Materialien zur Vorlesung

Einführung in die Ökonometrie

Sommersemester 2005

Prof. Dr. Klaus Neusser

Universität Bern

Inhalt

1 Einführung

Wozu Ökonometrie?

- Datenbeschreibung über den Zustand der Wirtschaft, Einkommensverteilung
- Quantifizierung von ökonomischen Zusammenhängen. Dies ist umso wichtiger, da die ökonomische Theorie in vielen Fällen nicht einmal qualitative Aussagen machen kann (z.B. Einkommens- versus Substitutionseffekt, Zusammenhang Menge und Preis, Wirtschaftswachstum und Arbeitslosigkeit, Absatz und Werbung, Zinssätze und Börsenkurse, Steuern, etc.
- Testen ökonomischer Theorien
- Prognosen: Inflation, Wirtschaftswachstum, Zinssätze, Verkäufe, Energiebedarf, etc.
- Politikberatung: Steuern, Staatsausgaben; Abschätzung des Marktpotentials

Ziel der Ökonometrie ist die Konstruktion ökonometrischer Modelle. Mittel der Ökonometrie:

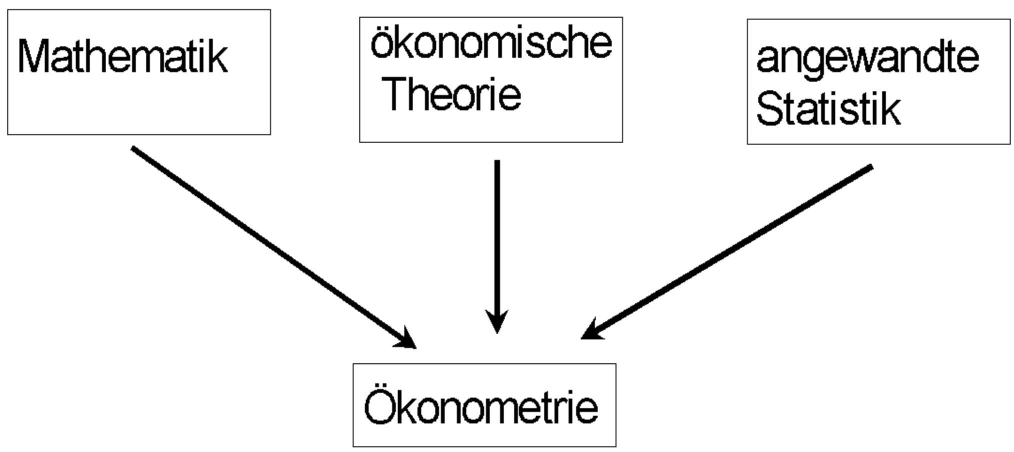


Abbildung aus Ragnar Frisch (1933): Gründung der „Econometric Society“; Zeitschrift „Econometrica“

1.1 Einige Lehrbücher

Das Buch zur Vorlesung:

Wooldridge, J. M. (2003), *Introductory Econometrics*, Ohio, South-Western.

Weitere Lehrbücher:

Amemiya, T. (1994): *Introduction to Statistics and Econometrics*, Cambridge, Massachusetts: Harvard University Press.

Davidson, R. und J.G. MacKinnon (1993): *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Greene, W.H. (1993): Econometrics, 2nd Edition, Upper Saddle River: Prentice-Hall and New York: Macmillan.

Greene, W.H. (1997): Econometrics, 3rd Edition, Upper Saddle River: Prentice-Hall.

Greene, W.H. (2000): Econometrics, 4th Edition, Upper Saddle River: Prentice-Hall.

Kennedy, P. (1998): *A Guide to Econometrics,* Oxford: Oxford University Press.

Maddala, G.S. (1992): *Introduction to Econometrics*, London: Macmillan.

Beispiele aus:

Berndt, E.R. (1991): *The Practice of Econometrics*, New York: Addison Wesley.

1.2 Einige ökonometrische Programmpakete

In den Übungen werden wir mit der Software EViews arbeiten. EViews hat ein sehr bequeme graphische Oberfläche und erlaubt die Schätzung aller hier behandelten Modelle. Wir verwenden EViews auch in den weiterführenden Vorlesungen Zeitreihenanalyse und Mikroökonometrie.

Weiter Statistik Programme:

EVIEWS	allgemeines Ökonometriepaket
STATA	allgemeines Ökonometriepaket
Micro-TSP	allgemeines Ökonometriepaket
TSP	allgemeines Ökonometriepaket
SHAZAM	allgemeines Ökonometriepaket
SAS	allgemeines Statistikpaket
SPSS	allgemeines Statistikpaket für Sozialwissenschaften
RATS	Zeitreihenanalyse
LIMDEP	qualitative Variable
GAUSS	matrix-orientierte Sprache
MATLAB	matrix-orientierte Sprache

2 Das klassische lineare Regressionsmodell

Literatur: Wooldridge (2003), Kapitel 1 oder Greene (1993), Kap. 5.

2.1 Ein Beispiel: capital asset pricing model CAPM

$$R_i = R_F + \beta_i(R_M - R_F) + \varepsilon_i$$

R_i Ertrag der i-ten Aktie

R_F Ertrag der risikolosen Anlage

R_M Ertrag des Marktes

β_i Beta der i-ten Aktie (Mass für das Risiko einer Anlage); bestimmbares Risiko (deterministisch)

ε_i das idiosynkratische Risiko (das unsystematische Risiko); unbestimmbares Risiko (stochastisch)

Das ökonometrische Modell

$$R_i - R_F = \alpha_i + \beta_i(R_M - R_F) + \varepsilon_i$$

Hypothesen: $\alpha_i = 0$ oder $\beta_i = 1$

$\beta_i = 1$ bedeutet, dass die Aktie i das gleiche Risiko wie der Markt hat.

