dh. Der systematische Faktor hat den Hauptbeitrag zum Risiko in einem grossen Portfolio.

3.4.3 Regressionsanalyse des Faktormodells

Wir schreiben das Modell als

$$X = a + BF_t + \varepsilon_t, \quad t = 1, \dots, n \quad (23)$$

wobei X_t und F_t die Vektoren von individuellen Returns und Faktoren zum Zeitpunkt t , Parameter a und b sind geschätzt werden.

Univariate Regression

Wir betrachten das univariate Regressionsmodell

$$X_{t,j} = a_j + b_j' F_t + \varepsilon_{t,j}, \quad t = 1, \dots, n$$

$X_{1,j}, \dots, X_{n,j}$ sind die Daten von j-ten Komponent über n Zeitpunkten.

Wir unterstellen, dass \hat{a}_j und \hat{b}_j durch Standard Ordinary

Least-Squares(OLS) Methode bekommen sind. Wir wiederholen den

Prozess für X_1, \dots, X_n , dann bekommen wir den ganzen Vektor a und b.

Multivariate Regression

Wir konstruieren einige grosse Matrizen:

$$X = \underbrace{\begin{pmatrix} X'_1 \\ \vdots \\ X'_n \end{pmatrix}}_{n \times d}, \quad F = \underbrace{\begin{pmatrix} 1 & F'_1 \\ \vdots & \vdots \\ 1 & F'_n \end{pmatrix}}_{n \times (p+1)}, \quad B_2 = \underbrace{\begin{pmatrix} a' \\ B' \end{pmatrix}}_{(p+1) \times d}, \quad E = \underbrace{\begin{pmatrix} \varepsilon'_1 \\ \vdots \\ \varepsilon'_n \end{pmatrix}}_{n \times d}.$$

$$\Rightarrow X = FB_2 + E \quad (24)$$

B_2 ist die Matrix der Regressionsparametern und geschätzt werden sein.

B_2 ist mit multivariate Ordinary Least-Squares(OLS) geschätzt,

$$\text{nämlich } \hat{B}_2 = (F'F)^{-1} F'X. \quad (25)$$

Jetzt testen wir ε auf die Unkorrelation.

$\hat{E} = X - F\hat{B}_2$ ist die Residuum-Matrix und die Schätzung von ε .

Wenn \hat{E} etwa diagonal ist, erfüllt $X = FB_2 + E$ die Eigenschaft von

Faktormodell. Aus (21) folgt $\hat{\Sigma}^{(F)} = \hat{B}\hat{\Omega}\hat{B}' + \hat{Y}$, wobei

$$\hat{\Omega} = \frac{1}{n-1} \sum_{i=1}^n (F_i - \bar{F})(F_i - \bar{F})'.$$

Bsp3.34(Einzel-Index Modell für die Dow Jones 30 Returns)

Die \hat{B} Zeile steht für B . Da $E(X_j) = a_j + B_j E(F)$, bedeutet die hohen

Werte hohe erwartete Returns. Das bedeutet auch hohe Risiken. Hier

sehen wir, dass HP, Intel, Microsoft von großen Risiken sind. Es entspricht

auch der Realität.

Die r^2 Zeile ist der Relationsmaßstab zwischen einzeln-Aktie und der Marktindex.

Die nächsten 10 Zeilen von der Tabelle gibt die Kovarianzmatrix der gesamten Daten an. Man bekommt sie mit Standard-Methode.

Die mittleren 10 Zeilen ist die Kovarianzmatrix von de Faktormodell, nämlich $\hat{\Sigma}^{(F)}$. Hier sehen wir, dass die ersten 10 Zeilen und die mittleren 10 Zeilen von ähnliche Struktur sind.

Die letzten 10 Zeilen ist die Fehlermatrix E . Nach Definition ist E diagonal, aber in diesem Beispiel ist es keine vollkommene diagonale Matrix. Es gibt enge Relationen unter Intel, Microsoft, IBM und HP. In dem Faktorenmodell ist das System nicht vollkommend repräsentiert.

Table 3.6. The first line gives estimates of B for a multivariate regression model fitted to 10 Dow Jones 30 stocks where the observed common factor is the return on the Dow Jones 30 index itself. The second row gives r^2 values for a univariate regression model for each individual time series. The next 10 lines of the table give the sample correlation matrix of the data R , while the middle 10 lines give the correlation matrix implied by the factor model. The final 10 lines show the estimated correlation matrix of the residuals from the regression model, with entries less than 0.1 in absolute value being omitted. See Example 3.36 for full details.

