

l_1 -Schätzer und robuste Alternativen

- Ausgewählte Kapitel der Regressionsanalyse -
Prof. Dr. Götz Trenkler

Jennifer Upmann

22.01.2009

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

- 1 Einleitung
 - Motivation
 - Vorliegen nicht-normalverteilter Residuen
 - Robuste Schätzer
- 2 Der gewichtete Kleinste-Quadrate-Schätzer
 - Robuste Schätzung von σ
 - Verschiedene Gewichte
 - Iterativ neugewichteter Kleinste-Quadrate-Schätzer
- 3 l_1 -Schätzer
 - Die Idee des l_1 -Schätzers
 - l_1 -Schätzer im Zweidimensionalen
 - Eigenschaften des l_1 -Schätzers
- 4 M-Schätzer
- 5 Robuster Ridge-Schätzer
- 6 Literatur

Einleitung

3/55

Auftretende Probleme beim Anwenden der KQ-Schätzung

- 1 Abhängigkeiten unter den Regressoren
- 2 Fehler, die nicht normalverteilt sind

⇒ Alternativen zum KQ-Schätzer betrachten

Einleitung

4/55

Modellannahmen für das lineare Regressionsmodell $y = X\beta + \epsilon$

- (i) X ist eine nicht-stochastische $n \times p$ Matrix mit $p < n$
- (ii) Die Matrix X hat vollen Spaltenrang, das heißt $\text{rg}(X) = p$
- (iii) Die Elemente des $n \times 1$ Vektors y sind beobachtbare Zufallsvariablen
- (iv) Die Elemente des $n \times 1$ Vektors ϵ sind nicht beobachtbare Zufallsvariablen mit
 $E(\epsilon) = 0$ und $\text{Cov}(\epsilon) = \sigma^2 I_n$ mit $\sigma^2 > 0$

Zusatzannahme

- (v) $\epsilon \sim N(0, \sigma^2 I_n)$

⇒ In diesem Vortrag nicht erfüllt!

Einleitung

4/55

Modellannahmen für das lineare Regressionsmodell $y = X\beta + \epsilon$

- (i) X ist eine nicht-stochastische $n \times p$ Matrix mit $p < n$
- (ii) Die Matrix X hat vollen Spaltenrang, das heißt $\text{rg}(X) = p$
- (iii) Die Elemente des $n \times 1$ Vektors y sind beobachtbare Zufallsvariablen
- (iv) Die Elemente des $n \times 1$ Vektors ϵ sind nicht beobachtbare Zufallsvariablen mit
 $E(\epsilon) = 0$ und $\text{Cov}(\epsilon) = \sigma^2 I_n$ mit $\sigma^2 > 0$

Zusatzannahme

- (v) $\epsilon \sim N(0, \sigma^2 I_n)$

⇒ In diesem Vortrag nicht erfüllt!

Motivation

5/55

Beispiel 1: Gegeben sei $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \dots, 12$

$$x = (2.9 \ 3.6 \ 3.9 \ 3.7 \ 4.4 \ 4.1 \ 4.7 \ 3.7 \ 4.5 \ 5.1 \ 5.7 \ 5.5)'$$

$$y = (4.0 \ 4.5 \ 5.1 \ 5.2 \ 6.0 \ 5.3 \ 6.3 \ 7.5 \ 6.0 \ 5.8 \ 6.9 \ 7.0)'$$

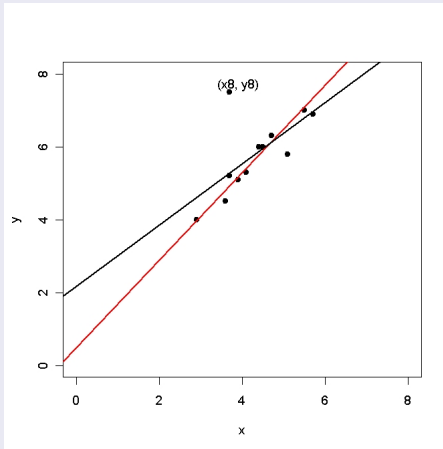
$$\beta = (0.5 \ 1.2)'$$

$$\hat{\beta}_{KQ} = (2.1785 \ 0.8389)'$$

$$L(\beta, \hat{\beta}_{KQ}) = (\hat{\beta}_{KQ} - \beta)'(\hat{\beta}_{KQ} - \beta) = 2.9478$$

Motivation

6/55



Motivation

7/55

Ausreißer (nach Rousseeuw und Leroy)

Falls eine Beobachtung stark von der linearen Beziehung der Mehrheit der Beobachtungen abweicht. Dabei werden sowohl die unabhängigen als auch die abhängigen Beobachtungen betrachtet.

Bemerkung

Ausreißer können von unbeobachtbaren Realisationen einzelner Fehler ϵ_i verursacht werden.

Motivation

7/55

Ausreißer (nach Rousseeuw und Leroy)

Falls eine Beobachtung stark von der linearen Beziehung der Mehrheit der Beobachtungen abweicht. Dabei werden sowohl die unabhängigen als auch die abhängigen Beobachtungen betrachtet.

Bemerkung

Ausreißer können von unbeobachtbaren Realisationen einzelner Fehler ϵ_i verursacht werden.