$\alpha_i > 0$ bedeutet, dass die Renditen der Aktie i ‚überdurchschnittlich' sind.

2.2 Das statistische Modell

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

Bezug zum CAPM:

y_t Beobachtung von R_i - R_F zum Zeitpunkt t

x_t Beobachtung von R_M - R_F zum Zeitpunkt t

Ziel ist die Schätzung von α und β für die i-te Aktie

Stichprobe:

Beobachtungen (y_t, x_t) für t = 1, . . ., T

Die Störungen (ε_t) sind nicht beobachtbar

y_t	x_t
Prädiktand	Prädiktor
Regressand	Regressor
erklärte Variable	erklärende Variable
abhängige Variable	unabhängige Variable
Effekt	Ursache
endogene Variable	exogene Variable
Zielgrösse	Kontrollvariable

Annahmen:

1. Linearer Zusammenhang zwischen Prädiktand und Prädiktor
2. $E\varepsilon_t = 0$ für alle t; Erwartungswert (Mittelwert) der Störungen ist Null
3. $V\varepsilon_t = \sigma^2$ für alle t; konstante Varianz, (Homoskedastizität)
4. $E\varepsilon_t\varepsilon_s = 0$ für alle $t \neq s$; Störungen sind unkorreliert in der Zeit, (unkorrelierte Störungen)
5. $Ex_t\varepsilon_t = 0$ für alle t; keine Korrelation zwischen Regressoren und Störungen, (Orthogonalität von Regressoren und Störungen)
6. x_t ist deterministisch und $\frac{s_x^2}{T-1} = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})^2 > 0$

   (keine Multikollinearität)
7. $\varepsilon_t \sim N(0, \sigma^2)$ (Normalität)

2.3 Schätzung: die Methode der kleinsten Quadrate (OLS)

Literatur: Wooldridge (2003), Anhang C.6 oder Greene (1997), Kapitel Example 6.7, p. 238.

Kleinstquadrateproblem („ordinary least squares“ OLS):

$$S(\alpha,\beta) = \sum_{t=1}^{T}\varepsilon_t^2 = \sum_{t=1}^{T}(y_t - \alpha - \beta x_t)^2 \rightarrow \min_{\alpha,\beta}$$

Lösung:

$$\hat{\beta} = \frac{s_{xy}}{s_x^2} \qquad \hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

mit

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \text{ und } \overline{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$$

$$s_{xy} = \sum_{t=1}^{T}(y_t - \overline{y})(x_t - \overline{x}) = \sum_{t=1}^{T}(y_t - \overline{y})x_t = \sum_{t=1}^{T} y_t x_t - T\overline{y}\overline{x}$$

$$s_x^2 = s_{xx} = \sum_{t=1}^{T}(x_t - \overline{x})^2 = \sum_{t=1}^{T}(x_t - \overline{x})x_t = \sum_{t=1}^{T} x_t^2 - T\overline{x}^2$$

DEFINITION:

Kleinstquadrateprädiktor: $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$

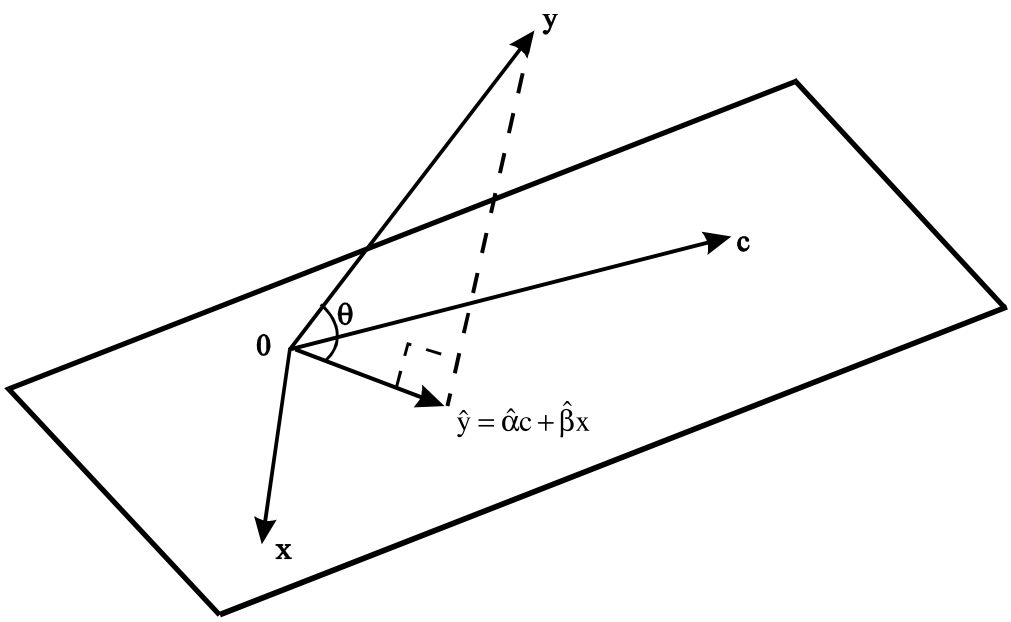
Kleinstquadrateresiduum: $\hat{\varepsilon}_t = e_t = y_t - \hat{y}_t$

Die Kleinstquadrateresiduen erfüllen die Orthogonalitätsbedingungen:

$$\sum \hat{\varepsilon}_t = 0$$

$$\sum \hat{\varepsilon}_t x_t = 0$$

Geometrische Interpretation des Kleinstquadrateschätzers:



Der Punkt $y = (y_1,...,y_T)'$ wird auf die durch die Konstante $c = (1,...,1)'$ und den Regressor $x = (x_1,...,x_T)$ aufgespannte Ebene projeziert. Dadurch bildet sich in $\hat{y}$ ein rechter Winkel. Der Abstand zwischen y und $\hat{y}$ ist dabei im Sinne der euklidischen Metrik minimal.

