

Iterationsschätzer

Ausgewählte Kapitel der Regressionsanalyse

Mark Neblik

11. Dezember 2008

- 1 Einführung
 - Motivation
 - Anwendungsmöglichkeiten
- 2 Iterationsschätzer
 - Voraussetzungen
 - Herleitung
 - Definition
 - Eigenschaften
- 3 Anwendung in \mathcal{R}
 - Programmierung
 - Daten
 - Vergleich mit KQ-Schätzer
- 4 Ausblick und Diskussion

Motivation

- Der gewöhnliche Kleinste-Quadrate-Schätzer ist nicht immer sinnvoll.
- Insbesondere bei kleinen Eigenwerten von $X'X$ sind Schätzungen und Varianz oft extrem groß.
- Dies tritt z.B. bei Multikollinearität auf.

Idee

- Auch verzerrte Schätzer als Alternativen in Betracht ziehen.
- Gesucht werden Folgen, die gegen die Moore–Penrose–Inverse konvergieren.

Anwendungsmöglichkeiten

- Iterationsschätzer als Alternative zum gewöhnlichen KQ–Schätzer.
- Multikollinearität tritt in vielen Bereichen auf, hauptsächlich in der Ökonometrie.

Lineares Modell

Wir setzen einen linearen Zusammenhang der Art

$$y = X\beta + u$$

voraus. Hierbei ist y die abhängige Variable mit T Beobachtungen und X die Regressormatrix, welche p erklärende Variablen enthält. Weiter sei β der wahre Koeffizientenvektor und u ein Fehlervektor mit Mittelwert 0 und Varianz σ^2 .

Die Eigenwerte von $X'X$ werden mit $\lambda_1, \dots, \lambda_r$ bezeichnet, wobei $\lambda_1 > \dots > \lambda_r$.

Herleitung des Iterationsschätzers

- Der Ridge-Schätzer $\hat{\beta}_k = (X'X + kI)^{-1}X'y$ konvergiert gegen den Minimum-Norm-Schätzer $\hat{\beta}^+ = X^+y$ für $k \rightarrow 0$.
- Da für $k \rightarrow 0$ $(X'X + kI)$ fast wieder singulär ist, ist der Vorteil des Ridge-Schätzers dahin.
- Suche Folge, die anders gegen X^+ konvergiert.

Definition des Iterationsschätzers

Definiere

$$X_{n,\alpha} := \alpha \sum_{i=0}^n (I - \alpha X'X)^i X' .$$

Dann ist

$$\hat{\beta}_{n,\alpha} := X_{n,\alpha} y$$

der Iterationsschätzer für den Parametervektor β .

Konvergenz gegen Moore-Penrose-Inverse

Für $\alpha \in (0, \frac{2}{\lambda_1})$ gilt:

$$\lim_{n \rightarrow \infty} X_{n,\alpha} = X^+$$

Konvergenz gegen Minimum-Norm-Schätzer

Für $\alpha \in (0, \frac{2}{\lambda_1})$ gilt:

$$\lim_{n \rightarrow \infty} \hat{\beta}_{n,\alpha} = \hat{\beta}^+$$

Erwartungswert, Varianz, etc.

Für den Iterationsschätzer gelten folgende Aussagen:

- (i) $E(\hat{\beta}_{n,\alpha}) = X_{n,\alpha}X\beta$
- (ii) $\text{Var}(\hat{\beta}_{n,\alpha}) = \sigma^2 \text{tr}(X'_{n,\alpha}X_{n,\alpha})$
- (iii) $\text{Bias}(\hat{\beta}_{n,\alpha})^2 = \beta'(X_{n,\alpha}X - I)^2\beta$
- (iv) $\text{Cov}(\hat{\beta}_{n,\alpha}) = \sigma^2 X_{n,\alpha}X'_{n,\alpha}$
- (v) $\text{MSE}(\hat{\beta}_{n,\alpha}) = \beta'(X_{n,\alpha}X - I)^2\beta + \sigma^2 \text{tr}(X'_{n,\alpha}X_{n,\alpha})$

Grenzwerte für Erwartungswert, Varianz, etc.