Vorliegen nicht-normalverteilter Residuen

8/55

Annahme: $\epsilon_i \text{ u.i.v. } \sim N(0, \sigma^2 I_n)$

Dann gilt: $P(-3\sigma \leq \epsilon_i \leq 3\sigma) = 0.9973, \quad i = 1, \dots, n$

\Rightarrow Extreme Realisationen von ϵ_i außerhalb von $[-3\sigma, 3\sigma]$ treten selten auf

Vorliegen nicht-normalverteilter Residuen

9/55

Annahme: Die Verteilung der Fehler ist eine Verteilung mit schweren Rändern

- Verteilung, dessen Ränder mehr Wahrscheinlichkeitsmasse aufweisen als die Ränder der Normalverteilung
- Extreme einzelne Fehler, die Ausreißer in den Daten produzieren, treten wesentlich häufiger auf
- Gilt nur für Ausreißer in den abhängigen Beobachtungen y_i , $i = 1, \dots, n$
- Hier: Grund für das Auftreten von Ausreißern

Vorliegen nicht-normalverteilter Residuen

10/55

1. Beispiel einer Verteilung mit schweren Rändern

$$F_X = (1-\alpha)\Phi\left(\frac{x-\mu}{\tau}\right) + \alpha\Phi\left(\frac{x-\mu}{c\tau}\right), \quad 0 \leq \alpha \leq 1, \quad 0 < c, \quad 0 < \tau$$

mit

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Es gilt:

$$E(X) = \mu \text{ und } Var(X) = (1-\alpha)\tau^2 + \alpha c^2\tau^2$$

Vorliegen nicht-normalverteilter Residuen

11/55

1. Beispiel einer Verteilung mit schweren Rändern

Gesetzt: $\mu = 0$, $\tau = 1$, $\alpha = 0.2$ und $c = 5$

$$\Rightarrow \sigma^2 := \text{Var}(X) = (1 - \alpha)\tau^2 + \alpha c^2 \tau^2 = 5.8$$

Vergleich: $X \sim 0.8\Phi(x) + 0.2\Phi(x/5)$ und $Y \sim N(0, 5.8)$

- $P(-\sigma \leq X \leq \sigma) = 0.8612$ und $P(-3\sigma \leq X \leq 3\sigma) = 0.9702$
- $P(-\sigma \leq Y \leq \sigma) = 0.6827$ und $P(-3\sigma \leq Y \leq 3\sigma) = 0.9973$

Vorliegen nicht-normalverteilter Residuen

11/55

1. Beispiel einer Verteilung mit schweren Rändern

Gesetzt: $\mu = 0$, $\tau = 1$, $\alpha = 0.2$ und $c = 5$

$$\Rightarrow \sigma^2 := \text{Var}(X) = (1 - \alpha)\tau^2 + \alpha c^2 \tau^2 = 5.8$$

Vergleich: $X \sim 0.8\Phi(x) + 0.2\Phi(x/5)$ und $Y \sim N(0, 5.8)$

- $P(-\sigma \leq X \leq \sigma) = 0.8612$ und $P(-3\sigma \leq X \leq 3\sigma) = 0.9702$
- $P(-\sigma \leq Y \leq \sigma) = 0.6827$ und $P(-3\sigma \leq Y \leq 3\sigma) = 0.9973$

Vorliegen nicht-normalverteilter Residuen

12/55

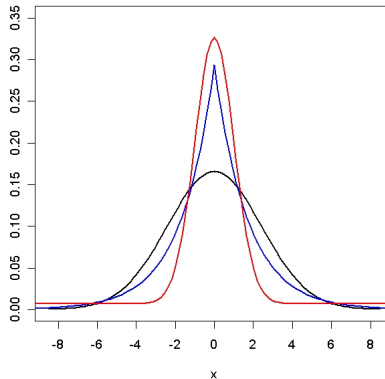
2. Beispiel einer Verteilung mit schweren Rändern: Laplace-Verteilung

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x - \alpha|}{\beta}}, -\infty < \alpha < \infty, 0 < \beta$$

- Hier: $\alpha = 0$ und $\beta = \sqrt{\sigma^2/2}$ mit $\sigma^2 = 5.8$

Vorliegen nicht-normalverteilter Residuen

13/55



Robuste Schätzer

14/55

Was soll das Verwenden von robusten Schätzer bewirken?
(nach Huber)

- Vernünftige Effizienz für das verwendete Modell
- Kleine Abweichungen von den Modellannahmen sollen das Ergebnis nur gering beeinträchtigen
- Trotz größere Abweichungen von den Modellannahmen sollen zufriedenstellende Ergebnisse erzielt werden

⇒ Schwerwiegende Fehler in einem kleinen Anteil der Beobachtungen sollen als kleine Abweichung angesehen werden

Der gewichtete Kleinste-Quadrate-Schätzer

15/55

Umgang mit Ausreißern

Eine Möglichkeit: Ausreißer vor der Durchführung der Analyse aus dem Datensatz entfernen

Aber: Sie können wichtige Informationen enthalten!