2.4 Güte der Anpassung: das Bestimmtheitsmass R^2

Wie gut ist die Regressionsgerade? („measure of fit“, Bestimmtheitsmass)

$\sum\left(y_t - \hat{\alpha} - \hat{\beta}x_t\right)^2 = \sum \hat{\varepsilon}_t^2$ kann beliebig skaliert werden.

„total sum of squares“: $SST = \sum\left(y_t - \overline{y}\right)^2 = s_y^2$

Es gilt: $y_t = \hat{y}_t + \hat{\varepsilon}_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\varepsilon}_t = \bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_t + \hat{\varepsilon}_t$

$$y_t - \bar{y} = \hat{\beta}(x_t - \bar{x}) + \hat{\varepsilon}_t$$

Durch quadrieren unter Berücksichtigung der Orthogonalitäts-bedingungen sowie anschliessendes aufsummieren ergibt sich:

$$s_y^2 = \underbrace{\sum (y_t - \bar{y})^2}_{\substack{\text{total sum of squares} \\ \text{SST}}} = \underbrace{\hat{\beta}^2 \sum (x_t - \bar{x})^2}_{\substack{\text{regression sum of} \\ \text{squares SSR}}} + \underbrace{\sum \hat{\varepsilon}_t^2}_{\substack{\text{error sum of} \\ \text{squares SSE}}} = \hat{\beta}^2 s_x^2 + \text{SSE}$$

Damit lässt sich das *Bestimmtheitsmass* R^2 folgendermassen definieren:

$$\boxed{R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\hat{\beta}^2 s_x^2}{s_y^2} = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum \hat{\varepsilon}_t^2}{s_y^2} \qquad \text{mit } 0 \le R^2 \le 1}$$

Geometrisch gesehen ist das Bestimmtheitsmass der Cosinus des Winkels θ zum Quadrat: $R^2 = \cos^2\theta$. Für θ gleich null ist der Cosinus eins, daher die Anpassung maximal. Für θ gleich $\pi/2$ (90 Grad) ist der Cosinus 0, die Regression erklärt nichts.

Korrelationskoeffizient und Bestimmtheitsmass

Der Korrelationskoeffizient zwischen x und y ist definiert als:

$$-1 \le \rho_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \, \sigma_y} \le 1$$

Das empirische Äquivalent dieses Korrelationskoeffizienten beträgt:

$$r_{x,y} = \frac{s_{xy}}{s_x \; s_y} = \hat{\beta}\,\frac{s_x}{s_y}$$

Die Steigung der Regressionsgeraden ($\hat{\beta}$) hat das selbe Vorzeichen wie r_{xy}. Das Bestimmtheitsmass entspricht dem quadrierten Korrelationskoeffizienten:

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\beta}^2 s_x^2}{s_y^2} = r_{xy}^2 \Rightarrow 0 \le R^2 \le 1$$

Die Beziehung des empirischen Äquivalents des Korrelationskoeffizienten zwischen y_t und $\hat{y}_t$, mit $r_{y,\hat{y}}$ bezeichnet, zum Bestimmmtheitsmass ist gegeben durch:

$$r_{y,\hat{y}}^2 = \frac{s_{y,\hat{y}}^2}{s_y^2 s_{\hat{y}}^2} = \frac{\left[\sum (y_t - \bar{y})(\hat{y}_t - \bar{y})\right]^2}{s_y^2 SSR} = \frac{\left[\sum (y_t - \bar{y})\hat{\beta}(x_t - \bar{x})\right]^2}{s_y^2 s_x^2 \hat{\beta}^2}$$

$$= \frac{\hat{\beta}^2\left[\sum (y_t - \bar{y})(x_t - \bar{x})\right]^2}{s_y^2 s_x^2 \hat{\beta}^2} = \frac{s_x^2 \hat{\beta}^2}{s_y^2} \times \frac{s_{xy}^2}{\hat{\beta}^2 s_x^4} = R^2$$

Zusammengefasst:

$$\boxed{r_{x,y}^2 = r_{y,\hat{y}}^2 = R^2}$$

2.5 Eigenschaften in kleinen Stichproben

SATZ (Gauss-Markov): Unter den Annahmen 1-6 ist der OLS-Schätzer der beste lineare erwartungstreue Schätzer für α und β. Der OLS-Schätzer ist BLUE ("best linear unbiased estimator").

