Kapitel 11

Heteroskedastizität

Der Sachverhalt

Das (wahres) Modell ist $y = X\beta + u$, Dimension von X: $n \times k$

$$\boxed{\text{Annahme A6: } \mathrm{Var}\{u\} = \sigma^2 I}$$

Annahme 6 sagt, dass die Varianz des Fehles konstant ist (Homoskedastizität):

$$\mathrm{Var}\{u_t\} = \sigma^2, \quad t = 1,\ldots,n$$

In der Realität trifft diese Annahme nicht immer zu. Man spricht dann von Heteroskedastizität.

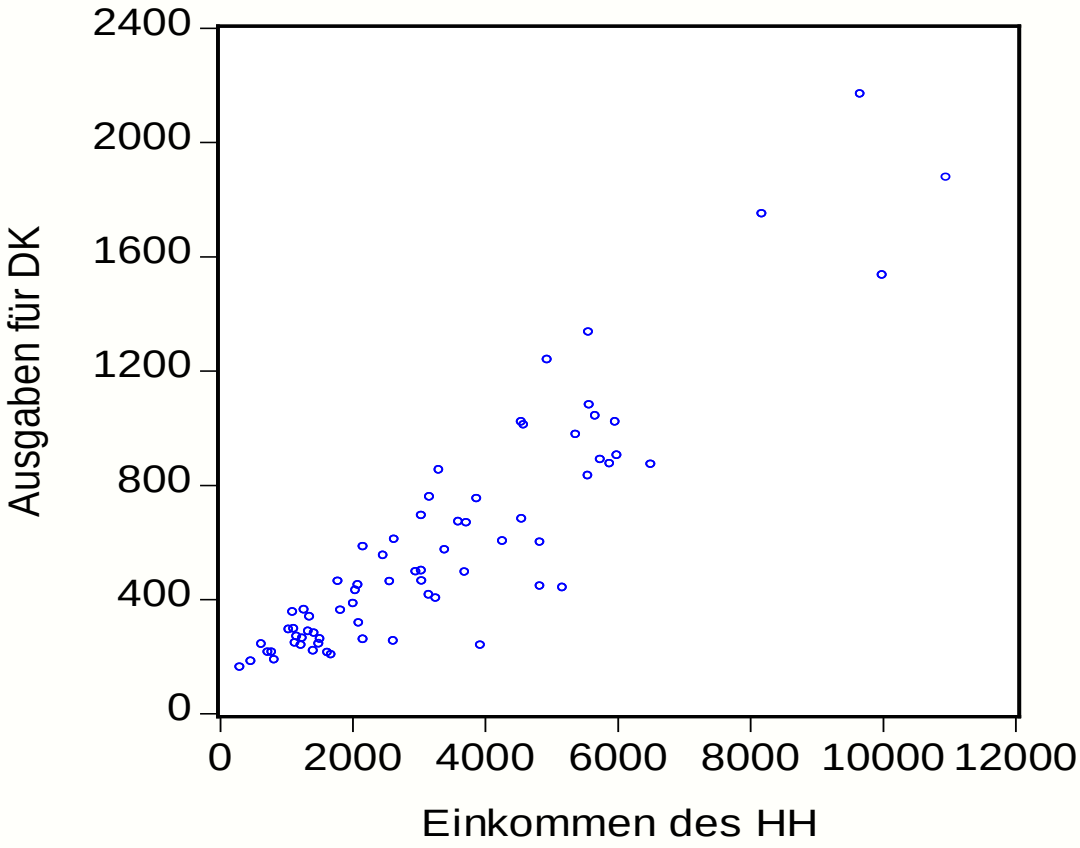
$$\mathrm{Var}\{u\} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2) = \sigma^2\Omega = \sigma^2\,\mathrm{diag}(\omega_1, \ldots, \omega_n)$$

Fragestellungen:

- Konsequenzen von Heteroskedastizität
- Möglichkeiten zum Identifizieren von Heteroskedastizität
- Alternative Schätzverfahren bei Vorliegen von Heteroskedastizität