Für den Iterationsschätzer gelten folgende Grenzwertaussagen:

$$(ia) \lim_{n \rightarrow \infty} E(\hat{\beta}_{n,\alpha}) = \beta$$

$$(iia) \lim_{n \rightarrow \infty} \text{Var}(\hat{\beta}_{n,\alpha}) = \sigma^2 \text{tr}((X'X)^{-1})$$

$$(iiia) \lim_{n \rightarrow \infty} \text{Bias}(\hat{\beta}_{n,\alpha})^2 = 0$$

$$(iva) \lim_{n \rightarrow \infty} \text{Cov}(\hat{\beta}_{n,\alpha}) = \sigma^2 (X'X)^{-1}$$

$$(va) \lim_{n \rightarrow \infty} \text{MSE}(\hat{\beta}_{n,\alpha}) = \sigma^2 \text{tr}((X'X)^{-1})$$

$$(vi) \lim_{n \rightarrow \infty} \hat{\beta}'_{n,\alpha} \hat{\beta}_{n,\alpha} = \hat{\beta}' \hat{\beta}$$

Iterationsschätzer in Rao's Klasse

Der Iterationsschätzer gehört zu Rao's Klasse von linearen homogenen Schätzer, d.h. es gelten

- (i) $\hat{\beta}_{n,\alpha}$ ist ein homogener linearer Bayes-Schätzer für $\beta \in \mathbb{R}^p$
- (ii) $\hat{\beta}_{n,\alpha}$ ist ein genereller Ridge-Schätzer für $\beta \in \mathbb{R}^p$
- (iii) $\hat{\beta}_{n,\alpha}$ ist ein Minimax-Schätzer für $\beta \in \Theta_N$

wobei $\Theta_N = \{\beta | \beta' N \beta \leq \sigma^2\}$ und N eine passende positiv definite Matrix ist.

Quadratische Länge und Varianz des Iterationsschätzers

Seien $\hat{\beta}^{(1)}$ und $\hat{\beta}^{(2)}$ lineare homogene Schätzer, dann gilt mit Wahrscheinlichkeit 1:

$$\hat{\beta}^{(1)'}\hat{\beta}^{(1)} < \hat{\beta}^{(2)'}\hat{\beta}^{(2)} \Rightarrow \text{Var} \left(\hat{\beta}^{(1)} \right) < \text{Var} \left(\hat{\beta}^{(2)} \right)$$

Zusätzlich gilt, dass im Sinne der euklidischen Norm der Iterationsschätzer echt kleiner als der KQ-Schätzer ist, d.h. es gilt:

$$\hat{\beta}_{n,\alpha}'\hat{\beta}_{n,\alpha} < \hat{\beta}'\hat{\beta} \quad \forall n \in \mathbb{N} \text{ und } \alpha \in \left(0, \frac{2}{\lambda_1} \right)$$

aus beiden Aussagen folgt, dass der Iterationsschätzer eine kleinere Varianz als der KQ-Schätzer besitzt.

Bestimmung von n_c

Sei c eine bekannte Konstante, so dass $\beta' \beta \leq c \sigma^2$. Sei n_c die kleinste natürliche Zahl, so dass

$$(1 - \alpha \lambda_r)^{n_c+1} < \frac{2}{c \lambda_1 + 1}$$

mit $\alpha \in \left(0, \frac{1}{\lambda_1}\right)$ erfüllt ist. Dann gilt:

$$MSE(\hat{\beta}_{n,\alpha}) < \text{Var}(\hat{\beta}) \quad \forall n \geq n_c .$$

Damit ist gezeigt, dass ein n_c existiert, ab dem der Iterationsschätzer einen kleineren mittleren quadratischen Fehler besitzt als der gewöhnliche KQ-Schätzer.

