Kapitel 2
Das Lineare Regressionsmodell

Angewandte Ökonometrie
WS 2010/11

Nikolaus Hautsch
Humboldt-Universität zu Berlin

Inhalt I

2.1 Modell und Annahmen
- 2.1.1 Modellgleichung und Variablen
- 2.1.2 Standardannahmen

2.2 Eigenschaften des KQ-Schätzers
- 2.2.1 Generelle Eigenschaften
- 2.2.2. Unter Normalität

2.3 Lineare Hypothesentests

2.4 Generalisierte Fehlertermstrukturen
- 2.4.1 KQ-Schätzung
- 2.4.2 Generalisierte KQS
- 2.4.3 FGLS Schätzung

2.1.1 Modellgleichung und Variablen

- N Beobachtungen einer abhängigen (metrischen) Variable y: $y_1, \ldots, y_N$
- K erklärende Variablen $x^{(1)}, x^{(2)}, \ldots, x^{(K)}$: $x_i := (x_{i1}, \ldots, x_{iK})'$ für $i = 1, \ldots, N$
- Ziel: Bestimmung des Zusammenhanges zwischen y und den erklärenden Variablen $x^{(1)}, \ldots, x^{(K)}$.

Modellgleichung

- Summennotation

$$y_i = \beta_1 + x_i^{(2)}\beta_2 + ... + x_i^{(K)}\beta_K + \varepsilon_i, \quad i = 1, ..., N$$

- Vektornotation

$$y_i = \sum_{k=1}^{K} \beta_k x_i^{(k)} + \varepsilon_i = x_i'\beta + \varepsilon_i$$

- Matrixnotation: $y = X\beta + \varepsilon,$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{(N\times 1)}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}_{(N\times 1)}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}_{(K\times 1)}$$

$$\begin{aligned} X &= \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1K} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2K} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{NK} \end{bmatrix}_{(N\times K)} = \begin{bmatrix} x_1' \\ x_2' \\ \cdots \\ x_N' \end{bmatrix} \\ &= [x^{(1)}, \ldots, x^{(K)}], \end{aligned}$$

wobei $x_i^{(1)} = 1$ für $i = 1, \ldots, N$.

Variablen

- y_i - beobachtbare zufällige Variable:
  Output, Zielgröße, abhängige Variable, Regressand, ...
  ("endogene" Variable)
- x_{ik} - beobachtbare deterministische Variable:
  Input, Eingangsgrößen, unabhängige Variable, Regressoren, ...
  ("exogene" Variable)
- $\beta_1, \beta_2, ..., \beta_K$ - unbekannte Parameter (konstant)
- ε_i - unbeobachtbare Zufallsgrößen, Fehler, Störungen, ...

2.1.2 Standardannahmen

I. **Annahmen über die Regressoren**

(a) Die Regressoren $x_{i1}, \ldots, x_{iK}$ sind deterministisch
$\Rightarrow X$ ist deterministisch

(b) X hat vollen Rang K, d.h., die Spalten von X sind linear unabhängig:

$$rk(X) = K \qquad \Rightarrow \qquad \text{erfordert } N \geq K!$$

II. Annahmen über die Fehlerterme

$$\varepsilon \sim (0, \sigma^2 I_N)$$

(a) Die Fehler haben den Erwartungswert Null: $\mathbb{E}[\varepsilon_i] = 0 \; \forall \, i$
(b) Die Fehler sind homoskedastisch: $\mathrm{V}[\varepsilon_i] = \sigma^2 \quad \forall \, i$
(c) Die Fehler sind unkorreliert: $\mathbb{E}[\varepsilon_i \varepsilon_j] = 0$ für $i \neq j$
(d) Manchmal (für Inferenz): Die Fehler sind normalverteilt.

$$\varepsilon \sim N(0, \sigma^2 I_N)$$
$$\varepsilon_i \sim N(0, \sigma^2) \quad \text{i.i.d.}$$

III. Annahmen über die Modellstruktur

(a) Die unbekannten Parameter β und σ^2 sind konstant.
(b) Das lineare Modell ist adäquat (korrekt spezifiziert), d.h.