Weitere Möglichkeit: Methode finden, die Ausreißer bewertet und ihnen ein angemessenes Gewicht gibt

⇒ Anwendung des gewichteten Kleinsten-Quadrate-Schätzers

Der gewichtete Kleinste-Quadrate-Schätzer

16/55

Der gewichtete Kleinste-Quadrate-Schätzer

Minimierung der Funktion $f_W(\beta_*) = (y - X\beta_*)'W(y - X\beta_*)$

$$\Rightarrow \hat{\beta}_W = (X'WX)^{-1}X'Wy$$

- W : Diagonalmatrix mit den Diagonalelementen w_{ii}
- w_{ii} : hängen von dem i-ten Element $\hat{\epsilon}_i$ des KQ-Residuums $\hat{\epsilon} = y - X\hat{\beta}_{KQ}$ ab

Der gewichtete Kleinste-Quadrate-Schätzer

17/55

Die Gewichtsfunktion

$$w_{ii} = \begin{cases} \frac{\Psi(\hat{\epsilon}_i / \hat{\sigma}_{rob})}{\hat{\epsilon}_i / \hat{\sigma}_{rob}} & \text{falls } \hat{\epsilon}_i \neq 0 \\ 1 & \text{falls } \hat{\epsilon}_i = 0 \end{cases}$$

- $\hat{\sigma}_{rob}$ ein robuster Schätzer für σ
- $\Psi(\cdot)$ erzielt Wert mit dem gleichen Vorzeichen und einen absoluten Wert nicht größer als $\hat{\epsilon}_i / \hat{\sigma}_{rob} \Rightarrow 0 \leq w_{ii} \leq 1$
- Diagonalelemente der Matrix W sind stochastisch
- Matrix W ist positiv definit

Robuste Schätzung von σ

18/55

Robuste Schätzung von σ

- $\hat{\sigma}_{rob} = \frac{1}{0.6745} \text{med}(|\hat{\epsilon}_i - \text{med}(\hat{\epsilon}_i)|) = \frac{MADM}{0.6745}$
- $\hat{\sigma}_{rob} = \frac{\text{med}(|\hat{\epsilon}_i|)}{0.6745} = \frac{MAD}{0.6745}$

Rechtfertigung des Vorfaktors $\frac{1}{0.6745}$

$$Z \sim N(0, \sigma^2)$$

Dann: $\text{med}(|Z - \text{med}(Z)|) = \text{med}(|Z|) = \sigma u_{0.75}$ mit
 $u_{0.75} = 0.6745$ dem 0.75-Quantil der Standardnormalverteilung

Robuste Schätzung von σ

18/55

Robuste Schätzung von σ

- $\hat{\sigma}_{rob} = \frac{1}{0.6745} \text{med}(|\hat{\epsilon}_i - \text{med}(\hat{\epsilon}_i)|) = \frac{MADM}{0.6745}$
- $\hat{\sigma}_{rob} = \frac{\text{med}(|\hat{\epsilon}_i|)}{0.6745} = \frac{MAD}{0.6745}$

Rechtfertigung des Vorfaktors $\frac{1}{0.6745}$

$$Z \sim N(0, \sigma^2)$$

Dann: $\text{med}(|Z - \text{med}(Z)|) = \text{med}(|Z|) = \sigma u_{0.75}$ mit
 $u_{0.75} = 0.6745$ dem 0.75-Quantil der Standardnormalverteilung

Verschiedene Gewichte

19/55

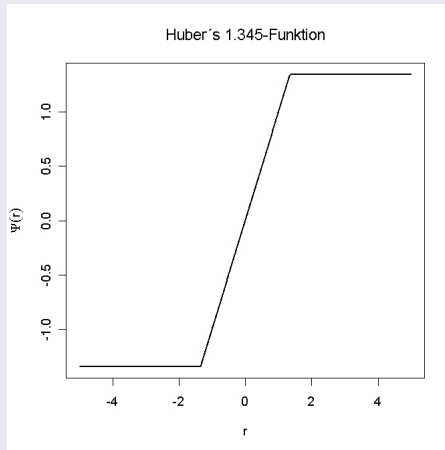
Huber's t -Funktion

$$\Psi(r) = \begin{cases} r & |r| \leq t \\ t \cdot \text{sign}(r) & |r| > t \end{cases}$$

- Absolutbetrag von $\hat{\epsilon}_i / \hat{\sigma}_{rob}$ größer als $t > 0 \Rightarrow w_{ii}$ sehr klein
- Fall $w_{ii} = 0$ ausgeschlossen
- Absolutbetrag von $\hat{\epsilon}_i / \hat{\sigma}_{rob}$ nicht größer als $t \Rightarrow w_{ii} = 1$
- Wahl des Parameters: oft $t = 1.345$, $t = 1.5$ und $t = 2.0$

Verschiedene Gewichte

20/55



Verschiedene Gewichte

21/55

2. Beispiel: Gewichteter KQ-Schätzer basierend auf Huber's 1.345-Funktion für die Daten aus Beispiel 1

Berechnung von...