Kriterien, so dass $\text{MSE}(\hat{\beta}_{n,\alpha}) < \text{MSE}(\hat{\beta})$ gilt

Die folgenden Ungleichungen sind ausreichende Kriterien dafür, dass $\text{MSE}(\hat{\beta}_{n,\alpha}) < \text{MSE}(\hat{\beta})$:

- (i) $\beta' P' \text{diag}\{\lambda_1(1 - \alpha\lambda_1)^{n+1}, \dots, \lambda_r(1 - \alpha\lambda_r)^{n+1}\} P \beta < \sigma^2$
- (ii) $(1 - \alpha\lambda_r)^{n+1} \beta' X' X \beta < \sigma^2$
- (iii) $\beta' X' X \beta < \sigma^2$
- (iv) $\lambda_1(1 - \alpha\lambda_r)^{n+1} \beta' \beta < \sigma^2$
- (v) $\lambda_1 \beta' \beta < \sigma^2$

Robustheit des MSE des Iterationsschätzers

Für den mittleren quadratischen Fehler des Iterationsschätzers gilt:

$$\lim_{\lambda_r \rightarrow 0} \text{MSE}(\hat{\beta}_{n,\alpha}) = c \text{ mit } c \in \mathbb{R}$$

d.h. für kleiner werdende Eigenwerte konvergiert der MSE gegen eine feste Grenze.

Bestimmung optimaler Parameter α und n

Um optimale Parameter für den Iterationsschätzer zu finden, sind folgende Aussagen nützlich:

- (i) $\widehat{\beta}'_{n,\alpha}\widehat{\beta}_{n,\alpha}$ ist streng monoton steigend mit n und α .
- (ii) $\text{Bias}(\widehat{\beta}_{n,\alpha})$ ist streng monoton fallend mit n und α .
- (iii) $\text{Var}(\widehat{\beta}_{n,\alpha})$ ist streng monoton steigend mit n und α .

Im Normalfall wird ein α im Konvergenzintervall fest gewählt und dann ein optimales n bestimmt.

Strategien zur Bestimmung von n

Es gibt im Wesentlichen zwei Strategien zur Bestimmung von n :

- Wähle n aus dem Bereich in dem sich die Iterationsspur stabilisiert hat (Hoerl–Kennard–Strategie)
- Wähle n so, dass $\hat{\beta}'_{n,\alpha}\hat{\beta}_{n,\alpha} = \hat{Q}$, wobei $\hat{Q} = \hat{\sigma}^2 \sum_{j=1}^p \frac{1}{\lambda_j}$ mit $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-p}$ und $\hat{u} = y - X\hat{\beta}$ (McDonald–Galarneau–Strategie)

Iterative Folge

Für die Folge von $X_{n,\alpha}$ gilt:

- (i) $X_{0,\alpha} = \alpha X'$
- (ii) $X_{n+1,\alpha} = (I - \alpha X'X)X_{n,\alpha} + \alpha X'$

Für den Iterationsschätzer gilt analog:

- (i) $\hat{\beta}_{0,\alpha} = \alpha X'y$
- (ii) $\hat{\beta}_{n+1,\alpha} = (I - \alpha X'X)\hat{\beta}_{n,\alpha} + \alpha X'y$

Der Algorithmus des Iterationsschätzers ist nicht selbst korrigierend.

Schnellere Folge

$X_{n,\alpha}$ konvergiert nur langsam gegen X^+ , außer wenn $\frac{\lambda_1}{\lambda_r}$ klein ist. Eine Folge die schneller konvergiert ist:

$$Y_{n+1,\alpha} = (2I - Y_{n,\alpha}X)Y_{n,\alpha} \text{ mit } Y_{0,\alpha} = \alpha X' \text{ mit } \alpha \in \left(0, \frac{2}{\lambda_1}\right).$$

Es gilt für diese Folge, dass $Y_{n,\alpha} = X_{2^n-1,\alpha}$.