	MO	KO	EK	HWP	INTC	MSFT	IBM	MCD	WMT	DIS
\hat{B}	0.87	1.01	0.77	1.12	1.12	1.11	1.07	0.86	1.02	1.03
r^2	0.17	0.33	0.14	0.18	0.17	0.21	0.22	0.23	0.24	0.26
MO	1.00	0.27	0.14	0.17	0.16	0.25	0.18	0.22	0.16	0.22
KO	0.27	1.00	0.17	0.22	0.21	0.25	0.18	0.36	0.33	0.32
EK	0.14	0.17	1.00	0.17	0.17	0.18	0.15	0.14	0.17	0.16
HWP	0.17	0.22	0.17	1.00	0.42	0.38	0.36	0.20	0.22	0.23
INTC	0.16	0.21	0.17	0.42	1.00	0.53	0.36	0.19	0.22	0.21
MSFT	0.25	0.25	0.18	0.38	0.53	1.00	0.33	0.22	0.28	0.26
IBM	0.18	0.18	0.15	0.36	0.36	0.33	1.00	0.20	0.20	0.20
MCD	0.22	0.36	0.14	0.20	0.19	0.22	0.20	1.00	0.26	0.26
WMT	0.16	0.33	0.17	0.22	0.22	0.28	0.20	0.26	1.00	0.28
DIS	0.22	0.32	0.16	0.23	0.21	0.26	0.20	0.26	0.28	1.00
MO	1.00	0.24	0.16	0.18	0.17	0.19	0.20	0.20	0.20	0.21
KO	0.24	1.00	0.22	0.24	0.23	0.26	0.27	0.28	0.28	0.29
EK	0.16	0.22	1.00	0.16	0.15	0.17	0.18	0.18	0.18	0.19
HWP	0.18	0.24	0.16	1.00	0.17	0.19	0.20	0.20	0.21	0.22
INTC	0.17	0.23	0.15	0.17	1.00	0.19	0.19	0.19	0.20	0.21
MSFT	0.19	0.26	0.17	0.19	0.19	1.00	0.22	0.22	0.22	0.23
IBM	0.20	0.27	0.18	0.20	0.19	0.22	1.00	0.23	0.23	0.24
MCD	0.20	0.28	0.18	0.20	0.19	0.22	0.23	1.00	0.23	0.24
WMT	0.20	0.28	0.18	0.21	0.20	0.22	0.23	0.23	1.00	0.25
DIS	0.21	0.29	0.19	0.22	0.21	0.23	0.24	0.24	0.25	1.00
MO	1.00									
KO		1.00					-0.12	0.12		
EK			1.00							
HWP				1.00	0.30	0.24	0.20			
INTC				0.30	1.00	0.43	0.20			
MSFT				0.24	0.43	1.00	0.14			
IBM		-0.12		0.20	0.20	0.14	1.00			
MCD		0.12						1.00		
WMT									1.00	
DIS										1.00

3.4.4 Hauptkomponenten Analyse (PCA)

Das Ziel von PCA ist Vorstellung von hoch korrelierte Daten durch
wenigere unkorrelierten lineare Kombination.

Sei B symmetrisch, dann gilt: $A = \Gamma \Lambda \Gamma'$ (26)

wobei (i) $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ die Diagonalmatrix der Eigenwert von A ist

und $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$.

(ii) Γ die Orthogonalmatrix ist und die Spalten von Γ standardisiert Eigenvektor von A sind.

Theoretische Hauptkomponenten

Sei X Zufallsvariable mit Erwartungsvektor μ und Kovarianzmatrix Σ , dann $Y = \Gamma'(X - \mu)$ (27)

definiert als Hauptkomponent Transformation von X ist.

Die einfache Berechnung zeigt, dass

$$E(Y) = 0, \quad \text{cov}(Y) = \Gamma'\Sigma\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda.$$

Y ist deswegen unkorreliert, die Varianz ist vom größten bis zum kleinsten sortiert.

Es gibt $\text{var}(\gamma_1'X) = \max\{\text{var}(a'X) : a'a = 1\}$, wobei γ_j die Eigenvektor von Σ zum j -te Eigenwert.

Betrachten wir, dass $\sum_{j=1}^d \text{var}(Y_j) = \sum_{j=1}^d \lambda_j = \text{spur}(\Sigma) = \sum_{j=1}^d \text{var}(X_j)$.

Die $\text{spur}(\Sigma) = \sum_{j=1}^d \text{var}(X_j)$ ist ein Maßstab für die gesamten Variation.