① $\hat{\beta}_{KQ} = (2.1785 \ 0.8389)'$

② $\hat{\epsilon}_i = y_i - (\hat{\beta}_{KQ})_0 - (\hat{\beta}_{KQ})_1 x_i$

③ $\hat{\sigma}_{rob} = \frac{1}{0.6745} \text{med}(|\hat{\epsilon}_i - \text{med}(\hat{\epsilon}_i)|) = 0.3920$

④ w_{ii} mit Huber's 1.345-Funktion:

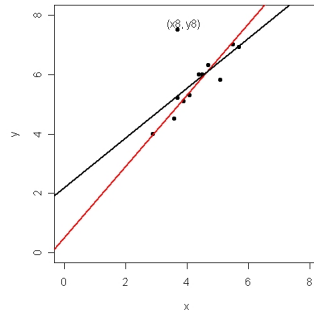
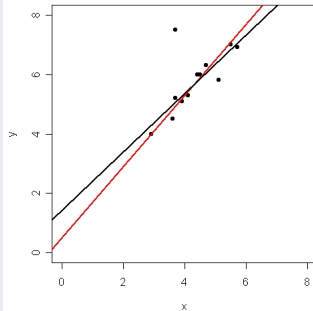
$$w_{ii} = (0.862 \ 0.755 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0.238 \ 1 \ 0.802 \ 1 \ 1)'$$

⑤ $\hat{\beta}_W = (1.4172 \ 0.9848)'$

$\Rightarrow L(\beta, \hat{\beta}_W) = 0.8875$

Verschiedene Gewichte

22/55



Verschiedene Gewichte

23/55

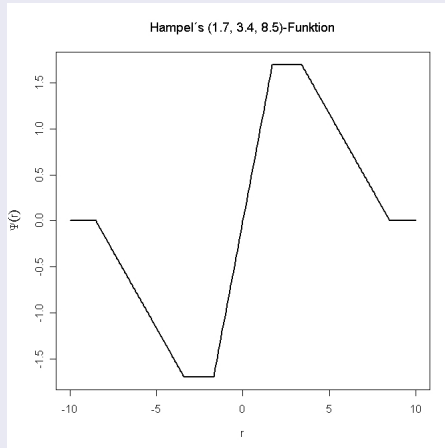
Hampel's (a, b, c) -Funktion

$$\Psi(r) = \begin{cases} r & |r| \leq a \\ a \cdot \text{sign}(r) & a < |r| \leq b \\ a \cdot \text{sign}(r) \left(\frac{c - |r|}{c - b} \right) & b < |r| \leq c \\ 0 & |r| > c, \end{cases}$$

- Gewöhnliche Wahl der Parameter:
 $a = 1.7$, $b = 3.4$ und $c = 8.5$

Verschiedene Gewichte

24/55



Verschiedene Gewichte

25/55

Ramsay's E_a -Funktion

$$\Psi(r) = e^{-a|r|}r$$

- Wahl des Parameters: $a = 0.3$

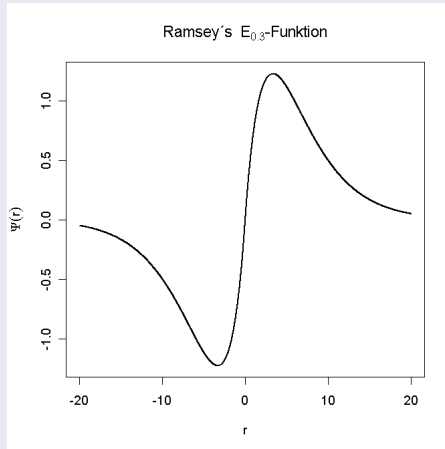
Andrews $\sin(a)$ -Funktion

$$\Psi(r) = \begin{cases} \sin(r/a) & |r| \leq \pi a \\ 0 & |r| > \pi a \end{cases}$$

- Wahl des Parameters: $a = 1.0$ oder $a = 1.339$

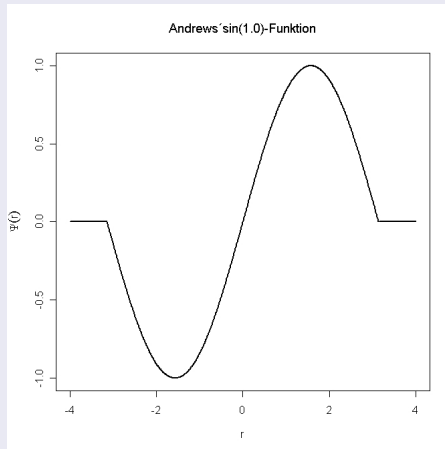
Verschiedene Gewichte

26/55



Verschiedene Gewichte

26/55



Verschiedene Gewichte

27/55

3. Beispiel: Gewichteter KQ-Schätzer basierend auf Andrew's $\sin(a)$ -Funktion für die Daten aus Beispiel 1

Berechnung von...

① $\hat{\beta}_{KQ} = (2.1785 \ 0.8389)'$

② $\hat{\epsilon}_i = y_i - (\hat{\beta}_{KQ})_0 - (\hat{\beta}_{KQ})_1 x_i$

③ $\hat{\sigma}_{rob} = \frac{1}{0.6745} \text{med}(|\hat{\epsilon}_i - \text{med}(\hat{\epsilon}_i)|) = 0.3920$

④ w_{ii} mit Andrew's $\sin(a)$ -Funktion :

$$w_{ii} = (0.64 \ 0.55 \ 0.87 \ 0.99 \ 0.98 \ 0.89 \ 0.97 \ 0 \ 0.99 \ 0.59 \ 0.99 \ 0.95)',$$

⑤ $\hat{\beta}_W = (1.1441 \ 1.0398)'$

$\Rightarrow L(\beta, \hat{\beta}_W) = 0.4405$

Der gewichtete Kleinste-Quadrate-Schätzer

28/55

Anmerkung 1

Ein gewichteter Kleinste-Quadrate-Schätzer mit einer robusten Festlegung der Gewichte ist ein nicht-linearer Schätzer für β .

