Lehrstuhl fuer Statistik, Oekonometrie und empirische Wirtschaftsforschung, Universitaet Tuebingen

Modelltransformationen, Heteroskedastie und Autokorrelation

Dr. S. Prohl

27. April 2007/ SoSe 2007

Modellannahmen der Multiplen Regressionen

- A.1 $y_i = \beta_0 + \beta_1 x_i + e_i$.
- A.2 $E(e_i) = 0$.
- A.3 $Cov(e_i, x_i) = 0$.
- A.4 $Var(e_i) = \sigma^2$.
- A.5 $Cov(e_i, e_{i+k}) = 0$, mit $i \neq k$.
- A.6 Keine perfekte Multikollinearitaet.
- A.7 Stoergroessen $e_1, ..., e_i$ sind normalverteilt.
- Unter Annahmen (A.1) - (A.6) liefert KQ-Methode lineare Schaetzfunktionen fuer die Regressionsparameter, die unverzerrt (erwartungstreu) und effizient sind (Best Linear Unbiased Estimators).

Modelltransformationen

Ausgangspunkt: Das Multiple Regressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j + e$$

- lineare Transformationen
- nicht-lineare Transformationen
  - Logarithmieren
  - Polynome
  - Interaktionen

Lineare Transformationen

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e.$
- $y = (\beta_0 + \beta_1 a + \beta_2 b) + \beta_1 (x_1 - a) + \beta_2 (x_2 - b) + e.$
- $ay = a\beta_0 + a\beta_1 x_1 + a\beta_2 x_2 + ae.$
- $ay = a\beta_0 + (a\beta_1 / b)(bx_1) + a\beta_2 x_2 + ae.$
- Die Modelle (1)-(4) sind aequivalent, aber die Interpretation der Parameter ist unterschiedlich.

Interpretation und Anwendung von Transformationen

- Modell (2): sei $a = E(x_1)$ und $b = E(x_2)$. Dann ist die Konstante gleich $E(y)$.
- Modell (3): $\beta_1 = \partial E(ay|ax_1, ax_2)/\partial ax_1$ (z.B., transformierte Einheiten).
- Modell (4): sei $a = \sigma_y{}^{-1}$, und $b = \sigma_{x_1}{}^{-1}$. Dann gilt, dass $\beta_1 \sigma_{x_1}/\sigma_y = \partial(E(y|x_1)/\sigma_y)/\partial x_1/\sigma_{x_1}$ der sogenannte standardisierte Koeffizient. Er gibt uns an, um wieviele Standardabweichungen sich $E(y|x)$ aendert, wenn x_1 um eine Standardabweichung erhoeht wird

Schaetzung des transformierten Modells

Kleinste-Quadrate (KQ) Schaetzer fuer β_1

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Wass passiert wenn wir:

- multiplizieren y mit einem Faktor?
- multiplizieren x mit einem Faktor?
- multiplizieren y und x mit einem Faktor?

Nicht-lineare Transformation: Logarithmus

Taylor-Approximation an der Stelle $z = 1 : Ln(z) \approx -1 + z$.

- Interpretation
- Sei $z = \frac{y_1}{y_0}$ (in der 'Naehe' von 1)
- $\Rightarrow Ln(\frac{y_1}{y_0}) \approx -1 + \frac{y_1}{y_0}$
- $\Leftrightarrow Ln(y_1) - Ln(y_0) \approx \frac{y_1 - y_0}{y_0}$
- $\Delta Ln(y) \approx$ relative Veraenderung in y
- $100 \Delta Ln(y) \approx \% \Delta y$.

Anwendung in der Regressionsanalyse

$Ln(y) = \beta_0 + \beta_1 x + e$
wobei $e = 0$.

- Definiere:
- $Ln(y_1) = \beta_0 + \beta_1(x + \Delta x)$
- $Ln(y_0) = \beta_0 + \beta_1 x$
- $Ln(y_1) - ln(y_0) = \beta_1 \Delta x \approx \frac{y_1 - y_0}{y_0}$
- Approximation ist umso besser, je kleiner $\beta_1 \Delta x$ ist.

