

# Thema: Hauptkomponentenschätzer

## Seminarvortrag „Ausgewählte Kapitel der Regressionsanalyse“

Dozent: Prof. Dr. Götz Trenkler

04.12.08

# Gliederung

- 1 Vorüberlegungen
  - Lineares Regressionsmodell
  - Motivation
- 2 Herleitung des Hauptkomponentenschätzers
  - Beispiel
- 3 Alternative Herleitung
- 4 Eigenschaften
  - Erwartungswert und Kovarianz
  - Vergleich mit dem KQ-Schätzer
  - Weitere Bemerkungen
- 5 Anzahl der Hauptkomponenten
- 6 Marquardt-Schätzer
- 7 Zusammenfassung
- 8 Literatur

# 1. Vorüberlegungen

## Lineares Regressionsmodell

$y = X\beta + \varepsilon$  mit den vier Annahmen:

- (i)  $X$  nichtstochastische  $n \times p$  Matrix mit  $p < n$ ,
- (ii)  $X$  hat vollen Spaltenrang,
- (iii)  $y$  beobachtbarer Zufallsvektor der Länge  $n$ ,
- (iv)  $\varepsilon$  nichtbeobachtbarer Zufallsvektor der Länge  $n$ , so dass  $E(\varepsilon) = 0$  und  $\text{Cov}(\varepsilon) = \sigma^2 I_n$  mit  $\sigma^2 > 0$ .

# 1. Vorüberlegungen

## Lineares Regressionsmodell

$y = X\beta + \varepsilon$  mit den vier Annahmen:

- (i)  $X$  nichtstochastische  $n \times p$  Matrix mit  $p < n$ ,
- (ii)  $X$  hat vollen Spaltenrang,
- (iii)  $y$  beobachtbarer Zufallsvektor der Länge  $n$ ,
- (iv)  $\varepsilon$  nichtbeobachtbarer Zufallsvektor der Länge  $n$ , so dass  $E(\varepsilon) = 0$  und  $\text{Cov}(\varepsilon) = \sigma^2 I_n$  mit  $\sigma^2 > 0$ .

## Gewöhnlicher Kleinste-Quadrate-Schätzer

$$\hat{\beta} = (X'X)^{-1}X'y$$

# Motivation

## Multikollinearität

- Starke Korrelation zwischen einigen Variablen in  $X$
- Bsp.: Aktienkurse, Gesundheitsdaten, sozialer Status
- Führt zu großer Varianz des KQ-Schätzers

# Motivation

## Multikollinearität

- Starke Korrelation zwischen einigen Variablen in  $X$
- Bsp.: Aktienkurse, Gesundheitsdaten, sozialer Status
- Führt zu großer Varianz des KQ-Schätzers

## Ausweg

- Einige Variablen ausschließen
- Alternative Schätzer, die eine kleinere Varianz haben, aber verzerrt sind: Ridge-, Shrinkage- und **Hauptkomponentenschätzer**

## Ausgangspunkt

Multikollinearität führt dazu, dass einige Eigenwerte von  $X'X$  sehr klein sind:

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3.1 & 3 & 3 \\ 4 & 4 & 4.1 \end{pmatrix} \Rightarrow \lambda_1 = 91.410, \quad \lambda_2 = 0.010, \quad \lambda_3 = 0.001$$

## Ausgangspunkt

Multikollinearität führt dazu, dass einige Eigenwerte von  $X'X$  sehr klein sind:

$$X_1 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3.1 & 3 & 3 \\ 4 & 4 & 4.1 \end{pmatrix} \Rightarrow \lambda_1 = 91.410, \quad \lambda_2 = 0.010, \quad \lambda_3 = 0.001$$

$$X_2 = \begin{pmatrix} 1 & 1 & -2 \\ 2 & 0 & 1 \\ -1 & 1 & 2 \\ -2 & 0 & -1 \end{pmatrix} \Rightarrow \lambda_1 = 10, \quad \lambda_2 = 10, \quad \lambda_3 = 2$$