Anmerkung 2

Die exakten statistischen Eigenschaften sind schwer anzugeben.

Der gewichtete Kleinste-Quadrate-Schätzer

28/55

Anmerkung 1

Ein gewichteter Kleinste-Quadrate-Schätzer mit einer robusten Festlegung der Gewichte ist ein nicht-linearer Schätzer für β .

Anmerkung 2

Die exakten statistischen Eigenschaften sind schwer anzugeben.

Iterativ neugewichteter Kleinste-Quadrate-Schätzer 29/55

Iterativ neugewichteter Kleinste-Quadrate-Schätzer (IRLS-Schätzer)

Durch Iteration:

- 1: Berechnung von $\hat{\beta}_W = (X'WX)^{-1}X'Wy$
- 2: Berechnung von $\hat{\epsilon}_W = y - X\hat{\beta}_W$
- 3: Berechnung von $\hat{\sigma}_{rob}$ und neuen Gewichte w_{ii}
- 4: Berechnung von $\hat{\beta}_{W_1} = (X'W_1X)^{-1}X'W_1y$
- 5: Berechnung von $\hat{\epsilon}_{W_1} = y - X\hat{\beta}_{W_1}$
- 6: ...

Die Idee des l_1 -Schätzers

30/55

Erinnerung

KQ-Schätzung:

$$f(\beta_*) = \|y - X\beta_*\|_2^2 = (y - X\beta_*)'(y - X\beta_*) = \sum_{i=1}^n (y_i - x'_{[i]}\beta_*)^2$$

\Rightarrow Minimum bei $\hat{\beta}_{KQ} = (X'X)^{-1}X'y$

Die Idee des l_1 -Schätzers

Betrachte $\|\cdot\|_1$

$\|x\|_1 = \sum_{i=1}^n |x_i|$ ist die l_1 -Norm

Die Idee des l_1 -Schätzers

30/55

Erinnerung

KQ-Schätzung:

$$f(\beta_*) = \|y - X\beta_*\|_2^2 = (y - X\beta_*)'(y - X\beta_*) = \sum_{i=1}^n (y_i - x'_{[i]}\beta_*)^2$$

\Rightarrow Minimum bei $\hat{\beta}_{KQ} = (X'X)^{-1}X'y$

Die Idee des l_1 -Schätzers

Betrachte $\|\cdot\|_1$

$\|x\|_1 = \sum_{i=1}^n |x_i|$ ist die l_1 -Norm

Die Idee des l_1 -Schätzers

31/55

l_1 -Schätzer

- **Ziel:** Finde Vektor $\hat{\beta}_{LAD}$, so dass $f(\beta_*) = \|y - X\beta_*\|_1 = \sum_{i=1}^n |y_i - x'_{[i]}\beta_*|$ minimiert wird
- Im Jahre 1757 von Roger Joseph Boscovich eingeführt
- Absicht dabei: Umfang der Erde schätzen

l_1 -Schätzer im Zweidimensionalen

32/55

Schätzen der Koeffizienten β_0 und β_1

- Wähle $(\hat{\beta}_{LAD})_0$ und $(\hat{\beta}_{LAD})_1$ so, dass die Summe der Absolutbeträge der Residuen $\sum_{i=1}^n |\hat{\epsilon}_i|$ so klein wie möglich wird
- LAD/ l_1 -Schätzer $(\hat{\beta}_{LAD})_0$ und $(\hat{\beta}_{LAD})_1$ sind die Werte von β_0 und β_1 die

$$\sum_{i=1}^n |y_i - (\beta_0 + \beta_1 x_i)|$$

minimieren

- $y_i - (\beta_0 + \beta_1 x_i)$: Abweichung des Punkts (x_i, y_i) zur Gerade $\hat{y} = X\beta$

l_1 -Schätzer im Zweidimensionalen

33/55

Anmerkung

Es reicht die Linien zu betrachten, die durch mindestens zwei Datenpunkte verlaufen, um die Linie mit der kleinsten absoluten Abweichung zu finden.

Begründung

- 1 $y = a + bx$ in Streudiagramm eingezeichnet:
Soll zunächst durch keinen Punkt verlaufen
- 2 Bewegt wird die Linie solange bis sie durch einen Punkt verläuft
- 3 Linie kann in diesem Punkt rechts oder links herum gedreht werden bis sie durch einen weiteren verläuft

l_1 -Schätzer im Zweidimensionalen

33/55

Anmerkung

Es reicht die Linien zu betrachten, die durch mindestens zwei Datenpunkte verlaufen, um die Linie mit der kleinsten absoluten Abweichung zu finden.