Beispiel: Dauerhafter Konsum

70 Haushalte (HH): Monatliches HH-Einkommen und Ausgaben für Güter des dauerhaften Konsums (Querschnittsdaten)



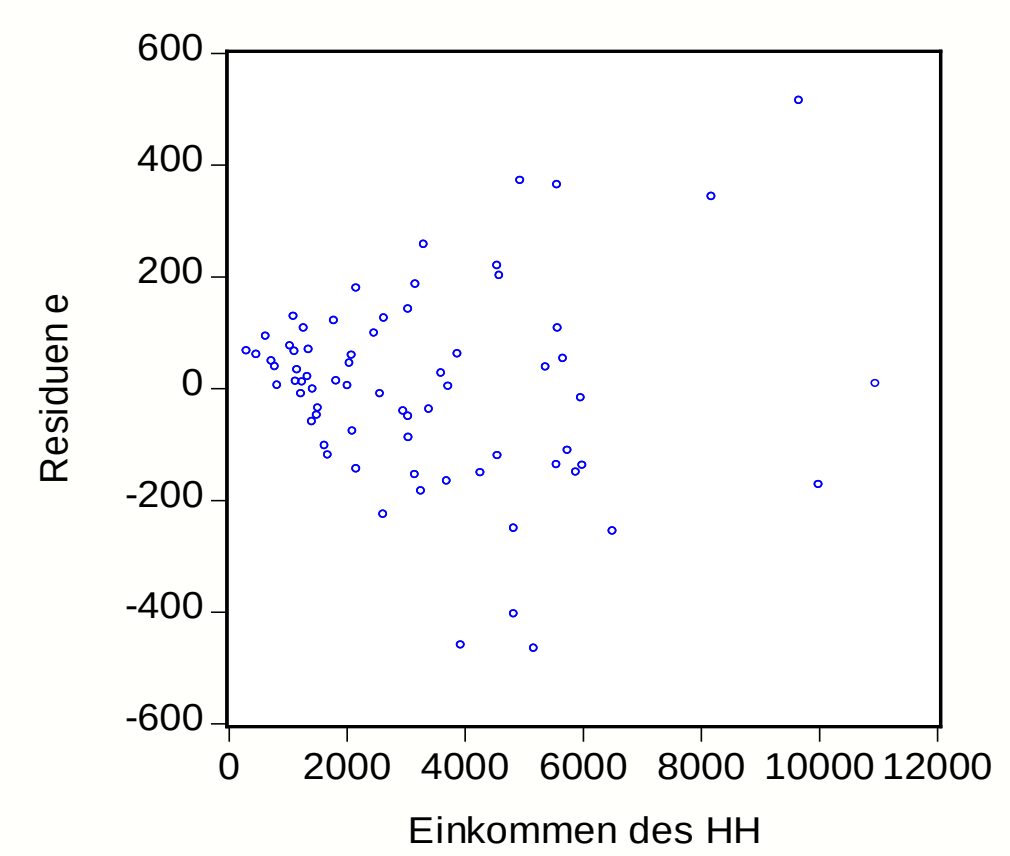
Bsp: Dauerhafter Konsum, Forts.

Plot der Residuen
$e = y - \hat{y}$ aus

$\hat{Y} = 44.18 + 0.17\,X$

X: Monatliches HH-Einkommen
Y: Ausgaben des dauerhaften Konsums

Je größer das Einkommen, umso mehr streuen die Residuen!



Heteroskedastizität: Typische Situationen

Heteroskedastizität tritt typischerweise auf bei

- Querschnittserhebungen, etwa von Haushaltsdaten. Z.B.: Die Heterogenität der HHe steigt mit dem Eink
- Querschnittserhebung über verschiedene Regionen (Regional/Geographische Daten) Z.B.: Unterschied Stadt/Land
- Daten mit geometrischem/exponentiellem Wachstum, wenn der Trend falsch spezifiziert ist (Box-Cox-Transformation zur Varianzstabilisierung) [$y \rightarrow \log(y)$]
- Modell mit stochastischen Regressionskoeffizienten
- Finanzmarkt-Daten wie Wechselkurse oder Renditen von Wertpapieren (bedingte Heteroskedastizität)

Beispiel: Stochastische Regressionskoeffizienten

Betrachten wir das Modell

$$Y_t = \alpha + \beta_t X_t + u_t, \qquad \beta_t = \beta + \varepsilon_t$$

Zwar ist $E(\beta_t) = \beta$, aber β_t ist hier nicht konstant.