Falls $\varepsilon_t \sim N(0, \sigma^2)$, dann gilt:

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} \sim N\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \underbrace{\sigma^2\begin{pmatrix}\frac{1}{T}+\frac{\bar{x}^2}{s_x^2} & \frac{-\bar{x}}{s_x^2}\\ \frac{-\bar{x}}{s_x^2} & \frac{1}{s_x^2}\end{pmatrix}}_{\text{Varianz-Kovarianz-Matrix}}\right)$$

Dies ist so zu lesen:

$$\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2}{T}+\frac{\sigma^2\bar{x}^2}{s_x^2}\right)$$

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{s_x^2}\right) \quad \text{bzw.} \quad \frac{\hat{\beta}-\beta}{\sqrt{\sigma^2/s_x^2}} \sim N(0,1)$$

$$\text{mit} \quad V\hat{\alpha} = \frac{\sigma^2}{T}+\frac{\sigma^2\bar{x}^2}{s_x^2}, \quad V\hat{\beta} = \frac{\sigma^2}{s_x^2} \quad \text{und} \quad Cov(\hat{\alpha},\hat{\beta}) = -\frac{\sigma^2\bar{x}}{s_x^2}$$

Der OLS-Schätzer von σ^2 ist definiert durch:

$$\tilde{\sigma}^2 = \frac{1}{T}\sum\hat{\varepsilon}_t^2$$

Der OLS-Schätzer von σ^2 ist nicht erwartungstreu. Ein erwartungstreuer Schätzer ist gegeben durch:

$$\hat{\sigma}^2 = \frac{1}{T-2}\sum \hat{\varepsilon}_t^2$$

Es gilt: $(T-2)\dfrac{\hat{\sigma}^2}{\sigma^2} = \dfrac{\sum \hat{\varepsilon}_t^2}{\sigma^2} \sim \chi^2(T-2)$ unabhängig von $\hat{\beta}$.

Bei bekannter Varianz ist $\dfrac{\hat{\beta}-\beta}{\sqrt{V\hat{\beta}}} = \dfrac{\hat{\beta}-\beta}{\sqrt{\sigma^2 / s_x^2}} \sim N(0,1)$

Bei geschätzter Varianz ist $\dfrac{\hat{\beta}-\beta}{\sqrt{\hat{V}\hat{\beta}}} = \dfrac{\hat{\beta}-\beta}{\sqrt{\hat{\sigma}^2 / s_x^2}} \sim t_{T-2}$

Einschub: Beziehungen zwischen einigen Verteilungen

Literatur: Wooldridge (2003), Anhang B.4 oder Kapitel Greene (1997, 2000), Kapitel 3.4.

1. $z \sim N(0,1) \Rightarrow z^2 \sim \chi^2(1)$

2. $x_1, \ldots, x_T$ sind T unabhängige Zufallsvariable mit $\chi^2(1)$ Verteilung, dann ist $x_1 + \cdots + x_T \sim \chi^2(T)$.

3. $z_1, \ldots, z_T$ sind T unabhängige Zufallsvariable mit $z_i \sim N(0, \sigma^2)$, dann gilt $\sum_{t=1}^{T} \left(\frac{z_t}{\sigma}\right)^2 \sim \chi^2(T)$.

4. $$\left.\begin{array}{l} x_1 \sim \chi^2(n_1) \\ x_2 \sim \chi^2(n_2) \\ x_1 \text{ und } x_2 \text{ unabhängig} \end{array}\right\} \Rightarrow \frac{x_1/n_1}{x_2/n_2} \sim F_{n_1,n_2}$$

5. $$\left.\begin{array}{l} z \sim N(0,1) \\ x \sim \chi^2(T) \end{array}\right\} \Rightarrow \frac{z}{\sqrt{x/T}} \sim t_T$$

6. $\frac{z^2}{x/T} \sim F_{1,T}$ da $z^2 \sim \chi^2(1)$ und $x \sim \chi^2(T)$

Einschub: Formulieren und Testen von Hypothesen

Literatur: Greene (1993), Kapitel 4.8, (1997, 2000), Kapitel 4.9.

Nullhypothese: H_0

Alternativhypothese oder Gegenhypothese: H_1

Die Nullhypothese stellt in der klassischen Prüfstatistik die Basis dar, von der aus entschieden wird, ob die Alternativhypothese akzeptiert werden kann oder nicht. Nur wenn die Daten "praktisch" nicht mit der Nullhypothese erklärt werden können, darf sie zugunsten der Alternativhypothese verworfen werden.