$$\mathbb{E}[y_i] = x_i'\beta \quad \text{oder} \quad \mathbb{E}[y] = X\beta$$

$\Rightarrow$ Beobachtungen sind unkorreliert mit gleicher Varianz:

$$\begin{aligned} & y_i \sim (x_i'\beta, \sigma^2) \text{ unkorreliert} \\ \Leftrightarrow \quad & y \sim (X\beta, \sigma^2 I_N) \end{aligned}$$

Inhalt I

2.1 Modell und Annahmen
2.1.1 Modellgleichung und Variablen
2.1.2 Standardannahmen

2.2 Eigenschaften des KQ-Schätzers
2.2.1 Generelle Eigenschaften
2.2.2. Unter Normalität

2.3 Lineare Hypothesentests

2.4 Generalisierte Fehlertermstrukturen
2.4.1 KQ-Schätzung
2.4.2 Generalisierte KQS
2.4.3 FGLS Schätzung

2.2.1 Generelle Eigenschaften

- Der KQS $\widehat{\beta} = (X'X)^{-1}X'y$ ist BLUE für β mit $\mathrm{V}[\widehat{\beta}] = \sigma^2(X'X)^{-1}$.
- Vektor der angepassten Werte: $\widehat{y} = X\widehat{\beta} = Py$ mit Projektionsmatrix $P = X(X'X)^{-1}X'$.
- Residuenvektor: $e := \widehat{\varepsilon} = y - \widehat{y} = y - X\widehat{\beta} = (I - P)y = (I - P)\varepsilon$
- Erwartungstreue Schätzung der Fehlervarianz: $\widehat{\sigma}^2 = \frac{1}{N-K}e'e = \frac{1}{N-K}\sum_{i=1}^{N} e_i^2$

- Numerische Eigenschaften der Residuen:

$$X'e = \sum_{i=1}^{n} x_i e_i = 0,$$

$$\sum_{i=1}^{n} x_{i1} e_i = \sum_{i=1}^{n} e_i = 0.$$

- Streuungszerlegung: $y'y = \widehat{y}'\widehat{y} + e'e$
- Bestimmtheitsmaß:

$$R^2 = \frac{\sum(\widehat{y}_i - \overline{\widehat{y}})^2}{\sum(y_i - \overline{y})^2}$$

2.2.2 Unter Normalität

- Verteilung des KQS:

$$\widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}).$$

- Varianz von $\widehat{\sigma}^2$:

$$\mathrm{V}[\widehat{\sigma}^2] = \frac{2\sigma^4}{n-k}.$$

- Der KQS $\widehat{\beta}$ ist ML-Schätzer und BUE für β.

Inhalt I

2.1 Modell und Annahmen
2.1.1 Modellgleichung und Variablen
2.1.2 Standardannahmen

2.2 Eigenschaften des KQ-Schätzers
2.2.1 Generelle Eigenschaften
2.2.2. Unter Normalität

2.3 Lineare Hypothesentests

2.4 Generalisierte Fehlertermstrukturen
2.4.1 KQ-Schätzung
2.4.2 Generalisierte KQS
2.4.3 FGLS Schätzung

- Test J linearer Hypothesen: $H_0: \; R\beta = r \quad \text{vs.} \quad H_1: \; R\beta \neq 0$
  (R ist $(J \times K)$ mit Rang J und r ist $(J \times 1)$)
- F-Test: α-Test lehnt H_0 ab, falls $F > F^{1-\alpha}_{J,N-K}$, wobei

$$F = \frac{(R\widehat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\widehat{\beta} - r)}{J\widehat{\sigma}^2} \overset{H_0}{\sim} F_{J,N-K}$$

- Fall $J = 1 \Rightarrow F$-Test ist äquivalent zum t-Test ($F_{1,q} = t_q^2$)

Inhalt I

2.1 Modell und Annahmen
2.1.1 Modellgleichung und Variablen
2.1.2 Standardannahmen

2.2 Eigenschaften des KQ-Schätzers
2.2.1 Generelle Eigenschaften
2.2.2. Unter Normalität

2.3 Lineare Hypothesentests

2.4 Generalisierte Fehlertermstrukturen
2.4.1 KQ-Schätzung
2.4.2 Generalisierte KQS
2.4.3 FGLS Schätzung

2.4.1 KQ-Schätzung

- *Generalisiertes* lineares Regressionsmodell:

$$y = X\beta + \varepsilon \; ; \qquad \mathbb{E}[\varepsilon|X] = 0$$

mit

$$\mathrm{V}[\varepsilon|X] = \mathbb{E}[\varepsilon\varepsilon'|X] = \Psi = \sigma^2\Omega \neq \sigma^2 I_n,$$

wobei Ω eine positiv-definite Matrix ist.