ε_t ist eine für alle t identisch und unabhängig verteilte Variable (i.i.d.) mit Mittel 0 und Varianz σ_e^2.

Das Modell kann geschrieben werden als

$$Y_t = \alpha + \beta X_t + v_t \quad \text{mit Störgrößen } v_t = u_t + X_t \varepsilon_t$$

Damit ist die Varianz der v_t nicht konstant. (ε_t, u_t unkorr)

$$\mathrm{Var}\{v_t\} = \sigma_u^2 + X_t^2 \sigma_e^2 = f(X_t^2)$$

OLS-Schätzer b unter Heteroskedastizität

- Für b gilt wie unter Homoskedastizität

  $$b = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u$$

  Mit $E\{u\} = 0$, ist b erwartungstreu: $E(b) = \beta$.

- Die Varianz von b nach der OLS Formel ist

  $$\sigma^2 (X'X)^{-1}$$

- Die korrekte Varianz von b ist allerdings mit

  $$\mathrm{Var}\{u\} = \sigma^2 \Omega = \sigma^2 \,\mathrm{diag}(\omega_1, \ldots, \omega_n)$$

  $$\mathrm{Var}\{b\} = E[(b-\beta)(b-\beta)'] = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

- Die Varianzen nach der OLS Formel sind verzerrt.

Heteroskedastizität und OLS

- Die OLS-Schätzer b für β sind
  - erwartungstreu
  - konsistent
  - haben die Kovarianzmatrix

$$\mathrm{Var}\{b\} = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

  - keine effizienten Schätzer (Es gibt bessere: WLS.)
  - unter allgemein erfüllten Bedingungen asymptotisch normalverteilt
- Der Schätzer $s^2 = e'e/(n-k)$ für die Varianz der Störgrößen σ^2 ist verzerrt
  (e: Vektor der OLS-Residuen)

Heteroskedastizität und OLS

Bei Anwendung von OLS, d.h. bei Vernachlässigung Vorliegender Heteroskedastizität, sind

- die aus $\sigma^2(X'X)^{-1}$ bestimmten Standardfehler verzerrt (nach oben und nach unten möglich), daher liefern
- die zugehörigen *t*- und *F*-Tests irreführende Ergebnisse!

Tests auf Heteroskedastizität

Daher ist es wichtig festzustellen, ob Heteroskedastizität vorliegt. Wir untersuchen dazu die Residuen.

Tests auf Basis der Residuen

- Goldfeld-Quandt-Test $\sigma_1^2 = \sigma_2^2$
- Breusch-Pagan-Test $\sigma_t^2 = \sigma^2(z_t' \delta)$
- Glejser-Test $\sigma_t = \sigma(z_t' \delta)$
- Harvey-Test $\log(\sigma_t^2) = z_t' \delta$ bzw. $\sigma_t^2 = \exp(z_t' \delta)$
- White-Test $\sigma_t^2 = \sigma^2(\ldots)$

Jeder dieser Tests unterstellt ein bestimmtes Modell für die Heteroskedastizität.

Goldfeld-Quandt-Test

Nullhypothese: Homoskedastizität

Alternative: Es liegen zwei Regime mit σ_1^2 und σ_2^2 als Varianz der Störgrößen vor. Die Zugehörigkeit zu Regimen wird durch die Variable Z angezeigt.

Beispiel (Z ist eine Regime-Dummy): (vgl. CWS09 F.14)

$y_1 = X_1\beta_1 + u_1$, $\mathrm{Var}\{u_1\} = \sigma_1^2 I_{n1}$ … Regime 1 (n_1 Beob.)

$y_2 = X_2\beta_2 + u_2$, $\mathrm{Var}\{u_2\} = \sigma_2^2 I_{n2}$ … Regime 2 (n_2 Beob.)

Nullhypothese: $\sigma_1^2 = \sigma_2^2$

F-Test:

$$F = \frac{S_1}{S_2}\frac{n_2 - k}{n_1 - k}$$

F vlt mit (n_1-k) und (n_2-k) FGen

S_i: Summe der quadrierten Residuen für i-tes Regime

Goldfeld-Quandt-Test, Forts.