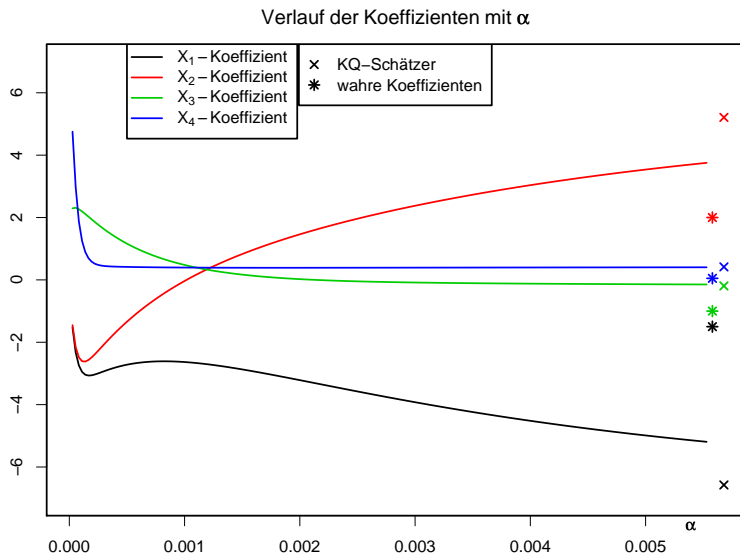
Für den Vergleich generierte Daten

Zum Vergleich des Iterations- und des KQ-Schätzers werden folgende Daten verwendet:

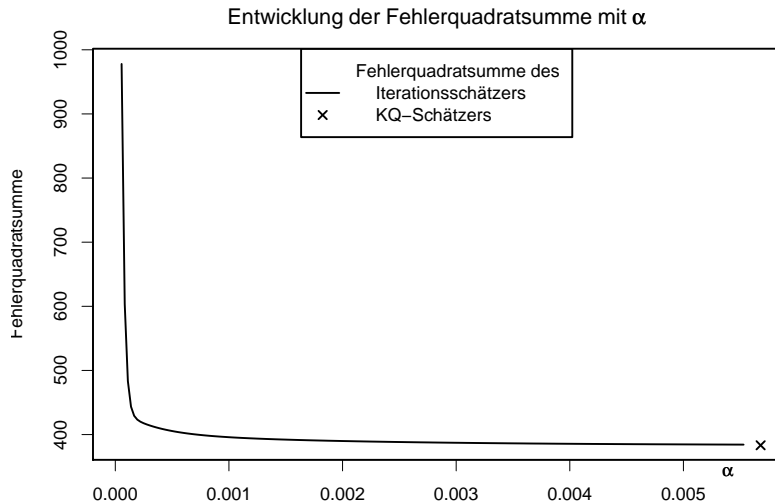
- $X_1 \sim \mathcal{N}(0, 1)$ mit 50 Elementen.
- $X_2 = X_1 + \mathcal{N}(0, \eta)$, wobei zunächst $\eta = 0.1$ ist.
- $X_3 = X_1 + X_2 + \mathcal{Re}(0, 1)$.
- $X_4 = -0.4 \cdot X_2 + \mathcal{Exp}(1)$.
- $X = (1_{50}, X_1, X_2, X_3, X_4)$.
- $y = 20 - 1.5 \cdot X_1 + 2 \cdot X_2 - X_3 + 0.05 \cdot X_4 + \mathcal{N}(0, 3)$.

