Vorlesung: Lineare Modelle

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

SoSe 2015

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN Institut für Statistik STA BLAB Statistisches Beratungs Labor

Das allgemeine lineare Modell

Das lineare Modell mit heteroskedastischen Störgrößen ist gegeben durch:

$$Y = X\beta + \varepsilon \tag{8.1}$$
$$\varepsilon \sim N(0, \sigma^2 V) \tag{8.2}$$
$$V = \text{diag}(v_1, v_2, \dots, v_n) \tag{8.3}$$

V bekannte Matrix zur Beschreibung der Varianzstruktur

Gewichtsmatrix: $W = V^{-1/2} = \text{diag}\,(v_1^{-1/2}, v_2^{-1/2}, \dots, v_n^{-1/2})$

Der gewichtete KQ-Schätzer

$$\hat{\beta}_W := (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{8.4}$$
$$\hat{\sigma}^2 = (\hat{\varepsilon}'V^{-1}\hat{\varepsilon})/(n - p') \tag{8.5}$$

$\hat{\beta}_W$ ist ML-Schätzer und minimiert die **gewichtete Residuenquadratsumme**

$$(Y - X\beta)'V^{-1}(Y - X\beta). \tag{8.6}$$

Herleitung durch Transformation

Das Modell (8.1)-(8.3) lässt sich in ein gewöhnliches lineares Modell transformieren:

$$
\begin{aligned}
Y^* &:= WY && (8.7)\\
X^* &:= WX && (8.8)\\
\varepsilon^* &:= W\varepsilon && (8.9)
\end{aligned}
$$

Dann gilt:

$$
\begin{aligned}
Y^* &= X^*\beta + \varepsilon^* && (8.10)\\
\varepsilon^* &\sim N(0, \sigma^2 I) && (8.11)
\end{aligned}
$$

$\hat{\beta}_W$ ist KQ-Schätzer im transformierten Modell

Verallgemeinerte KQ-Methode

Das lineare Modell mit allgemeiner Varianzstruktur ist gegeben durch:

$$
\begin{aligned}
Y &= X\beta + \varepsilon && (8.12)\\
\varepsilon &\sim N(0, \sigma^2 V) && (8.13)
\end{aligned}
$$

$V \in R^{n\times n}$: beliebige bekannte Kovarianzmatrix mit vollem Rang

Dann gibt es eine invertierbare Matrix T mit

$$TT' = V, \quad W = T^{-1} \textbf{ Gewichtsmatrix.}$$

Der **verallgemeinerte KQ-Schätzer** ist gegeben durch:

$$
\begin{aligned}
\hat{\beta}_W &:= (X'V^{-1}X)^{-1}X'V^{-1}Y && (8.14)\\
\hat{\sigma}^2 &= (\hat{\varepsilon}'V^{-1}\hat{\varepsilon})/(n-p') && (8.15)
\end{aligned}
$$

Das Modell (8.12)-(8.13) lässt sich wie oben in ein gewöhnliches lineares Modell transformieren:

$$
\begin{aligned}
Y^* &:= WY && (8.16)\\
X^* &:= WX && (8.17)\\
\varepsilon^* &:= W\varepsilon && (8.18)
\end{aligned}
$$

Dann gilt:

$$
\begin{aligned}
Y^* &= X^*\beta + \varepsilon^* && (8.19)\\
\varepsilon^* &\sim N(0, \sigma^2 I) && (8.20)
\end{aligned}
$$

Eigenschaften des verallgemeinerten KQ-Schätzers

Gegeben sei das Modell (8.12) bis (8.13).

Dann gilt:

$$
\begin{aligned}
E(\hat{\beta}_W) &= \beta && (8.21)\\
V(\hat{\beta}_W) &= \sigma^2(X'V^{-1}X)^{-1} && (8.22)
\end{aligned}
$$

Alle Testverfahren und Quadratsummenzerlegungen lassen sich im Modell $Y^* = X^*\beta + \varepsilon^*$ betrachten und damit auf den Fall homogener Varianzen zurückführen.

Allgemeines Gauss-Markov-Theorem

Sei das Modell

$$
\begin{aligned}
Y &= X\beta + \varepsilon, \quad \text{rg } X = p' \\
E(\varepsilon) &= 0 \\
V(\varepsilon) &= \sigma^2 V
\end{aligned}
$$

gegeben.

Dann ist $\hat{\beta}_W$ unter den erwartungstreuen linearen Schätzern derjenige mit der kleinsten Varianz:
$\hat{\beta}_W$ ist BLUE-Schätzer (**b**est **l**inear **u**nbiased **e**stimator).