Das Testverfahren läuft in folgenden Schritten ab:

1. Sortieren der Beobacht. nach steigenden Werten von Z
2. Entfernen von $2c$ Beobachtungen in der Mitte
3. Getrennte OLS-Anpassung an die ersten n_1 und die letzten n_2 Beobachtungen und Bestimmung der OLS-Schätzer b_i und der Summen der quadrierten Residuen S_i (i = 1,2)
4. Berechnen der Teststatistik F. Sie ist unter H_0 exakt oder näherungsweise F-verteilt mit $(n_1\text{-}c\text{-}k)$ und $(n_2\text{-}c\text{-}k)$ Freiheitsgraden. (Wir testen i.A. 1-seitig $s_1 > s_2$)

$$F = \frac{S_1}{S_2}\frac{n_2 - c - k}{n_1 - c - k}$$

Bsp: Dauerhafter Konsum, Forts.

Testen auf zwei Regime: (1) $X > 4000$ und (2) $X < 4000$.

A: beide Regime gemeinsam ($n = 70$): $\hat{Y} = 44.18 + 0.17\,X$

$S = 2{,}094.511$, $s = [S/(70\text{-}2)]^{0.5} = 175.5$

B(1): $X > 4000$ ($n_1 = 22$): $\hat{Y} = -155.34 + 0.20\,X$,

$S_1 = 1{,}331.777$, $s_1 = [S_1/(22\text{-}2)]^{0.5} = 258$

B(2): $X < 4000$ ($n_2 = 48$): $\hat{Y} = 119.71 + 0.13\,X$,

$S_2 = 627.648$, $s_2 = [S_2/(48\text{-}2)]^{0.5} = 117$

Bsp: Dauerhafter Konsum, Forts.

F-Teststatistik:

$$F = \frac{1331777}{627648} \frac{48-2}{22-2} = 4.88$$

$n_1 = 22$, $S_1 = 1{,}331.777$

$n_2 = 48$, $S_2 = 627.648$

p-Wert: 0.000004; Die Nullhypothese wird verworfen.

Implizit gilt die Annahme: $\beta_1 = \beta_2$!

Breusch-Pagan-Test, $H_0 : \sigma_t^2 = \sigma^2$

Modell für Heteroskedastizität mit f(..) > 0 und f(0)=1

$$\sigma_t^2 = \sigma^2 f(z_t' \delta)$$

mit *p*-Vektoren z_t und δ, Interzept δ_1, *p*-1 Variablen $Z_2, \ldots, Z_p$

Die Fehlervarianz wird durch die Variablen Z_j erklärt.

$$\sigma_t^2 = \delta_1 + \delta_2 Z_{t2} + \ldots + \delta_p Z_{tp}$$

Wir approximieren σ_t^2 durch die Residuen aus

$Y_t = \alpha + \beta X_t + u_t$, $\sigma_t^2 \approx e_t^2$ ($=\hat{u}_t^2$) und schätzen

$$e_t^2 = \delta_1 + \delta_2 Z_{t2} + \ldots + \delta_p Z_{tp} + v_t$$

H_0: $\delta_2 = \ldots = \delta_p = 0$, kein Z_j kann die Varianz erklären

H_A: ein $\delta_j \neq 0$, für $j > 1$.

Breusch-Pagan-Test, Forts.

Als Teststatistik können wir

den F-Test auf Signifikanz der Regression mit (p-1) und (n –p-1) Freiheitsgraden, F(p-1, n-p-1) verwenden.

(Der Test ist nur approximativ, da die Fehler v_t nicht normal verteilt sind.)

oder

dessen (asymptotische) Approximation durch die Chi-Quadrat-Verteilung. (Lagrange-Multiplikator-Test)

$LM(H) = nR_e^2 \sim \chi^2(p-1)$

Bsp: Dauerhafter Konsum, Forts.