$$\Rightarrow \beta = (20, -1.5, 2, -1, 0.05)'$$

Anwendung – Vergleich mit KQ-Schätzer

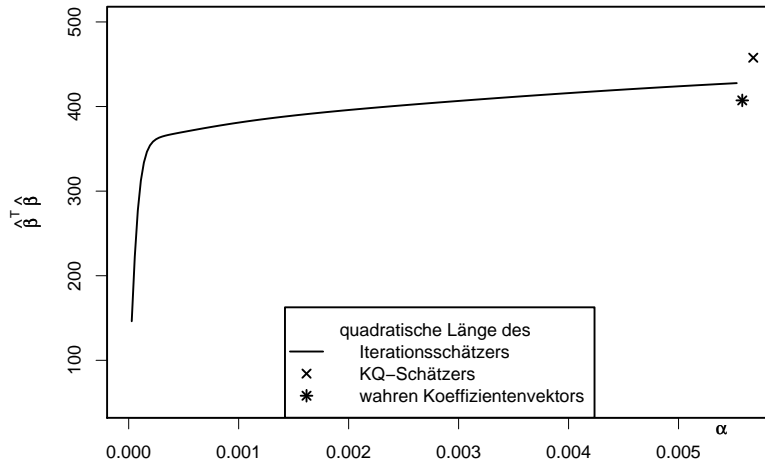


Anwendung – Vergleich mit KQ-Schätzer

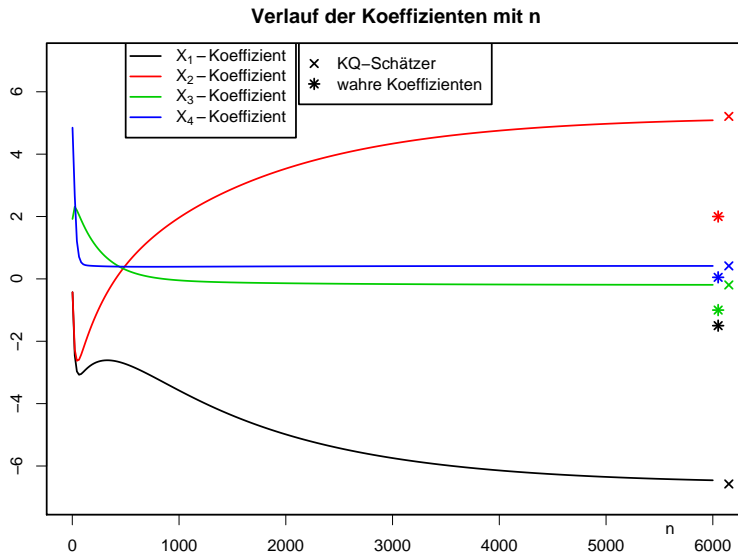


Anwendung – Vergleich mit KQ-Schätzer

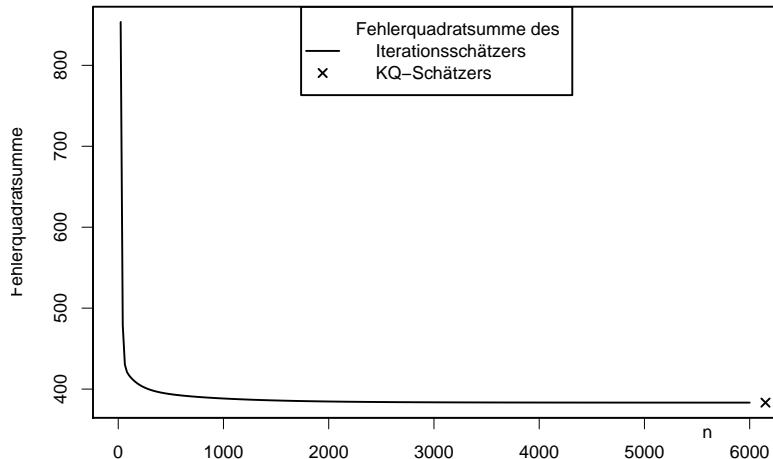
Entwicklung der quadratischen Länge des Iterationsschätzers mit α