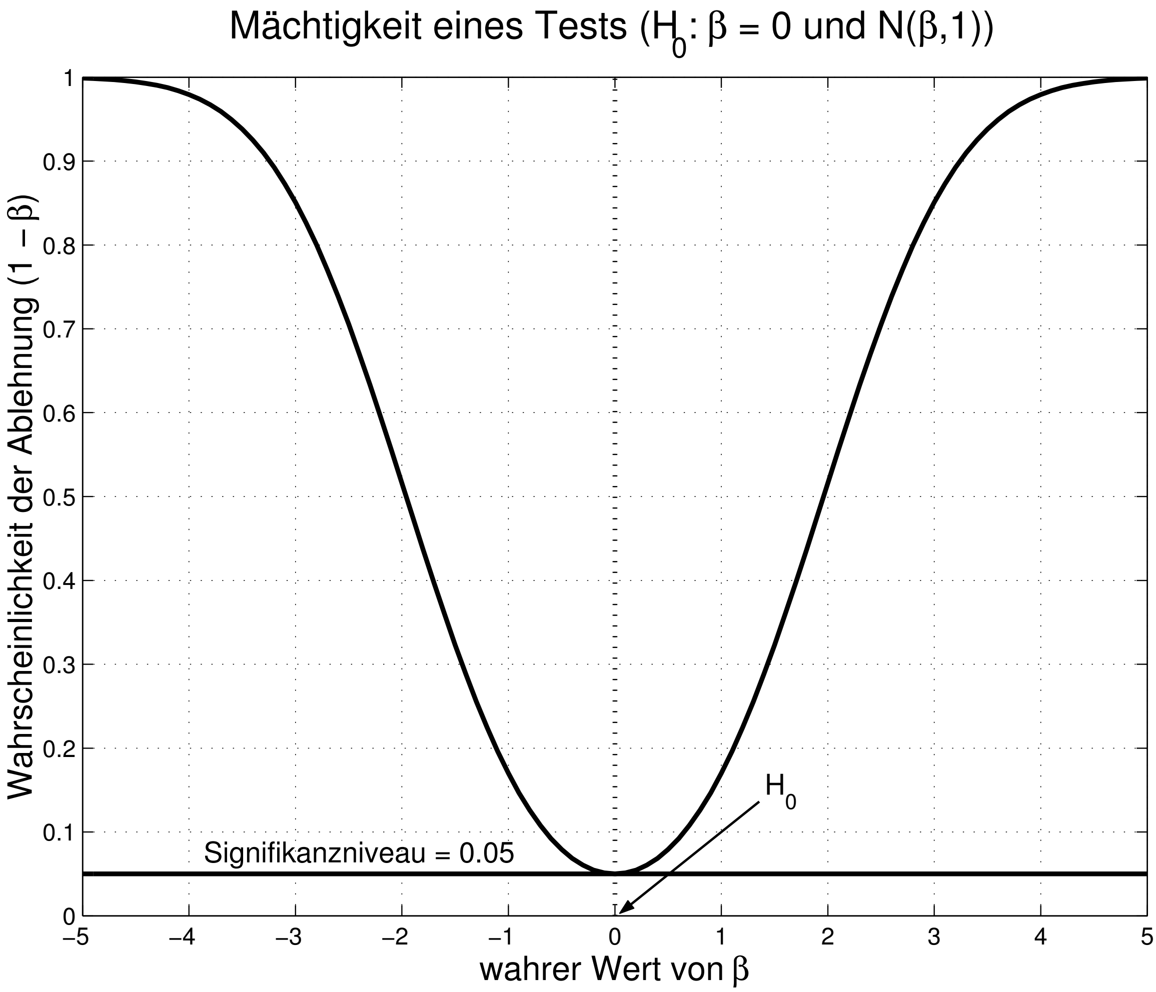
α-Fehler oder **Fehler erster Art** ("**type I error**"): Die Nullhypothese wird abgelehnt, obwohl sie wahr ist.

β-Fehler oder **Fehler zweiter Art** ("**type II error**"): Die Nullhypothese wird nicht abgelehnt, obwohl sie falsch ist.

	In der Population gilt	
	$\mathbf{H_0}$	$\mathbf{H_1}$
Entscheidung aufgrund der Stichprobe zugunsten von $\mathbf{H_0}$	richtige Entscheidung	β-Fehler Fehler 2. Art
Entscheidung aufgrund der Stichprobe zugunsten von $\mathbf{H_1}$	α-Fehler Fehler 1. Art	richtige Entscheidung

- Die Wahrscheinlichkeit einen Fehler erster Art zu begehen, heisst <u>Signifikanzniveau</u> ("size"; "significance level") des Tests und wird üblicherweise mit α bezeichnet. Dieses ist typischerweise 0.05.
- Die Wahrscheinlichkeit mit der eine falsche Nullhypothese verworfen wird, heisst <u>Mächtigkeit</u> des Tests ("power")

  Mächtigkeit = 1 - β = 1 - P(Fehlers zweiter Art)
- Bei gegebenem Signifikanzniveau α möchte man β so klein als möglich halten. Die Mächtigkeit eines Tests hängt allerdings vom wahren Wert des Parameters ab.



3 Das multiple lineare Regressionsmodell

Literatur: Wooldridge (2003), Kapitel 3 oder Greene (1997, 2000), Kapitel 6.1-6.6.

3.1 Das statistische Modell

Für die Beobachtungen t = 1,...,T wird der Zusammenhang postuliert:

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_K x_{Kt} + \varepsilon_t$$

Matrixschreibweise:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{x_{1T}}_{\text{1. Regressor}} & \underbrace{x_{2T}}_{\text{2. Regressor}} & \cdots & \underbrace{x_{KT}}_{\text{K-ter Regressor}} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} x_1' \} \text{ 1. Beobachtung} \\ x_2' \} \text{ 2. Beobachtung} \\ \vdots \\ x_T' \} \text{ T. Beobachtung} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

$$y = X\beta + \varepsilon$$

Ziel ist die Schätzung der unbekannten Parameter β und σ^2, sowie das Testen von Hypothesen über diese Parameter.

Das Regressionsproblem ist nur dann sinnvoll, wenn die K Regressoren linear unabhängig sind. In diesem Fall hat die Matrix X den Rang K. Eine notwendige Bedingung dafür ist, dass die Anzahl der Beobachtungen grösser als die Anzahl der Regressoren ist, d.h. $T \geq K$.

Annahmen:

1. funktionale Form

   lineares Modell: $y = X\beta + \varepsilon$

2. Erwartungswert null

   $E\varepsilon = 0$

3. konstante Varianz (Homoskedastizität) und keine Korrelation zwischen den Beobachtungen

   $E\varepsilon\varepsilon' = \sigma^2 I_T$; $(E\varepsilon_t^2 = \sigma^2$; $E\varepsilon_t\varepsilon_s = 0$ für $t \neq s)$

4. keine Information in X über ε

   $E(\varepsilon|X) = 0 \Rightarrow \mathrm{cov}(X, \varepsilon) = 0$

5. keine Multikollinearität

   X ist deterministisch mit rang(X) = K.