Begründung

- 1 $y = a + bx$ in Streudiagramm eingezeichnet:
Soll zunächst durch keinen Punkt verlaufen
- 2 Bewegt wird die Linie solange bis sie durch einen Punkt verläuft
- 3 Linie kann in diesem Punkt rechts oder links herum gedreht werden bis sie durch einen weiteren verläuft

l_1 -Schätzer im Zweidimensionalen

34/55

1. Algorithmus

Hauptteil dieser Prozedur: Für jeden gegebenen Punkt (x_0, y_0) soll die beste unter allen Linien gefunden werden.

- ❶ Suche für (x_1, y_1) die beste Line, die durch einen Punkt (x_2, y_2) verläuft
- ❷ Suche für (x_2, y_2) die beste Line, die durch einen Punkt (x_3, y_3) verläuft
- ❸ Suche für (x_3, y_3) die beste Linie, ...

⇒ Linien werden zunehmend besser

⇒ Ende des Algorithmus: Wenn sich die Linie nicht mehr ändert

l_1 -Schätzer im Zweidimensionalen

35/55

Was ist die beste Linie für (x_0, y_0) ? [Teil 1]

- 1 Berechne Steigungen $(y_i - y_0)/(x_i - x_0)$ für alle (x_i, y_i)
- 2 Sortiere die Steigungen:
$$(y_1 - y_0)/(x_1 - x_0) \leq (y_2 - y_0)/(x_2 - x_0) \leq \dots \leq (y_n - y_0)/(x_n - x_0)$$
- 3 Definiere T durch $T = \sum |x_i - x_0|$
- 4 Finde Index k mit $|x_1 - x_0| + \dots + |x_{k-1} - x_0| < \frac{1}{2}T$ und
$$|x_1 - x_0| + \dots + |x_{k-1} - x_0| + |x_k - x_0| > \frac{1}{2}T$$

l_1 -Schätzer im Zweidimensionalen

36/55

Was ist die beste Linie für (x_0, y_0) ? [Teil 2]

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- $\hat{\beta}_1 = \frac{y_k - y_0}{x_k - x_0}$
- $\hat{\beta}_0 = y_0 - \hat{\beta}_1 x$

l_1 -Schätzer im Zweidimensionalen

37/55

2. Algorithmus

- 1 Berechne die Summen der absoluten Abweichungen für jede mögliche Linie
- 2 Wähle diejenige mit der kleinsten Summe

Durchführbarkeit: Abhängigkeit von dem Stichprobenumfang n

Nichteindeutigkeit: Willkürliche Wahl einer dieser Linien oder Durchschnittsbildung

Eigenschaften des l_1 -Schätzers

38/55

Robustheit im Vergleich zum KQ-Schätzer

Betrachtet wird $\tilde{\epsilon}_i = y_i - x'_{[i]}\tilde{\beta}$

KQ-Schätzung: $|\tilde{\epsilon}_i|$ groß $\Rightarrow \tilde{\epsilon}_i^2$ noch größer \Rightarrow einzelnes großes Residuum kann KQ-Schätzer stark beeinflussen

l_1 -Schätzung: $|\tilde{\epsilon}_i|$ groß \Rightarrow Reaktion auf Ausreißer nicht so stark ausgeprägt

Aber: Robustheit gilt nicht im Hinblick auf starke Abweichung von einzelnen Beobachtungen in den unabhängigen Beobachtungen

Eigenschaften des l_1 -Schätzers

39/55

4. Beispiel: Vorliegen eines Ausreißers in der unabhängigen Variable

Gegeben: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \dots, 9$

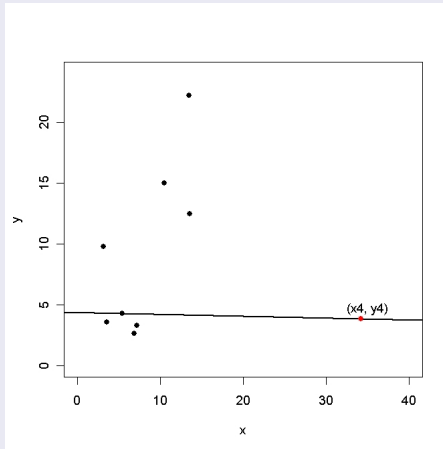
$$x = (3.58 \quad 7.15 \quad 13.58 \quad 34.15 \quad 3.14 \quad 6.83 \quad 5.43 \quad 13.48 \quad 10.50)'$$

$$y = (3.55 \quad 3.30 \quad 12.48 \quad 3.85 \quad 9.80 \quad 2.62 \quad 4.28 \quad 22.20 \quad 15.02)'$$

$$\hat{\beta}_{LAD} = (4.3664 \quad -0.0152)'$$

Eigenschaften des l_1 -Schätzers

40/55



Eigenschaften des l_1 -Schätzers

41/55

Berechnung

- Keine explizite Formel für den l_1 -Schätzer
- Statistische Software-Pakete bieten folgende Möglichkeiten der l_1 -Regression an:

S-PLUS: *l1fit*

R: Paket *quantreg* laden, Funktion *rq(·)* verwenden mit Einstellung *method = br*

Eigenschaften des l_1 -Schätzers

42/55

Nicht-Eindeutigkeit

Keine eindeutige Lösung $\hat{\beta}_{LAD}$ garantiert

5. Beispiel: Nicht-Eindeutigkeit des l_1 -Schätzers

Gegeben: (1, 1), (2, 1) und (2, 2)