Breusch-Pagan-Test, z.B. ob $\sigma_t^2 = \delta_1 + \delta_2 X_t$

1. Anpassen der Konsumfunktion: $\hat{Y}_t = 44.18 + 0.17\, X_t$
2. Berechnen von $e_t = Y_t - (44.18 + 0.17\, X_t)$
3. Anpassen der quadr. Residuen: $e_t^2 = -6385 + 10.9 X_t$

$R_e^2 = 0.2143$, $n = 70$, LM(H) = 70 (0.2143) = 15.0

p-Wert: 0.0001 aus der Chi-Quadrat-Verteilung mit (2-1)=1 Freiheitsgrad

Die Nullhypothese $\delta_2=0$ wird verworfen.

Bsp: Dauerhafter Konsum, Forts.

In diesem einfachen Modell können wir auch mit Hilfe des t-Tests auf δ_2 =0 entscheiden.

t-Test für δ_2: d_2=10.9, t = 4.3, *p*-Wert = 0.0001

Die Nullhypothese δ_2=0 kann nicht beibehalten werden.

Glejser-Test, $H_0 : \sigma_t^2 = \sigma^2$

Ein **Modell** für Heteroskedastizität mit f(..) > 0 und f(0)=1.

$$\sigma_t = \sigma f\,(z_t'\delta)$$

Mit (*p x 1*)-Vektoren z_t und δ: Interzept δ_1, *p*-1 Variablen $Z_2, \ldots, Z_p$

Die Fehlerstandardabweichung wird durch die Variablen Z_j erklärt.

$$\sigma_t = \delta_1 + \delta_2 Z_{t2} + \ldots + \delta_p Z_{tp}$$

Wir approximieren σ_t durch den Betrag der Residuen aus

$Y_t = \alpha + \beta X_t + u_t$, $\sigma_t \approx |e_t|$ und schätzen

$$|e_t| = \delta_1 + \delta_2 Z_{t2} + \ldots + \delta_p Z_{tp} + v_t$$

Glejser-Test, Forts.

H_0: $\delta_2 = \ldots = \delta_p = 0$, kein Z_j kann die Standardabweichung erklären.

H_A: ein $\delta_j \neq 0$, für $j > 1$.

Als Teststatistik können wir wie beim Breusch-Pagan-Test

den F-Test auf Signifikanz der Regression mit (p-1) und (n –p-1) Freiheitsgraden, F(p-1, n-p-1) verwenden.

oder

dessen (asymptotische) Approximation durch die Chi-Quadrat-Verteilung. (Lagrange-Multiplikator-Test)

$LM(H) = nR_e^2 \sim \chi^2(p\text{-}1)$

Harvey-Test, $H_0 : \log(\sigma_t^2) = \text{const}$

Ein **Modell** für Heteroskedastizität mit f(..) > 0 und f(0)=1.

$$\log(\sigma_t^2) = f(z_t'\delta)$$

Mit (*p x 1*)-Vektoren z_t und δ: Interzept δ_1, *p*-1 Variablen $Z_2, \ldots, Z_p$

Der log der Varianz wird durch die Variablen Z_j erklärt.

$$\log(\sigma_t^2) = \delta_1 + \delta_2 Z_{t2} + \ldots + \delta_p Z_{tp}$$

Approximation von σ_t^2 durch e_t^2, Nullhypothese, Alternative, wie auch die Testverteilungen sind analog zum Breusch-Pagan-Test.

White-Test, $H_0 : \sigma_t^2 = \sigma^2$

Test der Nullhypothese:

$$H_0: \quad \sigma_t^2 = \sigma^2 \quad \text{(konstant) für alle } t$$

gegen die unspezifizierte Alternative, H_0 sei unrichtig

Teststatistik:

$$W = n\, R_e^2$$

R_e^2 ist das Bestimmtheitsmaß aus der Hilfsregression der quadrierten Residuen e_t^2 auf *alle* Regressoren des Modells, ihrer Quadrate und gegebenenfalls auch auf ihrer Produkte.

W ist asymptotisch Chi-Quadrat-verteilt mit der Anzahl der geschätzten Koeffizienten minus eins als Freiheitsgrade.

White-Test

Vor-/Nachteile:

(1) Technische Überprüfung der Modelleigenschaften

(2) Man muss kein Verhaltensmodell für die Varianzen spezifizieren.