   Daraus folgt: $(X'X)^{-1}$ existiert.

6. Normalität

   $\varepsilon \sim N\left(0, \sigma^2 I_T\right)$

3.2 Schätzung: die Methode der kleinsten Quadrate (OLS)

Literatur: Wooldridge (2003), Kapitel 3.2 oder Greene (1997, 2000), Kapitel 6.4-6.4.2.

Der OLS-Schätzer („ordinary least squares“) minimiert die Summe der quadrierten Residuen:

$$S(\beta) = \varepsilon'\varepsilon = \sum \varepsilon_t^2 \to \min_{\beta}$$

OLS-Schätzer: $\hat{\beta} = (X'X)^{-1}X'y = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \cdot \sum_{t=1}^{T} x_t y_t$

OLS-Residuen:

$$e = \hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta} = \left(\underbrace{I_T - X(X'X)^{-1}X'}_{M}\right)y = My$$

M ist eine Projektion (M ist symmetrisch, d.h. M = M´; M ist idempotent, d.h. MM = M). Es gilt MX = 0: M projiziert auf den Raum, der orthogonal zu den Spalten von X aufgespannten Raum steht.

Der OLS-Prädiktor von y:

$$\hat{y} = X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{P}y = Py$$

P ist eine Projektion (P ist symmetrisch und idempotent). Es gilt PX = X: P projiziert auf den von den Spalten von X aufgespannten Raum.

3.3 Güte der Anpassung: das Bestimmtheitsmass R^2

Literatur: Wooldridge (2003), Kapitel 3.2 oder Greene (1997, 2000), Kapitel 6.5.

Mittelwertbereinigung: $x_t - \bar{x}$; $x = \left(x_1 \ x_2 \cdots x_T\right)'$

Mit Hilfe von $\iota = \underbrace{\left(1 \quad 1 \quad \cdots \quad 1\right)}_{\text{T Einsen}}'$ lässt sich ein Operator M^0 zur Mittelwertbereinigung generieren:

$$M^0 = I_T - \iota\left(\iota'\iota\right)^{-1}\iota' = I_T - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \frac{1}{T}\left(1 \quad 1 \quad \cdots \quad 1\right) = I_T - \frac{1}{T}\iota\iota'$$

$$M^0 x = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_T - \bar{x} \end{pmatrix}$$; M^0x entspricht den mittelwertbereinigten Daten

Das Bestimmtheitsmass R^2 ist definiert als:

$$\boxed{R^2 = 1 - \frac{e'e}{y'M^0y} = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}}$$

SATZ: $0 \le R^2 \le 1$, falls X eine Konstante enthält.

BEWEIS: $M^0y = M^0X\hat{\beta} + M^0e = M^0X\hat{\beta} + e$

$$\begin{aligned} y'M^0y &= \left(\hat{\beta}'X' + e'\right)M^0X\hat{\beta} + \left(\hat{\beta}'X' + e'\right)e \\ &= \hat{\beta}'X'M^0X\hat{\beta} + e'M^0X\hat{\beta} + \hat{\beta}'X'e + e'e = \hat{\beta}'X'M^0X\hat{\beta} + e'e \end{aligned}$$

$$\Rightarrow 1 = \underbrace{\frac{\hat{\beta}'X'M^0X\hat{\beta}}{y'M^0y}}_{>0} + \underbrace{\frac{e'e}{y'M^0y}}_{>0} \geq 0$$

BEMERKUNG: Falls X keine Konstante enthält, ist $M^0 e \neq e$ und $e'M^0X \neq 0$. Das R^2 ist in diesem Fall beliebig, es kann sogar negativ werden.

Das korrigierte Bestimmtheitsmass $\overline{R}^2$

R^2 steigt mit der Anzahl der Regressoren

Korrigiertes R^2:

$$\boxed{\overline{R}^2 = 1 - \frac{e'e/(T-K)}{y'M^0y/(T-1)} = 1 - \frac{T-1}{T-K}\left(1-R^2\right)}$$

Das korrigierte R^2 kann fallen, falls ein zusätzlicher Regressor in die Gleichung eingebaut wird; es kann sogar negativ werden.

3.4 Eigenschaften in kleinen Stichproben

Literatur: Wooldridge (2003), Kapitel 3.5 oder Greene (1997, 2000), Kapitel 6.6.1-6.6.2.

Erwartungswert des OLS-Schätzers:

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$$
$$\Rightarrow E\hat{\beta} = \beta + (X'X)^{-1}X'E\varepsilon = \beta$$

d.h. der OLS-Schätzer ist erwartungstreu.

Varianz des OLS-Schätzers:

$$V\hat{\beta} = E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'\right] = (X'X)^{-1}X'(E\varepsilon\varepsilon')X(X'X)^{-1} = \underbrace{\sigma^2(X'X)^{-1}}_{\text{Varianz-Kovarianz-Matrix}}$$

Gauss-Markov Theorem: Der OLS-Schätzer ist der beste lineare erwartungstreue Schätzer von β. Dies gilt auch für jede beliebige Linearkombination $w'\beta$. D.h. der OLS-Schätzer ist auch bezüglich jedes einzelnen Koeffizienten BLUE.

3.5 Das Testen von Hypothesen: t-, Wald- und F-Test

Literatur: Wooldridge (2003), Kapitel 3.2 oder Greene (1997, 2000), Kapitel 6.6.3-6.6.6, 7.2.