## Spektralzerlegung

- Spektralzerlegung  $X'X = U\Lambda U'$  mit
  - ▶  $\Lambda$  : Diagonalmatrix der Eigenwerte  $\lambda_1, \dots, \lambda_p$  von  $X'X$
  - ▶  $U$  : Orthogonale Matrix der Eigenvektoren  $u_1, \dots, u_p$  von  $X'X$

## Spektralzerlegung

- Spektralzerlegung  $X'X = U\Lambda U'$  mit

- ▶  $\Lambda$  : Diagonalmatrix der Eigenwerte  $\lambda_1, \dots, \lambda_p$  von  $X'X$
- ▶  $U$  : Orthogonale Matrix der Eigenvektoren  $u_1, \dots, u_p$  von  $X'X$

$$\begin{aligned}(X'X)^{-1} &= (U\Lambda U')^{-1} \\ &= U'^{-1}\Lambda^{-1}U^{-1} \\ &= U\Lambda^{-1}U' \quad (\text{da } U \text{ orthogonal: } U' = U^{-1}) \\ &= \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i'.\end{aligned}$$

## Spektralzerlegung

- Spektralzerlegung  $X'X = U\Lambda U'$  mit

- ▶  $\Lambda$  : Diagonalmatrix der Eigenwerte  $\lambda_1, \dots, \lambda_p$  von  $X'X$
- ▶  $U$  : Orthogonale Matrix der Eigenvektoren  $u_1, \dots, u_p$  von  $X'X$

$$\begin{aligned}(X'X)^{-1} &= (U\Lambda U')^{-1} \\ &= U'^{-1}\Lambda^{-1}U^{-1} \\ &= U\Lambda^{-1}U' \quad (\text{da } U \text{ orthogonal: } U' = U^{-1}) \\ &= \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i'\end{aligned}$$

$$\begin{aligned}\Rightarrow \text{Cov}(\hat{\beta}) &= \sigma^2(X'X)^{-1} \\ &= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i'\end{aligned}$$

## Spektralzerlegung

- Spektralzerlegung  $X'X = U\Lambda U'$  mit

- ▶  $\Lambda$  : Diagonalmatrix der Eigenwerte  $\lambda_1, \dots, \lambda_p$  von  $X'X$
- ▶  $U$  : Orthogonale Matrix der Eigenvektoren  $u_1, \dots, u_p$  von  $X'X$

$$\begin{aligned}(X'X)^{-1} &= (U\Lambda U')^{-1} \\ &= U'^{-1}\Lambda^{-1}U^{-1} \\ &= U\Lambda^{-1}U' \quad (\text{da } U \text{ orthogonal: } U' = U^{-1}) \\ &= \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i'\end{aligned}$$

$$\begin{aligned}\Rightarrow \text{Cov}(\hat{\beta}) &= \sigma^2(X'X)^{-1} \\ &= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i' \quad \leftarrow \text{kleine Eigenwerte: große Varianz!}\end{aligned}$$

## 2. Herleitung des Hauptkomponentenschätzers

### Spektralzerlegung

- Betrachte nun Spektralzerlegung von  $X'X$ :

$$X'X = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} U_1' \\ U_2' \end{pmatrix}$$

- $\Lambda_1$  Diagonalmatrix mit  $r \leq p$  „großen“ Eigenwerten
- $\Lambda_2$  Diagonalmatrix mit  $p - r$  „kleinen“ Eigenwerten
- $U_1$  enthält die ersten  $r$  Spalten von  $U$
- $U_2$  enthält die verbleibenden  $p - r$  Spalten von  $U$

## Idee des Hauptkomponentenschätzers

- Eliminiere die  $p - r$  kleinen Eigenwerte
- Große Varianz des Schätzers soll so verhindert werden

## Idee des Hauptkomponentenschätzers

- Eliminiere die  $p - r$  kleinen Eigenwerte
- Große Varianz des Schätzers soll so verhindert werden