$\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^d \lambda_j}$ bedeutet, wie stark X durch die ersten k Komponenten repräsentiert ist.

Stichprobe der Hauptkomponenten

Der Unterschied zwischen Theoretische prinzipielle Komponent und der Stichprobe Hauptkomponent ist nur die unbekannte Kovarianzmatrix Σ durch Stichprobe Kovarianzmatrix

$S_x = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ geschätzt wird.

Hauptkomponenten als Faktoren

Angenommen, dass die ersten k Komponenten am wichtigsten für die gesamten Variation sind. Wir zerlegen Y in $(Y_1', Y_2')'$, wobei $Y_1 \in \mathbb{R}^k$ und $Y_2 \in \mathbb{R}^{d-k}$ ist; analog zerlegen wir Γ , dann gibt die Darstellung

$$X = \mu + \Gamma_1 Y_1 + \Gamma_2 Y_2 = \mu + \Gamma_1 Y_1 + \varepsilon. \quad (28)$$

$\Gamma_2 Y_2$ kann als Fehler gesehen werden.

Die Faktor sind nämlich $F_t = (\gamma_1' X, \dots, \gamma_k' X)'$.

Bsp 3.35 (PCAbasiertes Faktorenmodell für Dow Jones 30 Returns)

In Abbildung 3.9 sehen wir, dass die ersten beiden Komponenten fast 50% der gesamten Variation ausmachen.

Die (a), (b) in Abbildung 3.10 stehen jeweilig für die ersten zwei Ladungsvektoren.

In der Tabelle 3.7, \hat{B} ist die B im Faktorenmodell. In diesem Fall haben wir zwei Faktoren in Betrachtung.

Die nächsten 10 Zeilen von der Tabelle sind die Korrelationsmatrix, die aus dem obigen Modell herauskommt. Im Vergleich mit der Tabelle 3.6 scheint sie mehr Struktur aufgenommen zu haben. Die letzten 10 Zeilen sind die Residuum-matrix, sie ist wenig korreliert als die originale Matrix.

In reale Welt ist es selten, dass man eine Diagonalmatrix bekommt.

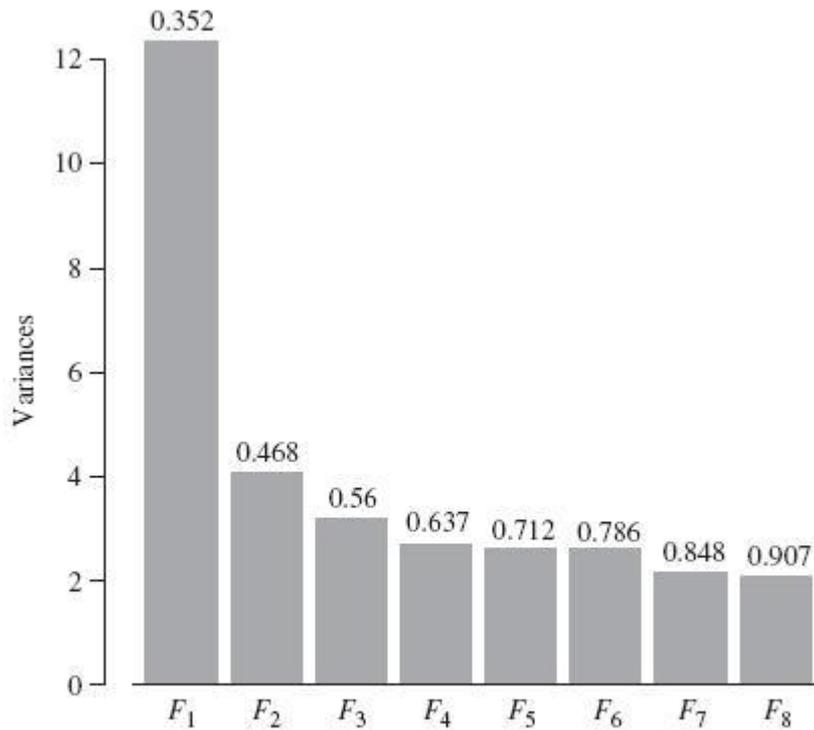


Figure 3.9. Barplot of the sample variances l_j of the first eight principal components; above each bar the cumulative proportion of the total variance explained by the components is given ($\sum_{j=1}^k l_j / \sum_{j=1}^{10} l_j, k = 1, \dots, 8$).