Anwendung – Vergleich mit KQ-Schätzer

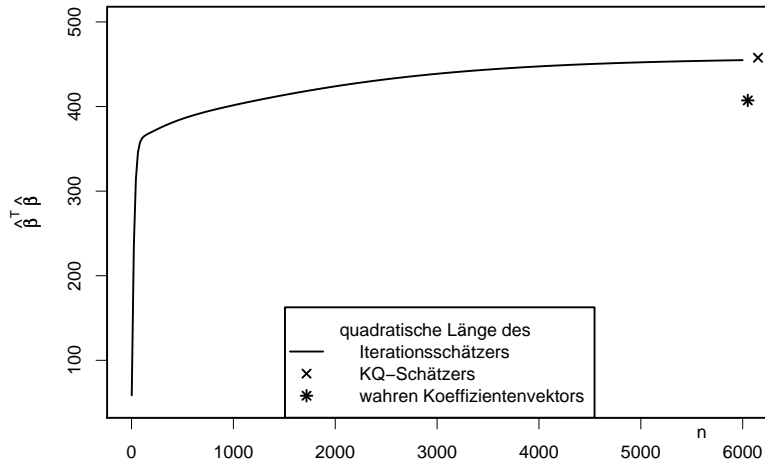


Entwicklung der Fehlerquadratsumme mit n



Anwendung – Vergleich mit KQ-Schätzer

Entwicklung der quadratischen Länge des Iterationsschätzers mit n



Für den Vergleich generierte Daten

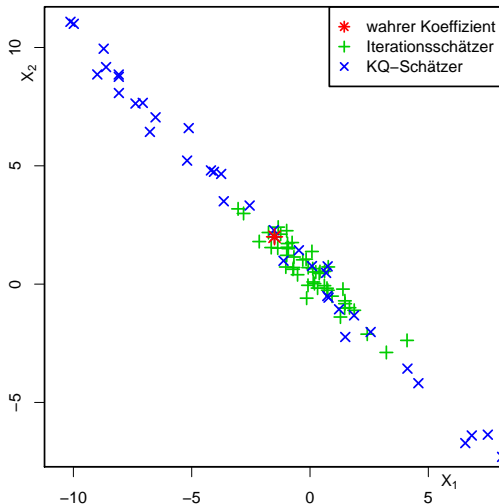
Nun werden die Schätzer $\hat{\beta}_{1000,0.0025}$ und $\hat{\beta}$ 50 Mal für die folgenden Daten geschätzt und verglichen mit den wahren Werten.

- $X_1 \sim \mathcal{N}(0, 1)$ mit 50 Elementen.
- $X_2 = X_1 + \mathcal{N}(0, \eta)$, wobei $\eta = 0.05$ bzw. $\eta = 0.2$ ist.
- $X = (1_{50}, X_1, X_2)$.
- $y = 20 - 1.5 \cdot X_1 + 2 \cdot X_2 + \mathcal{N}(0, 3)$.

$$\Rightarrow \beta = (20, -1.5, 2)'$$

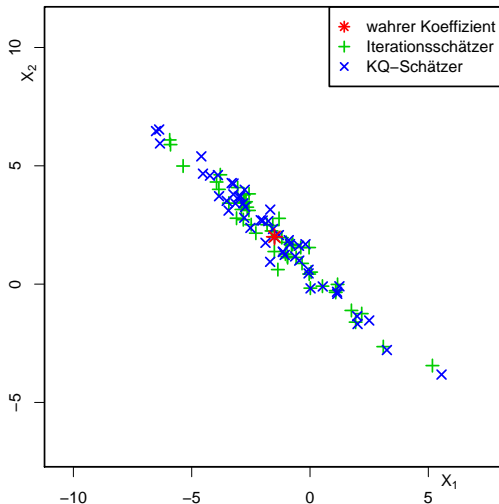
Anwendung – Vergleich mit KQ-Schätzer

Vergleich KQ – und Iterationsschätzer mit $\eta = 0.05$



Anwendung – Vergleich mit KQ-Schätzer

Vergleich KQ – und Iterationsschätzer mit $\eta = 0.2$



Ausblick

- Der Iterationsschätzer ist als Schätzer, der gegen $\hat{\beta}^+$ konvergiert, gut bei auftretender Multikollinearität.
- Weitere sinnvolle Schätzer, die vielleicht noch bessere oder andere Eigenschaften besitzen, können gefunden werden, indem Folgen bestimmt werden, die gegen X^+ konvergieren.

- Trenkler, G. (1981)
Biased estimators in the linear regression
Anton Hain Meisenheim GmbH, Königsstein.
- R Development Core Team (2008)
R: A Language and Environment for Statistical Computing
Version R 2.8.0., Wien.