## Restriktion

- Schätzer soll sich nicht in Richtung der kleinen Eigenwerte  $\lambda_{r+1}u_{r+1}, \dots, \lambda_p u_p$  ausbreiten
- Daraus ergibt sich Restriktion:

$$\begin{aligned}\beta &= U_1 \xi, \quad \xi \text{ beliebiger Vektor} \\ \Leftrightarrow U_1 U_1' \beta &= \beta \\ \Leftrightarrow U_2' \beta &= 0\end{aligned}$$

## Zur Erinnerung: Restringierter KQ-Schätzer

Restriktion:  $R\beta = a$ 

$$\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - a)$$

## Zur Erinnerung: Restringierter KQ-Schätzer

Restriktion:  $R\beta = a$

$$\hat{\beta}_R = \hat{\beta} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - a)$$

## Hauptkomponentenschätzer

- Hauptkomponentenschätzer ist spezieller restringierter Schätzer mit  $U_2'\beta = 0$  :

$$\hat{\beta}^{(r)} = \hat{\beta} - (X'X)^{-1}U_2[U_2'(X'X)^{-1}U_2]^{-1}U_2'\hat{\beta}$$

- andere Schreibweise:

$$\begin{aligned}\hat{\beta}^{(r)} &= U_1(U_1'X'XU_1)^{-1}U_1'X'y \\ &= U_1\Lambda_1^{-1}U_1'X'y\end{aligned}$$

## Beispiel

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad y = \begin{pmatrix} 6.0521 \\ 7.0280 \\ 7.1230 \\ 4.4441 \\ 5.0813 \end{pmatrix}$$

## Beispiel

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad y = \begin{pmatrix} 6.0521 \\ 7.0280 \\ 7.1230 \\ 4.4441 \\ 5.0813 \end{pmatrix}$$

## gewöhnlicher KQ-Schätzer

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} -4.2489 \\ 5.2013 \end{pmatrix}$$

Große Abweichung vom wahren  $\beta$

## Spektralzerlegung

Spektralzerlegung

$$X'X = U\Lambda U'$$

mit

$$U = ( U_1 \quad U_2 ) = \left( \begin{array}{c|c} -0.4537 & -0.8912 \\ -0.8912 & 0.4537 \end{array} \right)$$

und

$$\Lambda = \left( \begin{array}{cc} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{array} \right) = \left( \begin{array}{cc} 24.2493 & 0 \\ 0 & 0.0107 \end{array} \right)$$

## Spektralzerlegung

Spektralzerlegung

$$X'X = U\Lambda U'$$

mit

$$U = (U_1 \quad U_2) = \left( \begin{array}{c|c} -0.4537 & -0.8912 \\ -0.8912 & 0.4537 \end{array} \right)$$

und

$$\Lambda = \left( \begin{array}{cc} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{array} \right) = \left( \begin{array}{cc} 24.2493 & 0 \\ 0 & 0.0107 \end{array} \right)$$

## Hauptkomponentenschätzer mit $r = 1$

$$\hat{\beta}^{(1)} = U_1(U_1'X'XU_1)^{-1}U_1'X'y = \left( \begin{array}{c} 1.2284 \\ 2.4128 \end{array} \right)$$

Geringere Abweichung vom wahren  $\beta$  als bei  $\hat{\beta}$

## Hauptkomponentenschätzer mit $r = 2$

Für  $r = 2$  ergibt sich in diesem Beispiel der gewöhnliche KQ-Schätzer:

$$\begin{aligned}\hat{\beta}^{(2)} &= U(U'X'XU)^{-1}U'X'y \\ &= \underbrace{U\Lambda^{-1}U'}_{=(X'X)^{-1}}X'y \\ &= \begin{pmatrix} -4.2489 \\ 5.2013 \end{pmatrix} \\ &= \hat{\beta}\end{aligned}$$