- Spezialfälle:
  - Konditionale Heteroskedastie für $\mathrm{V}[\varepsilon_i|X] \equiv \Psi_{ii} \neq \Psi_{jj} \equiv \mathrm{V}[\varepsilon_j|X]$ für $i \neq j$.
  - Autokorrelation für $\mathrm{Cov}[\varepsilon_i, \varepsilon_j|X] = \Psi_{ij} \neq 0$ für $i \neq j$.

KQ-Schätzung wenn $\mathrm{V}[\varepsilon|X] \neq \sigma^2 I_n$

- ***Proposition***: Der KQS $\widehat{\beta} = (X'X)^{-1}X'y$ des linearen Modells mit verallgemeinerter Fehlertermstruktur hat die folgenden Eigenschaften:

  (i) $\mathbb{E}[\widehat{\beta}|X] = \beta$

  (ii) $$\begin{aligned}\mathrm{V}[\widehat{\beta}|X] &= (X'X)^{-1}X'\Psi X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \\ &\neq \sigma^2(X'X)^{-1}\end{aligned}$$

- White (1982) Varianz-Kovarianz-Schätzer:

$$\widehat{\mathrm{V}}[\widehat{\beta}|X] = (X'X)^{-1}X'\widehat{D}X(X'X)^{-1},$$

wobei $\widehat{D} = \mathrm{diag}[e_i^2]$ eine $n \times n$ Diagonalmatrix bezeichnet.

2.4.2 Generalisierte KQ-Schätzung

- Cholesky-Zerlegung: $\Omega^{-1} = P'P$ mit

$$P'P\Omega = P\Omega P' = I_n.$$

- Transformiertes Regressionsmodell:

$$y^* = X^*\beta + \varepsilon^*,$$

wobei $y^* = Py$, $X^* = PX$ und $\varepsilon^* = P\varepsilon$.

- Generalisierter KQ (GLS) Schätzer:

$$\begin{aligned} \widehat{\beta}^* &= (X^{*\prime}X^*)^{-1}X^{*\prime}y^* \\ &= (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y = \widehat{\beta}_{GLS} \end{aligned}$$

Eigenschaften

- Das transformierte Modell erfüllt die vollen idealen Bedingungen:

$$\mathrm{V}[\varepsilon^*|X] = \sigma^2 P(P'P)^{-1}P' = \sigma^2 I_n.$$

- Erwartungswert und Varianz des GLS-Schätzers:

$$\mathbb{E}[\widehat{\beta}_{GLS}|X] = \beta$$
$$\mathrm{V}[\widehat{\beta}_{GLS}|X] = (X'\Psi^{-1}X)^{-1}$$

- $\widehat{\beta}_{GLS}$ ist BLUE:

$$\mathrm{V}[\widehat{\beta}|X] - \mathrm{V}[\widehat{\beta}_{GLS}|X] = \sigma^2(X'X)^{-1} - \sigma^2(X'\Omega^{-1}X)^{-1}$$

ist positiv semi-definit.

2.4.3 FGLS Schätzung

- Problem: Ψ muß geschätzt werden!
- "Verfügbare generalisierte KQ-Schätzung" (feasible generalized least squares, FGLS):

$$\widehat{\beta}_{FGLS} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y,$$

  wobei $\widehat{\Omega}$ ein konsistenter Schätzer von Ω ist.
- Eigenschaften:
  - (i) $\widehat{\beta}_{FGLS}$ ist generell ein nichtlinearer Schätzer.
  - (ii) $\mathbb{E}[\widehat{\beta}_{FGLS}] \neq \beta$ in endlichen Stichproben.
  - (iii) Asymptotisch äquivalent zum ML-Schätzer