Bsp: Dauerhafter Konsum, Forts.

White-Test, ob $\sigma_t^2 = \sigma^2$ für alle t

1. Anpassen der Konsumfunktion: $\hat{Y} = 44.18 + 0.17\,X$
2. Anpassen der Res: $e^2 = -18185 + 18.4\,X - 0.0008\,X^2$

$R_e^2 = 0.226$, $W = 70\,(0.226) = 15.82$,

p-Wert: 0.0004 aus der Chi-Qudrat-Vtlg mit (3-1) = 2 FGen.

Die Nullhypothese konstanter Fehlervarianz wird verworfen.

Inferenz bei Heteroskedastizität

Kovarianzmatrix des OLS-Schätzers b ist

$$\mathrm{Var}\{b\} = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Die Verwendung von $\sigma^2 (X'X)^{-1}$ führt zu verzerrten Testergebnissen.

Falsche Standardfehler können vermieden werden durch

1. Verwendung der korrekten Varianzen (Heteroskedastizitäts-konsistente Kovarianzmatrizen)
2. Transformation des Modells, sodass die Störgrößen homoskedastisch werden.

Heteroskedastie-konsistente Standardfehler (White)

Die theoretische Kovarianzmatrix für b ist

$$\mathrm{Var}\{b\} = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Die *heteroskedasticity consistent* geschätzte Kovarianzmatrix nach White (Ω diagonal, x_t' ist Zeilenvektor von X)

$$\frac{1}{n-k}(X'X)^{-1}\left(\sum_t e_t^2 x_t x_t^{'}\right)(X'X)^{-1}$$

(Es werden die e_t^2 als Näherung für σ_t^2 verwendet.)

Sie liefert die passenden „White-Standardfehler“ für die b_i

Simulationen zeigen, dass die White-Standardfehler die tatsächlichen Standardfehler unterschätzen.

Heteroskedastie-konsistente Standardfehler (White)

EViews bietet White-Standardfehler als Option bei der OLS-Schätzung.

Bsp: Dauerhafter Konsum, Forts.

Wir schätzen die Konsumfunktion mit OLS:

$$\hat{Y} = 44.18 + 0.17\,X$$

einmal mit

- OLS-Standardfehler: Der Standardfehler des Koeffizienten von X beträgt 0.009.
- White-/Heteroskedastie-konsistente Standardfehler: 0.012

Der nicht korrigierte Standardfehler (OLS) unterschätzt um mehr als 30%.

Variablen-Transformation

Ist die funktionale Form der Abhängigkeit der σ_t^2 bekannt, oder haben wir ein gutes Modell dafür, so wählen wir eine geeignete Transformation des Modells.

Beispiel:

- Das Modell sei $Y_t = \alpha + \beta X_t + u_t$ mit heteroskedast. Störgrößen: $\sigma_t^2 = \delta Z_t^2$ für $t = 1, \ldots, n$.
- Der transformierte Fehler $v_t = u_t/Z_t$ ist nun

  $\mathrm{Var}\{v_t\} = \mathrm{Var}\{u_t\}/Z_t^2 = \sigma^2 \; (= \delta^2)$

  homoskedastisch.
- In einem Schritt zusammengefasst bedeutet das die Transformation des Modells $Y_t = \alpha + \beta X_t + u_t$ zu

$$\frac{Y_t}{Z_t} = \alpha \frac{1}{Z_t} + \beta \frac{X_t}{Z_t} + v_t$$

Variablen-Transformation, Forts.

Wir schätzen nun das transformierte Modell. Es hat die Eigenschaften, dass

- die Fehler (wie gewünscht) homoskedastisch sind, und
- die Koeffizienten mit denen im ursprünglichen Modell übereinstimmen, also direkt ablesbar sind.
- Allerdings: Die R^2 der Schätzungen mit und ohne Gewichtung sind nicht vergleichbar, da verschiedene *y*-Variable modelliert werden!
- In EViews kann man die Option gewichtete LS-Schätzung wählen.

Gewichtete LS-Schätzer

Die Schätzung des transformierten Modells kann als gewichteter LS-Schätzer interpretiert werden.