→ Daher Wahl von  $r$  in diesem Fall trivial

### 3. Alternative Herleitung

#### Hauptkomponentenanalyse

- $X$  großer Datensatz mit vielen zusammenhängenden Variablen
- Transformation zu Hauptkomponenten  $Z = X \cdot U$
- Hauptkomponenten mit den größten zugehörigen Eigenwerten enthalten Großteil der Variation von  $X$

### 3. Alternative Herleitung

#### Hauptkomponentenanalyse

- $X$  großer Datensatz mit vielen zusammenhängenden Variablen
- Transformation zu Hauptkomponenten  $Z = X \cdot U$
- Hauptkomponenten mit den größten zugehörigen Eigenwerten enthalten Großteil der Variation von  $X$

#### Umformulierung des Linearen Modells

$$\begin{aligned}
 y &= X\beta + \varepsilon \\
 \Leftrightarrow y &= \underbrace{XU}_Z \underbrace{U'\beta}_\gamma + \varepsilon \\
 \Leftrightarrow y &= Z\gamma + \varepsilon.
 \end{aligned}$$

## Hauptkomponentenschätzer

- $\hat{\gamma} = (Z'Z)^{-1}Z'y \Rightarrow \hat{\beta} = U\hat{\gamma}$
- Für  $\hat{\beta}^{(r)}$  wird  $Z$  auf die ersten  $r$  Hauptkomponenten beschränkt!  
(Statt Variablen in  $X$  werden Hauptkomponenten weggelassen!)

## Hauptkomponentenschätzer

- $\hat{\gamma} = (Z'Z)^{-1}Z'y \Rightarrow \hat{\beta} = U\hat{\gamma}$
- Für  $\hat{\beta}^{(r)}$  wird  $Z$  auf die ersten  $r$  Hauptkomponenten beschränkt!  
(Statt Variablen in  $X$  werden Hauptkomponenten weggelassen!)

## Betrachtung als Summe

$$\hat{\beta} = \sum_{i=1}^p \frac{1}{\lambda_i} u_i u_i' X' y.$$

$$\hat{\beta}^{(r)} = \sum_{i=1}^r \frac{1}{\lambda_i} u_i u_i' X' y.$$

Äquivalent zur Definition aus Kapitel 2:  $\hat{\beta}^{(r)} = U_1 \Lambda_1^{-1} U_1' X' y.$

## 4. Eigenschaften

### Erwartungswert und Kovarianz

Eigenschaften vom restringierten Schätzer ableitbar:

- Linearer Schätzer
- Kovarianz:

$$\text{Cov}(\hat{\beta}^{(r)}) = \sigma^2 \sum_{i=1}^r \frac{1}{\lambda_i} u_i u_i'.$$

- Erwartungswert:

$$E(\hat{\beta}^{(r)}) = \beta - \sum_{i=r+1}^p u_i u_i' \beta.$$

## Vergleich mit dem KQ-Schätzer

Für  $\Delta = \text{MSE}(\beta, \hat{\beta}) - \text{MSE}(\beta, \hat{\beta}^{(r)})$  gilt:

$$\Delta \text{ ist n.n.d.} \Leftrightarrow \beta' U_2 \Lambda_2 U_2' \beta \leq \sigma^2.$$

## Schlussfolgerung

Für Überlegenheit von  $\hat{\beta}^{(r)}$  über  $\hat{\beta}$  nötig:

- Kleine Diagonalelemente in  $\Lambda_2$  (Multikollinearität)
- $\sigma^2$  nicht zu klein

## Vergleich mit dem KQ-Schätzer

Für  $\Delta = \text{MSE}(\beta, \hat{\beta}) - \text{MSE}(\beta, \hat{\beta}^{(r)})$  gilt:

$$\Delta \text{ ist n.n.d.} \Leftrightarrow \beta' U_2 \Lambda_2 U_2' \beta \leq \sigma^2.$$

## Schlussfolgerung

Für Überlegenheit von  $\hat{\beta}^{(r)}$  über  $\hat{\beta}$  nötig:

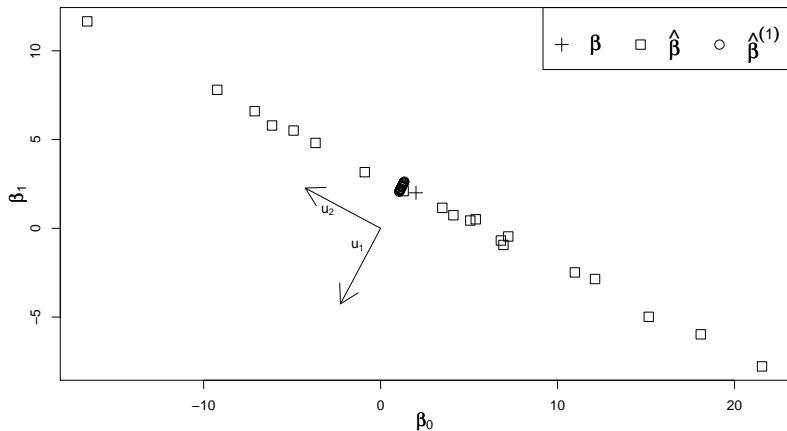
- Kleine Diagonalelemente in  $\Lambda_2$  (Multikollinearität)
- $\sigma^2$  nicht zu klein