Das Problem derRuecktransformation

Das Ausgangsmodell
$Ln(y) = \beta_0 + \beta_1 x + e.$

- Es gilt $E(Ln(y)|x) = \beta_0 + \beta_1 x$.
- Wie ist $E(y|x)$ definiert?
- $E(y|x) = \exp(\beta_0 + \beta_1 x)E[exp(e)]$
- Beachte: $E[\exp(e)] \neq \exp(E(e)) = 1$
- Normalverteilung und Homoskedastie
  $E[exp(e)] = \exp(1/2\sigma^2)$
  Also $E(y|x) = \exp(1/2\sigma^2)\exp(\beta_0 + \beta_1 x)$
- Konsistente Schaetzung:
  $E(\hat{y}|x) = \exp(1/2\hat{\sigma}^2)\exp(\hat{\beta}_0 + \hat{\beta}_1 x)$

Heteroskedastie

Auswirkungen der Heteroskedastie auf die Parameter:

- OLS ist weiterhin unverzerrt: $E(\hat{\beta}_j) = \beta_j$.
- $Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-r_j{}^2)\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$.
- OLS Standardfehler (und die t-Werte) sind verzerrt.
- Falsche Inferenz: Fehler 1.Art bei z-Test $\neq \alpha$.
- OLS ist ineffizient.

Testverfahren fuer Heteroskedastie

Goldfeld und Quandt Test fuer Heteroskedastie:

- $F_{emp} = \frac{\hat{\sigma}_2{}^2}{\hat{\sigma}_1{}^2} = \frac{\sum_{i_1=1}^{N_1} \hat{e}_{i_1}^2 / (N_1 - J - 1)}{\sum_{i_2=1}^{N_2} \hat{e}_{i_2}^2 / (N_2 - J - 1)}$
  - Unter H_0 : (Homoskedastie) $\Rightarrow F_{emp}$ ist F-verteilt.
  - Falls $F_{emp} > F_{theor} \Rightarrow H_0$ wird verworfen.

Testverfahren fuer Heteroskedastie

Breusch-Pagan Test fuer Heteroskedastie:

- Annahme: Heteroskedastie der Form $Var(e_i) = f(x_i, \gamma)$.
- Schaetze das Modell mit OLS und extrahiere die Residuen $\hat{e}_i$
- Schaetze die Hilfsregression $\hat{e}_i^{\,2} = \gamma_0 + \gamma_1 x_{1i} + ... + \gamma_k x_{kj} + u_i$
- Teste $H_0 : \gamma_1 = ... = \gamma_k = 0$ mit F-Test.

Zwei Strategien

- Verwende OLS mit der korrigerten Standardfehler $\Rightarrow$ White (robuste) Sandardfehler.
  - Vorteil: kann fuer jede Form der Heteroskedastie angewendet werden
  - Nachteil: asymptotisch und Effizienzverlust.
- Modifiziere das Schaetzverfahren $\Rightarrow$ Generalised Least Squares (GLS)
  - Beobachtungen mit groesserer Varianz werden weniger stark gewichtet als Beobachtungen mit kleinerer Varianz.
  - der Gewichtungsfaktor ergibt sich aus der vermuteten Form der Heteroskedastie.

White (konsistente) Standardfehler

Im Falle der einfachen Regression

- $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$.
- Es gilt $Var(\hat{\beta}_1|x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 Var(y_i|x_i)}{[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2}$.
- Wie soll die Varianz geschaetzt werden?
- OLS unter Homoskedastieannahme: $Var(\hat{y}_i|x_i) = \hat{\sigma}^2 \Rightarrow$ Vereinfachung.
- White-Varianz: $Var(\hat{y}_i|x_i) = Var(\hat{e}_i|x_i) = \hat{e}_i{}^2$

Generalised Least Squares (Weighted LS)

- Es gibt zwei Faelle:
  - σ_i ist bekannt (zumindest bis auf Konstante)$\Rightarrow$ GLS
  - σ_i ist unbekannt $\Rightarrow$ FGLS
- Fall 1: Das transformierte Modell:
  $\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_1}{\sigma_i} + \ldots + \beta_j \frac{x_j}{\sigma_i} + \frac{e_i}{\sigma_i}$.
- Es gilt: $Var(\frac{e_i}{\sigma_i}) = {\sigma_i}^{-2} Var(e_i) = 1$.
- $\Rightarrow$ Homoskedastie ist erfuellt.
- Fall 2: Wie im Fall 1, nur werden die Variable mit $\hat{\sigma}_i$ gewichtet. $\hat{\sigma}_i$ muss auf Vorstufe geschaetzt werden, z.B., nach Breusch-Pagan Methode.
- Nachteil: Feasible GLS ist nicht unverzerrt, nur konsistent.