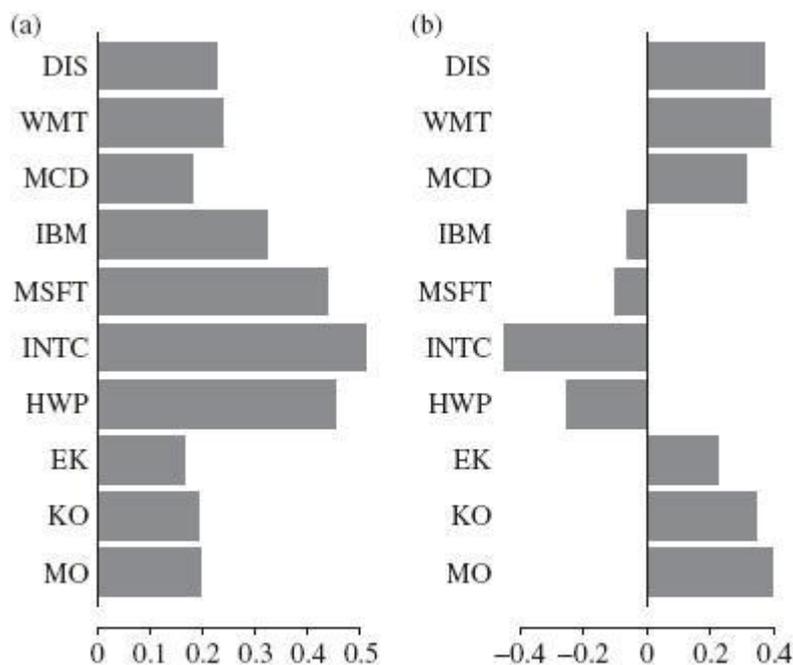


Figure 3.10. Barplot summarizing the loadings vectors g_1 and g_2 defining the first two principal components: (a) factor 1 loadings; and (b) factor 2 loadings.

Table 3.7. The first two lines give estimates of the transpose of B for a factor model fitted to 10 Dow Jones 30 stocks, where the factors are constructed from the first two sample principal components. The third row gives r^2 values for the univariate regression model for each individual time series. The next 10 lines give the correlation matrix implied by the factor model. The final 10 lines show the estimated correlation matrix of the residuals from the regression model, with entries less than 0.1 in absolute value omitted. See Example 3.38 for full details.

	MO	KO	EK	HWP	INTC	MSFT	IBM	MCD	WMT	DIS
\hat{B}'	0.20	0.19	0.16	0.45	0.51	0.44	0.32	0.18	0.24	0.22
r^2	0.39	0.34	0.23	-0.26	-0.45	-0.10	-0.07	0.31	0.39	0.37
	0.35	0.42	0.18	0.55	0.75	0.56	0.35	0.34	0.42	0.41
MO	1.00	0.39	0.25	0.17	0.13	0.25	0.20	0.35	0.38	0.38
KO	0.39	1.00	0.28	0.21	0.17	0.29	0.23	0.38	0.42	0.42
EK	0.25	0.28	1.00	0.18	0.15	0.22	0.18	0.25	0.28	0.27
HWP	0.17	0.21	0.18	1.00	0.64	0.55	0.43	0.20	0.23	0.23
INTC	0.13	0.17	0.15	0.64	1.00	0.61	0.48	0.16	0.19	0.18
MSFT	0.25	0.29	0.22	0.55	0.61	1.00	0.44	0.27	0.31	0.30
IBM	0.20	0.23	0.18	0.43	0.48	0.44	1.00	0.21	0.25	0.24
MCD	0.35	0.38	0.25	0.20	0.16	0.27	0.21	1.00	0.38	0.37
WMT	0.38	0.42	0.28	0.23	0.19	0.31	0.25	0.38	1.00	0.41
DIS	0.38	0.42	0.27	0.23	0.18	0.30	0.24	0.37	0.41	1.00
MO	1.00	-0.19	-0.15					-0.19	-0.37	-0.26
KO	-0.19	1.00	-0.15		0.11				-0.16	-0.17
EK	-0.15	-0.15	1.00					-0.15	-0.16	-0.16
HWP				1.00	-0.63	-0.37	-0.14			
INTC		0.11		-0.63	1.00	-0.24	-0.31			
MSFT				-0.37	-0.24	1.00	-0.22			
IBM				-0.14	-0.31	-0.22	1.00			
MCD	-0.19		-0.15					1.00	-0.19	-0.19
WMT	-0.37	-0.16	-0.16					-0.19	1.00	-0.23
DIS	-0.26	-0.17	-0.16					-0.19	-0.23	1.00