## Beispiel aus Kapitel 2

$$0.0082 \leq 1.$$

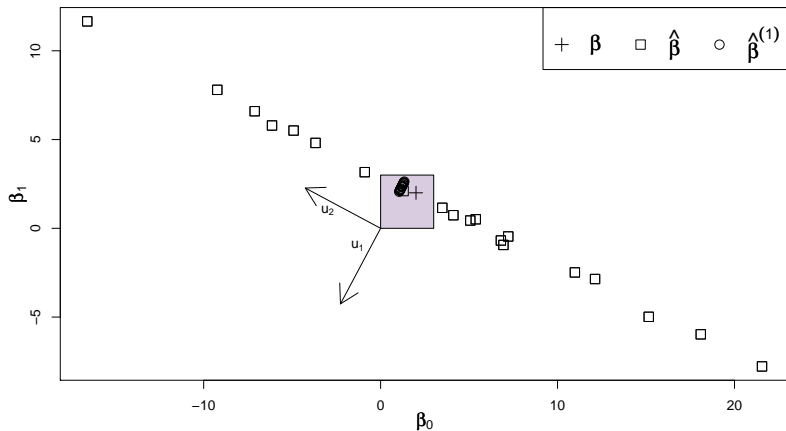
## 20 Simulationen für Beispiel

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad \sigma^2 = 1$$



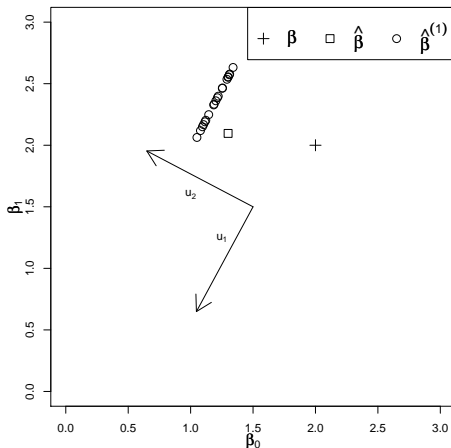
## 20 Simulationen für Beispiel

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad \sigma^2 = 1$$



## 20 Simulationen für Beispiel, höhere Auflösung

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad \sigma^2 = 1$$

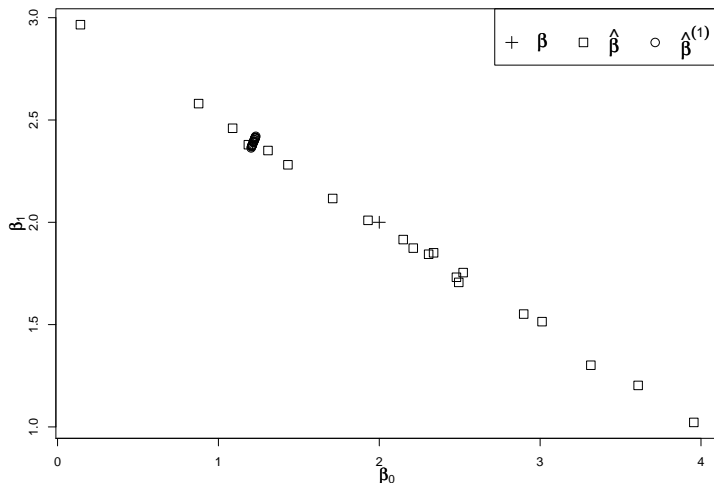


## 20 Simulationen bei kleinerer Fehlervarianz

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad \sigma^2 = 0.1$$

## 20 Simulationen bei kleinerer Fehlervarianz

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix}, \quad \sigma^2 = 0.1$$

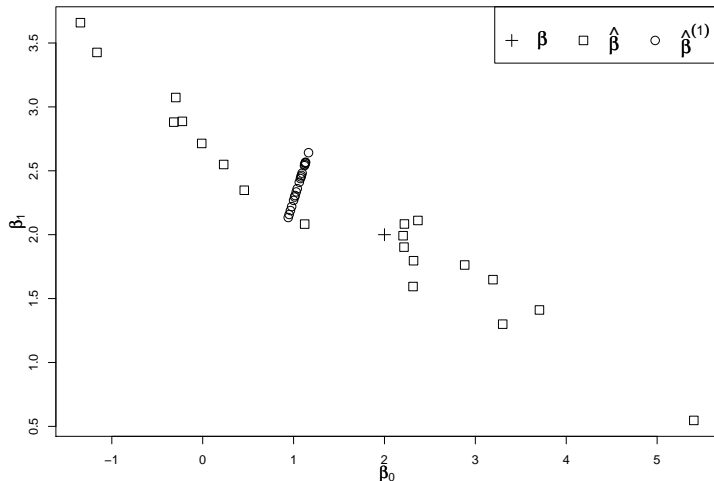


## 20 Simulationen bei schwächerer Multikollinearität

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.3 & 3 & 2.5 & 1.4 & 2.3 \end{pmatrix}, \quad \sigma^2 = 1$$

## 20 Simulationen bei schwächerer Multikollinearität

$$\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.3 & 3 & 2.5 & 1.4 & 2.3 \end{pmatrix}, \quad \sigma^2 = 1$$



## Weitere Bemerkungen

- Auch ähnliche Vergleiche zwischen der Methode, Variablen auszuschließen und Hauptkomponentenschätzer möglich
  - Trenkler (1991). Mean square error matrix comparisons between biased restricted least squares estimators. *Sankhya: The Indian Journal of Statistics* 53, S. 317-318

## Weitere Bemerkungen

- Auch ähnliche Vergleiche zwischen der Methode, Variablen auszuschließen und Hauptkomponentenschätzer möglich
  - Trenkler (1991). Mean square error matrix comparisons between biased restricted least squares estimators. *Sankhya: The Indian Journal of Statistics* 53, S. 317-318
- Es können Beispiele konstruiert werden, bei denen die ersten  $(p - 1)$  Hauptkomponenten, die fast 100 % der Varianz beinhalten, nichts zur Anpassung von  $y$  beitragen, wohingegen die letzte ignorierte Hauptkomponente  $y$  perfekt anpasst
  - Hadi und Ling (1998). Some cautionary notes on the use of principal components regression. *The American Statistician* 52 (1), S. 15-19

## 5. Anzahl der Hauptkomponenten

### Wahl von $r$

- Wahl von  $r$  meist nicht trivial
- Einerseits soll  $\hat{\beta}^{(r)}$  so wenig wie möglich beschränkt werden  
→ Wähle  $r$  so groß wie möglich
- Andererseits soll Effekt der Multikollinearität verhindert werden  
→ Wähle  $r$  so klein wie nötig
- In Literatur werden viele Methoden zur Wahl von  $r$  diskutiert