Gesucht: $\min_{\beta_{0*}, \beta_{1*}} \sum_{i=1}^3 |y_i - \beta_{0*} - \beta_{1*}x_i|$

Es gilt: $0 \leq (\hat{\beta}_{LAD})_0 \leq 1$ und $(\hat{\beta}_{LAD})_1 = 1 - (\hat{\beta}_{LAD})_0$

Eigenschaften des l_1 -Schätzers

42/55

Nicht-Eindeutigkeit

Keine eindeutige Lösung $\hat{\beta}_{LAD}$ garantiert

5. Beispiel: Nicht-Eindeutigkeit des l_1 -Schätzers

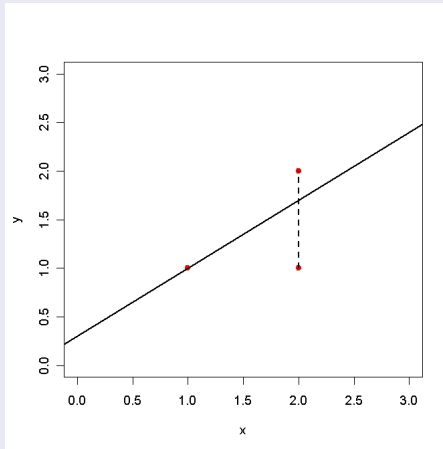
Gegeben: (1, 1), (2, 1) und (2, 2)

Gesucht: $\min_{\beta_{0*}, \beta_{1*}} \sum_{i=1}^3 |y_i - \beta_{0*} - \beta_{1*}x_i|$

Es gilt: $0 \leq (\hat{\beta}_{LAD})_0 \leq 1$ und $(\hat{\beta}_{LAD})_1 = 1 - (\hat{\beta}_{LAD})_0$

Eigenschaften des l_1 -Schätzers

43/55



Eigenschaften des l_1 -Schätzers

44/55

LAD Gleichung und p Datenpunkten

Gegeben: lineares Regressionsmodell mit Achsenabschnitt:

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n)$$

\Rightarrow Beobachtungsmatrix besteht aus n Zeilen

$(x_{i2}, \dots, x_{ip}, y_i), i = 1, \dots, n$

Annahme an $\hat{\beta}_{LAD}$: Mindestens p von n der p -dimensionalen Beobachtungen liegen auf der durch die Lösung generierten Hyperebene:

$$(x_{i2}, \dots, x_{ip}, y_i) \in H_{LAD}$$

Eigenschaften des l_1 -Schätzers

45/55

Einschub: Hyperebene

Betrachte: $v'_* = (1, v_2, \dots, v_p)$

Hyperebene bezüglich $\hat{\beta}_{LAD}$:

$$H_{LAD} = \{(v_2, \dots, v_p, u) : u = v'_* \hat{\beta}_{LAD}\}$$

$p = 2$: Linie

$p = 3$: Ebene im \mathbb{R}^3

Eigenschaften des l_1 -Schätzers

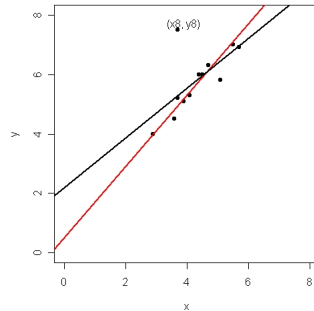
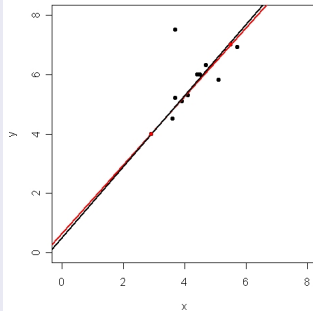
46/55

6. Beispiel: l_1 -Schätzer für β für die Daten aus Beispiel 1

- $\hat{\beta}_{LAD} = (0.6538 \quad 1.1538)'$
- $L(\beta, \hat{\beta}_{LAD}) = 0.258$
- Hyperebene H_{LAD} , die zu $\hat{\beta}_{LAD}$ gehört:
Menge aller Punkte (x, y) mit $y = 0.6538 + 1.1538x$

Eigenschaften des l_1 -Schätzers

47/55



M-Schätzer

Idee des M -Schätzers

- Von Peter Huber 1964 eingeführt
- KQ-Schätzer: $\min_{\beta_*} \sum_{i=1}^n \epsilon_{*i}^2$
- l_1 -Schätzer: $\min_{\beta_*} \sum_{i=1}^n |\epsilon_{*i}|$

⇒ Kombination der beiden Methoden!