Beispiel: Modell $Y_t = a + b\,X_t + u_t$ mit $\mathrm{Var}\{u_t\} = \sigma_t^2 = \delta\, Z_t^2$

- Wir minimieren im nicht-transformierten Modell

  $\sum_t (Y_t - a - b\,X_t)^2$ $\qquad u_t = Y_t - a - b\,X_t$

- und im transformierten Modell

  $\sum_t w_t\,(Y_t - a - b\,X_t)^2$ $\qquad v_t = (1/Z_t)(Y_t - a - b\,X_t)$

  mit den *Gewichten* $w_t = (1/Z_t^2)$

[$w_t = 1/\omega_t$ mit $\omega_t = Z_t^2$ aus Ω. $\mathrm{Var}\{u\} = \delta\,\Omega = \delta\,\mathrm{diag}(Z_1^2, \ldots, Z_n^2)$]

Diese gewichtete LS-Schätzung ist ein Spezialfall der GLS-Schätzung, *generalized* LS-Schätzung.

Bsp: Dauerhafter Konsum, Forts.

Unser Modell ist $Y_t = a + b X_t + u_t$ mit $\sigma_t^2 = \delta_1 X_t$.

Die zugehörige Transformation des Modells ist

$$(1/\sqrt{X_t})$$

Das entspricht Gewichten $w_t = 1/X_t$

Die angepasste Funktion ist

$$\frac{\hat{Y}_t}{\sqrt{X_t}} = \frac{90.42}{\sqrt{X_t}} + 0.153\sqrt{X_t}$$

- Eviews: LS -> Options -> Weight type: Variance , Series: X_t

GLS, generalised least squares

Angenommen unser Modell liegt in Matrixform vor

$$y = X\beta + u \quad \text{mit} \quad \mathrm{Var}\{u\} = \mathrm{E}(uu') = \mathrm{diag}(\sigma_t^2) = \Omega$$

u ist heteroskedastisch.
y ist (n x 1), X ist (n x k), β ist (k x 1) und Ω ist (n x n).

Die Gewichte w_t (von oben) sind

$$w_t = 1/\sigma_t^2.$$

Wir multiplizieren die Gleichung in jedem Zeitpunkt mit

$$w_t^{1/2} = 1/\sigma_t$$

$$y_t/\sigma_t = x'_t/\sigma_t\, \beta + u_t/\sigma_t \quad \text{bzw.} \quad y_t^* = x_t^{*\prime}\, \beta + v_t$$

GLS, generalised least squares

In Matrixschreibweise ist das äquivalent zur Multiplikation mit $\Omega^{-1/2} = \mathrm{diag}(1/\sigma_t)$ von links

$$\Omega^{-1/2} y = \Omega^{-1/2} X \beta + \Omega^{-1/2} u \quad \text{bzw.} \quad y^* = X^* \beta + v$$

v ist homoskedastisch. Der OLS-Schätzer für das transformierte Modell (*) ist einfach

$$b^* = (X^{*\prime} X^*)^{-1} X^{*\prime} y^* \quad \text{mit} \quad \mathrm{Var}(b^*) = (X^{*\prime} X^*)^{-1}$$

OLS liefert den besten unverzerrten Schätzer für β, b^*, da die v_t homoskedastisch sind.

GLS-Schätzer

b^* in den beobachteten Variablen y und X angeschrieben gibt den GLS-Schätzer, $b^* = b_{GLS}$ des ursprünglichen Modells

$$y = X\beta + u$$

$b_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ und

$\mathrm{Var}(b_{GLS}) = (X'\Omega^{-1}X)^{-1}$

Die Elemente von Ω sind allerdings (noch) Parameter der Grundgesamtheit und daher unbekannt.

FGLS-Schätzer

Den FGLS (feasible GLS) Schätzer (feasible = ausführbar) erhält man, wenn für die Elemente aus Ω Werte aus der Stichprobe gewählt werden.

Da wir nur Varianzen benötigen (Hauptdiagonalelemente) setzen wir wie oben (siehe White heteroscedasticity consistent Standardfehler und 2-stufige Schätzung) für σ_t^2 die quadrierten Residuen aus der OLS Schätzung (1.Stufe) ein.

$$\sigma_t^2 \approx e_t^2$$