## 1. Methode

- Teste, ob ein bereits gewähltes  $r$  zu klein ist
- Teste dazu die Hypothesen

$$H_0 : U_2' \beta = 0 \quad \text{gegen} \quad H_1 : U_2' \beta \neq 0$$

mit Teststatistik

$$F = \frac{(\hat{\beta} - \hat{\beta}^{(r)})' X' X (\hat{\beta} - \hat{\beta}^{(r)})}{y' M y} \cdot \frac{n - p}{p - r}$$

mit  $M = I_n - X(X'X)^{-1}X'$ ,  $I_n$   $n$ -dim. Einheitsmatrix.

Lehne  $H_0$  ab, falls  $F \geq F_{p-r; n-p; 1-\alpha}$

- Falls  $H_0$  abgelehnt wird, wähle  $r$  größer

## Erinnerung: Beispiel aus Kapitel 2

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix},$$

$$y = ( 6.0521 \quad 7.0280 \quad 7.1230 \quad 4.4441 \quad 5.0813 )',$$

$$\hat{\beta} = ( -4.2489 \quad 5.2013 )', \quad \hat{\beta}^{(1)} = ( 1.2284 \quad 2.4128 )',$$

$$n = 5, \quad p = 2 \text{ und } r = 1$$

## Erinnerung: Beispiel aus Kapitel 2

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1.9 & 2.1 & 2 & 2 & 1.8 \end{pmatrix},$$

$$y = ( 6.0521 \quad 7.0280 \quad 7.1230 \quad 4.4441 \quad 5.0813 )',$$

$$\hat{\beta} = ( -4.2489 \quad 5.2013 )', \quad \hat{\beta}^{(1)} = ( 1.2284 \quad 2.4128 )',$$

$$n = 5, \quad p = 2 \text{ und } r = 1$$

## 1. Methode: Beispiel aus Kapitel 2

Teststatistik:

$$F = 0.2918$$

$$\alpha = 0.05 \rightarrow F_{2-1;5-2;1-0.05} = 10.128$$

Da  $F < F_{1;3;0.95}$ , lehne  $H_0$  nicht ab

## 2. Methode

- Vergleiche zwei Hauptkomponentenschätzer mit unterschiedlichem  $r$  ( $r_1 > r_2$ )
- Es gilt:

$\hat{\beta}^{(r_2)}$  ist nicht schlechter als  $\hat{\beta}^{(r_1)}$   $\Leftrightarrow \gamma_j = 0, j = r_2 + 1, \dots, r_1$

mit  $\gamma = (\gamma_1 \dots \gamma_p)' = U\beta$

## Erinnerung: Beispiel aus Kapitel 2

$$U = \begin{pmatrix} -0.4537 & -0.8912 \\ -0.8912 & 0.4537 \end{pmatrix}$$

## 2. Methode: Beispiel aus Kapitel 2

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = U\beta = \begin{pmatrix} -2.6898 \\ -0.8750 \end{pmatrix}$$

- $\gamma_2 \neq 0$ , aber Vermutung liegt nahe, dass  $\hat{\beta}^{(1)}$  dennoch nicht schlechter ist als  $\hat{\beta}^{(2)}$
- Mögliche Erklärung: Vergleich hier schwierig, weil  $\hat{\beta}^{(2)}$  gewöhnlicher KQ-Schätzer
- Vermutlich müssen Dimensionen für diesen Vergleich höher sein

### 3. Methode

- Finde optimales  $r$  mittels
  - ▶ schrittweiser Vorwärtsselektion
  - ▶ schrittweiser Rückwärtsselektion
- Bei beiden Verfahren werden Hauptkomponenten in Reihenfolge der Größe ihrer Eigenwerte betrachtet