Der OLS-Schätzer ist eine lineare Funktion von ε. Aus Annahme 6 folgt daher, dass $\hat{\beta} \sim N\left(\beta, \sigma^2(X'X)^{-1}\right)$, insbesondere gilt $\hat{\beta}_k \sim N\left(\beta_k, \sigma^2(X'X)^{-1}_{kk}\right)$, wobei $\sigma^2(X'X)^{-1}_{kk} = \sigma^2 S_{kk} = V\hat{\beta}_k$. Daraus folgt für die standardisierte Verteilung:

$$z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 S_{kk}}} \sim N(0,1)$$

wobei S_{kk} das k-te Diagonalelement von $(X'X)^{-1}$ bezeichnet. σ^2 ist jedoch nicht bekannt, kann aber aus den Residuen geschätzt werden:

$e = My = M(X\beta+\varepsilon) = M\varepsilon$. Daher ist $e'e = \varepsilon'M'M\varepsilon = \varepsilon'M\varepsilon$

$$E(e'e) = E[\mathrm{tr}(e'e)] = E[\mathrm{tr}(\varepsilon'M\varepsilon)] = E[\mathrm{tr}(M\varepsilon'\varepsilon)] = \mathrm{tr}\left(M\underbrace{E\varepsilon'\varepsilon}_{\sigma^2 I_T}\right) = \sigma^2\mathrm{tr}(M)$$

$$\mathrm{tr}(M) = \mathrm{tr}\left(I_T - X(X'X)^{-1}X'\right)$$
$$= \mathrm{tr}(I_T) - \mathrm{tr}\left(X(X'X)^{-1}X'\right) = T - \mathrm{tr}\left(\underbrace{(X'X)^{-1}(X'X)}_{I_K}\right) = T - K$$

Ein erwartungstreuer Schätzer von σ^2 ist daher:

$$\hat{\sigma}^2 = s^2 = \frac{e'e}{T-K} = \frac{y'y - y'X\hat{\beta}}{T-K} = \frac{y'My}{T-K}$$

Die geschätzte Varianz von $\hat{\beta}$ ist demnach $s^2(X'X)^{-1}$.

Es gilt:

$$(T-K)\frac{s^2}{\sigma^2} = \frac{e'e}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)' M\left(\frac{\varepsilon}{\sigma}\right) \sim \chi^2_{\mathrm{tr}(M)} = \chi^2_{T-K}.$$

Diese Verteilung ist unabhängig von β_k, so dass

$$t_k = \frac{\dfrac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 S_{kk}}}}{\sqrt{\dfrac{(T-K)s^2/\sigma^2}{T-K}}} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{s^2 S_{kk}}} \sim t_{T-K}$$

Der Hypothesentest bei einzelnen Koeffizienten erfolgt analog zum einfachen linearen Regressionsmodell.

Test einer linearen Restriktion

H_0: $$r_1\beta_1 + r_2\beta_2 + \ldots + r_K\beta_K = r'\beta = q$$

t-Statistik: $$t = \frac{r'\hat{\beta} - r'\beta}{\sqrt{V\left(r'\hat{\beta}\right)}} = \frac{\hat{q} - q}{\sqrt{V\hat{q}}} \sim t_{T\text{-}K}\ ,$$

wobei $$V\hat{q} = V\left(r'\hat{\beta}\right) = s^2 r'(X'X)^{-1} r$$

Test mehrerer linearer Restriktionen

H_0: $$\underbrace{R}_{J\times K}\beta = \underbrace{q}_{J\times 1}$$

Die Anzahl der Restriktionen J entspricht der Anzahl der Zeilen von R.

- Test eines einzelnen Koeffizienten: H_0: $\beta_k = 0$

$$R = \begin{pmatrix} 0 & \cdots & 0 & \underbrace{1}_{\substack{k-\text{te}\\ \text{Stelle}}} & 0 & \cdots & 0 \end{pmatrix} \qquad q = 0$$

- Test mehrerer Koeffizienten auf null: H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \quad q = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- Test auf Gleichheit: H_0: $\beta_1 = \beta_2$

$$R = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \end{pmatrix} \quad q = 0$$

Der OLS-Schätzer ist normal verteilt $\Rightarrow$ $R\hat{\beta} - q \sim$ normal verteilt

$$V\left(R\hat{\beta} - q\right) = RV\left(\hat{\beta}\right)R' = \sigma^2 R(X'X)^{-1} R'$$

<u>Wald-Test</u>:

$$\begin{aligned} W &= \left(R\hat{\beta} - q\right)' V\left(R\hat{\beta} - q\right)^{-1}\left(R\hat{\beta} - q\right) \\ &= \left(R\hat{\beta} - q\right)' \left[\sigma^2 R(X'X)^{-1} R'\right]^{-1}\left(R\hat{\beta} - q\right) \sim \chi^2_J \end{aligned}$$

σ^2 ist nicht bekannt, wird aber durch den unverzerrten Schätzer s^2 ersetzt. Die so ermittelte neue Testgrösse ist als Verhältnis von zwei χ^2-verteilten Variablen $F_{J,T\text{-}K}$-verteilt:

$$F = \frac{\left(R\hat{\beta} - q\right)' \left[\sigma^2 R(X'X)^{-1} R'\right]^{-1}\left(R\hat{\beta} - q\right) / J}{\left[(T-K)s^2 / \sigma^2\right] / (T-K)} \sim F_{J,T-K}$$

Durch geeignete Umformung lässt sich σ^2 eliminieren und es ergibt sich die Testgrösse zu

$$\begin{aligned} F &= \frac{\left(R\hat{\beta} - q\right)' \left[R(X'X)^{-1} R'\right]^{-1}\left(R\hat{\beta} - q\right) / J}{e'e / (T-K)} \\ &= \frac{\left(R\hat{\beta} - q\right)' \left[s^2 R(X'X)^{-1} R'\right]^{-1}\left(R\hat{\beta} - q\right)}{J} \end{aligned}$$

J = Anzahl Restriktionen, Anzahl Zeilen von R

T = Anzahl Beobachtungen

K = Anzahl Regressoren

4 Relevante Regressoren

Literatur: Wooldridge (2003), Kapitel 3.3 oder Greene (1997, 2000), Kapitel 8.4.