M-Schätzer

$$\min_{\beta_*} \sum_{i=1}^n \varrho(\epsilon_{*i})$$

- $\varrho(r)$ kann derart gewählt werden, dass sie
 - ① „im Wesentlichen“ mit r^2 übereinstimmt, wenn $|r| \leq t$
 - ② „im Wesentlichen“ mit $|r|$ übereinstimmt, wenn $|r| > t$

Beispiel für $\varrho(r)$

$$\varrho(r) = \begin{cases} r^2/2 & |r| \leq t \\ t|r| - t^2/2 & |r| > t \end{cases}$$

- Ableitung ist Huber's t -Funktion
- zugehörige IRLS als (annäherende) Lösung des Problems

M-Schätzer

M-Schätzer

$$\min_{\beta_*} \sum_{i=1}^n \varrho(\epsilon_{*i})$$

- $\varrho(r)$ kann derart gewählt werden, dass sie
 - ① „im Wesentlichen“ mit r^2 übereinstimmt, wenn $|r| \leq t$
 - ② „im Wesentlichen“ mit $|r|$ übereinstimmt, wenn $|r| > t$

Beispiel für $\varrho(r)$

$$\varrho(r) = \begin{cases} r^2/2 & |r| \leq t \\ t|r| - t^2/2 & |r| > t \end{cases}$$

- Ableitung ist Huber's t -Funktion
- zugehörige IRLS als (annäherende) Lösung des Problems

Verwenden anderer Funktionen $\varrho(\cdot)$

- Ergibt andere Schätzer
- $\varrho(\hat{\epsilon}_i)$ ist das Maß des Umfangs des i-ten Residuums. Soll folgenden Eigenschaften genügen:
 - 1 Residuum ist Null \Rightarrow Umfang ist Null
 - 2 Residuum weiter von der Null entfernt als ein anderes \Rightarrow Umfang ist größer
 - 3 Alle Umfänge sind nichtnegativ
- $\hat{\beta}_{KQ}$ und $\hat{\beta}_{LAD}$ spezielle M-Schätzer mit $\varrho(r) = r^2$ und $\varrho(r) = |r|$

Buchstabe „M“

- Bezieht sich auf die Maximum-Likelihood-Schätzung
- $\varrho(\epsilon_{*i}) = -\log f(\epsilon_{*i})$ ergibt den gewöhnlichen Maximum-Likelihood-Schätzer
- Manche M-Schätzer sind ebenfalls Maximum-Likelihood-Schätzer unter einer speziellen Verteilung von ϵ

Robuster Ridge-Schätzer

Erinnerung: Ridge-Schätzer

$$\hat{\beta}_k = (X'X + kI_p)^{-1}X'X\hat{\beta}_{KQ}$$

Die Idee des robusten Ridge-Schätzers

- $\hat{\beta}_{KQ}$ durch robusten Schätzer $\hat{\beta}^{rob}$ ersetzen
 $\Rightarrow \hat{\beta}_k^{rob} = (X'X + kI_p)^{-1}X'X\hat{\beta}^{rob}$
- Holland: Bestimmung eines robusten gewichteten Schätzers, dann Anwendung einer Methode der Ridge-Schätzung
- Askin und Montgomery: Festlegung des Ridge-Parameters, dann Anwenden eines robusten Schätzers

Robuster Ridge-Schätzer

Erinnerung: Ridge-Schätzer

$$\hat{\beta}_k = (X'X + kI_p)^{-1}X'X\hat{\beta}_{KQ}$$

Die Idee des robusten Ridge-Schätzers

- $\hat{\beta}_{KQ}$ durch robusten Schätzer $\hat{\beta}^{rob}$ ersetzen
 $\Rightarrow \hat{\beta}_k^{rob} = (X'X + kI_p)^{-1}X'X\hat{\beta}^{rob}$
- Holland: Bestimmung eines robusten gewichteten Schätzers, dann Anwendung einer Methode der Ridge-Schätzung
- Askin und Montgomery: Festlegung des Ridge-Parameters, dann Anwenden eines robusten Schätzers

Robuster Ridge-Schätzer



7. Beispiel: Ridge-Schätzer für die Daten aus Beispiel 1

- Basierend auf dem gewichteten KQ-Schätzer, der mit Huber's t -Funktion

$$\hat{\beta}_k^{rob} = (X'X + kI_2)^{-1} X'X \begin{pmatrix} 1.4172 \\ 0.9848 \end{pmatrix}$$

- Ridge-Parameter: mit Hilfe der Ridge-trace; hier: $k = 3$
 $\hat{\beta}_3^{rob} = (0.4019 \ 1.1967)'$
- $L(\beta, \hat{\beta}_k^{rob}) = 0.0096$

Literaturangabe

-  Askin, R.G., Montgomery, D. C. (1980)
Augmented robust estimators
Technometrics, 22, 333 - 341
-  Birkes, D., Dodge, Y. (1993)
Alternative Methods of Regression
Wiley, New York
-  Groß, J. (2003)
Linear Regression
Springer Verlag, Heidelberg
-  Holland, P. (1973)
Weighted ridge and robust regression methods
NBER Working Paper No.11, Cambridge, MA
-  Huber, P. J. (1981)
Robust Statistics
Wiley & Sons, Inc., USA

Literaturangabe



Ramsay, J.O. (1977)

A comparative study of several robust estimates of slope, intercept and scale in linear regression

Journal of the American Statistical Association, 72, 608-615



Rousseeuw, P. J., Leroy, A. J. (1987)

Robust Regression and Outlier Detection

Wiley, New York



Silvapulle (1991)

Robust ridge regression based on an M-estimator

Australian Journal of Statistics, 33, 319-333



Simpson, J. R., Montgomery, D.C (1998)

The development and evaluation of alternative generalized M-estimation techniques

Communications in Statistics, Simulation, 27, 999-1018