## 4. Methode

- Größe des Eigenwerts sagt nichts über Einfluss auf abhängige Variable aus
- Betrachte daher Hauptkomponenten nicht nur in Reihenfolge der Größe ihrer Eigenwerte
- Prüfe Einfluss der Hauptkomponenten z.B. mittels  $t$ -Test und wähle danach die Hauptkomponenten aus (siehe Depczynski *et al.*, 2000)

## 6. Marquardt-Schätzer

### Eigenschaften: Marquardt-Schätzer

- Abwandlung des Hauptkomponentenschätzers
- Eliminieren Einfluss der kleinen Eigenwerte nicht ganz
- Mischung aus Hauptkomponentenschätzer und gewöhnlichem KQ-Schätzer
- Marquardt-Schätzer ist kein restringierter KQ-Schätzer

## Erinnerung

- gewöhnlicher KQ-Schätzer:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (U\Lambda^{-1}U')X'y\end{aligned}$$

- Hauptkomponentenschätzer:

$$\hat{\beta}^{(r)} = U_1\Lambda_1^{-1}U_1'X'y$$

## Erinnerung

- gewöhnlicher KQ-Schätzer:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (U\Lambda^{-1}U')X'y\end{aligned}$$

- Hauptkomponentenschätzer:

$$\hat{\beta}^{(r)} = U_1\Lambda_1^{-1}U_1'X'y$$

## Definition: Marquardt-Schätzer

$$\hat{\beta}^{(r,c)} = (U_1\Lambda_1^{-1}U_1' + c \cdot U_2\Lambda_2^{-1}U_2')X'y, \quad 0 \leq c \leq 1$$

- $c = 1 \Rightarrow$  gewöhnlicher KQ-Schätzer
- $c = 0 \Rightarrow$  Hauptkomponentenschätzer

## Marquardt-Schätzer: Beispiel aus Kapitel 2

- Wähle erneut  $r = 1$
- Wähle für  $c$  exemplarisch zwei Werte:

▶  $c = 0.1$  :

$$\hat{\beta}^{(1,0.1)} = \begin{pmatrix} 0.6898 \\ 2.6872 \end{pmatrix}$$

▶  $c = 0.3$  :

$$\hat{\beta}^{(1,0.3)} = \begin{pmatrix} -0.3874 \\ 3.2356 \end{pmatrix}$$

- $\hat{\beta}^{(1,0.1)}$  liegt näher am wahren  $\beta = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$  als  $\hat{\beta}^{(1,0.3)}$

## 7. Zusammenfassung

### Vorteile des Hauptkomponentenschätzers

- Wesentlich kleinere Varianz als gewöhnlicher KQ-Schätzer bei Vorliegen von Multikollinearität
- Geringer Informationsverlust

## 7. Zusammenfassung

### Vorteile des Hauptkomponentenschätzers

- Wesentlich kleinere Varianz als gewöhnlicher KQ-Schätzer bei Vorliegen von Multikollinearität
- Geringer Informationsverlust

### Nachteile des Hauptkomponentenschätzers

- Verzerrtheit, vor allem bei schwächerer Multikollinearität oder kleiner Fehlervarianz
- Entscheidungskriterien, welche Hauptkomponenten ausgeschlossen werden sollen, sind umstritten

# Literatur



Depczynski, U., V. J. Frost und K. Molt (2000).

Genetic algorithms applied to the selection of factors in principal component regression.

*Analytica Chimica Acta* 420, 217–227.



Groß, J. (2003).

*Linear regression*.

Berlin [u.a.]: Springer.



Hadi, A. S. und R. F. Ling (1998, Februar).

Some cautionary notes on the use of principal components regression.

*The American Statistician* 52(1), 15–19.



Jolliffe, I. T. (1986).

*Principal Component Analysis*.

New York: Springer.



Sufian, A. J. M. (2005).

Analyzing collinear data by principal component regression approach - An example from developing countries.

*Journal of Data Science* 3, 221–232.