Beispiele für Varianzstrukturen

- AR(1) (allgemeine Zeitreihenstruktur)
- Longitudinale Daten (Mehrdimensionale Zeitreihen) Blockdiagonale Struktur
- Symmetrische Struktur (gemischte Modelle)

Weitere Schätzstrategien

Im Allgemeinen müssen die Parameter der Varianzstruktur geschätzt werden.
Dazu gibt es verschiedene Verfahren:

- ML
- REML (Restricted Maximum Likelihood)
- Robuste Varianzschätzung mit „Working correlation“

ML und REML-Schätzung I

Sei das Modell

$$
\begin{aligned}
Y &= X\beta + \varepsilon, \quad \text{rg } X = p' \\
E(\varepsilon) &= 0 \\
V(\varepsilon) &= \sigma^2 V(\vartheta)
\end{aligned}
$$

gegeben. Der Parameter(vektor) ϑ ist zu schätzen.
Als Log-Likelihood ergibt sich (von additiven Konstanten abgesehen):

$$
l(\beta, \boldsymbol{\vartheta}) = -\frac{1}{2}\left(\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{Y} - X\beta)' V^{-1}(\boldsymbol{\vartheta})(\boldsymbol{Y} - X\beta)\right) \tag{8.23}
$$

ML und REML-Schätzung II

Ist $\boldsymbol{\vartheta}$ bekannt, so ist der MLE von $\boldsymbol{\beta}$ bedingt auf $\boldsymbol{\vartheta}$ (gewichteter KQ-Schätzer:)

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = \left(X'V(\boldsymbol{\vartheta})^{-1}X\right)^{-1}X'V^{-1}(\boldsymbol{\vartheta})Y. \tag{8.24}$$

Einsetzen liefert die Profil-Log-Likelihood:

$$l(\boldsymbol{\vartheta}) = -\frac{1}{2}\left(\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{Y} - X\beta(\boldsymbol{\vartheta}))'V^{-1}(\boldsymbol{\vartheta})(\boldsymbol{Y} - X\beta(\boldsymbol{\vartheta}))\right) \tag{8.25}$$

ML und REML-Schätzer

Maximieren von (10.22) bezüglich $\boldsymbol{\vartheta}$ liefert ML-Schätzer.
Da dieser nicht erwartungstreu ist, verwendet man häufig den sogenannten restringierten ML-Schätzer:

Dieser maximiert

$$L_R(\boldsymbol{\vartheta}) = l(\boldsymbol{\vartheta}) - \frac{1}{2}ln|X'V(\boldsymbol{\vartheta})^{-1}X| \tag{8.26}$$

Im einfachen linearen Modell entspricht der REML-Schätzer dem erwartungstreuen Schätzer von σ^2.

Inferenz bezüglich von β

Es gilt: $\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ einer multivariaten Normalverteilung mit Erwartungswert $\boldsymbol{\beta}$ und Kovarianzmatrix

$$\text{var}(\hat{\boldsymbol{\beta}}) = (X'V^{-1}X)^{-1} \tag{8.27}$$

Da V unbekannt ist, wird es durch den (RE)ML-Schätzer $V(\hat{\vartheta})$ ersetzt.

Zur Konstruktion von Konfidenzintervallen und entsprechenden Tests nimmt man an, dass β asymptotisch normalverteilt ist. Für spezielle Modelle ist dies bewiesen, aber eine allgemeingültige asymptotische Normalverteilungsaussage ist nicht nachgewiesen.

Da die Varianzmatrix V nur geschätzt wird, werden in der Praxis deshalb häufig approximative t-Tests und entsprechende Konfidenzintervalle benutzt, die die Verteilung von $(\hat{\beta}_j - \beta_j)/\hat{s.e.}(\hat{\beta}_j)$ durch eine t-Verteilung approximieren und die zugehörigen Freiheitsgrade geeignet schätzen.

Beispiele

- Wildzeitreihen:
  Varianzstruktur: Unabhängigkeit der Einzelzeitreihen, aber AR(1)-Struktur für jede einzelne Zeitreihe
- Schuldaten mit Korrelation innerhalb einer Klasse

Korrelationsmatrix Wilddaten

$$
\begin{pmatrix}
\boxed{\begin{matrix} 1 & \rho & \rho^2 & \dots & \rho^k \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{matrix}} & 0 & 0 \\
0 & \boxed{\begin{matrix} 1 & \rho & \rho^2 & \dots & \dots & \rho^k \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{k-1} \\ & & & & & \vdots \\ \rho^2 & & 1 & & & \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \rho^k & \dots & \dots & \dots & \dots & 1 \end{matrix}} & 0 \\
 & & \ddots \\
0 & 0 & \boxed{}
\end{pmatrix}
$$