4.1 Partielle Regression

Literatur: Greene (1997, 2000), Kapitel 6.4.3.

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

wobei $X = (X_1, X_2)$ ist. Der OLS-Schätzer ist gegeben durch:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

Die Lösung für β_2 ist gegeben durch:

$$\boxed{\begin{aligned} \hat{\beta}_2 &= \left[X_2'\left(I_T - X_1(X_1'X_1)^{-1}X_1'\right)X_2\right]^{-1} X_2'\left(I_T - X_1(X_1'X_1)^{-1}X_1'\right)y \\ &= (X_2'M_1X_2)^{-1}X_2'M_1y = \left(\tilde{X}_2'\tilde{X}_2\right)^{-1}\tilde{X}_2'\tilde{y} \end{aligned}}$$

mit $\tilde{X}_2 = M_1X_2$ und $\tilde{y} = M_1y$.

4.2 Ausgelassene Regressoren ("omitted variables")

Literatur: Wooldridge (2003), Kapitel 3.3 oder Greene (1997, 2000), Kapitel 8.4.2.

Das "wahre" Modell ist: $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

In der Schätzung wird allerdings X_2 vernachlässigt und das *fehlspezifizierte* Modell $y = X_1\beta_1 + \varepsilon$ geschätzt.

$$\begin{aligned}\hat{\beta}_1 &= (X_1'X_1)^{-1}X_1'y = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon\end{aligned}$$

$$E\hat{\beta}_1 = \beta_1 + \underbrace{(X_1'X_1)^{-1}X_1'X_2}_{P_{1.2}}\beta_2$$

$$V\hat{\beta}_1 = \sigma^2(X_1'X_1)^{-1}$$

Der OLS-Schätzer ist verzerrt bzw. nicht erwartungstreu. Die Richtung der Verzerrung ("bias") hängt sowohl von X'_1X_2, der Kovarianz zwischen X_1 und X_2, als auch vom Koeffizienten β_2 ab.

In der korrekten Regression wird β_1 geschätzt als $\hat{\beta}_{1.2} = (X_1'M_2X_1)^{-1}X_1'M_2y$ mit $M_2 = I_T - X_2(X_2'X_2)^{-1}X_2'$.

$$\begin{aligned}\hat{\beta}_{1.2} &= (X_1'M_2X_1)^{-1}X_1'M_2\underbrace{(X_1\beta_1 + X_2\beta_2 + \varepsilon)}_{y} \\ &= \beta_1 + \underbrace{(X_1'M_2X_1)^{-1}X_1'M_2X_2\beta_2}_{=O \text{ da } M_2X_2=0} + (X_1'M_2X_1)^{-1}X_1'M_2\varepsilon \\ &= \beta_1 + (X_1'M_2X_1)^{-1}X_1'M_2\varepsilon\end{aligned}$$

$$\begin{aligned}E\hat{\beta}_{1.2} &= \beta_1 \\ V\hat{\beta}_{1.2} &= E\left[(X_1'M_2X_1)^{-1}X_1'M_2\varepsilon\varepsilon'M_2X_1(X_1'M_2X_1)^{-1}\right] = \sigma^2(X_1'M_2X_1)^{-1}\end{aligned}$$

Ein Vergleich der Kehrwerte der beiden Varianzen zeigt:

$$\left(V\hat{\beta}_1\right)^{-1} - \left(V\hat{\beta}_{1.2}\right)^{-1} = \frac{1}{\sigma^2}(X_1'X_1) - \frac{1}{\sigma^2}(X_1'M_2X_1)$$
$$= \frac{1}{\sigma^2}X_1'(I_T - M_2)X_1 = \frac{1}{\sigma^2}X_1'X_2(X_2'X_2)^{-1}X_2'X_1 \geq 0$$
$$\Rightarrow \quad V\hat{\beta}_1 \leq V\hat{\beta}_{1.2}$$

Der Schätzer $\hat{\beta}_1$ ist zwar verzerrt, hat aber die kleinere Varianz als der erwartungstreue Schätzer $\hat{\beta}_{1.2}$.

Ein zusätzliches Problem besteht in der Schätzung von σ^2.

$$\hat{\sigma}^2 = s^2 = \frac{e_1'e_1}{T - K_1}$$
$$e_1 = M_1 y = M_1(X_1\beta_1 + X_2\beta_2 + \varepsilon) = M_1X_2\beta_2 + M_1\varepsilon$$

$$Ee_1'e_1 = E(\beta_2'X_2'M_1 + \varepsilon'M_1)(M_1X_2\beta_2 + M_1\varepsilon)$$
$$= \beta_2'X_2'M_1X_2\beta_2 + E\beta_2'X_2'M_1\varepsilon + E\varepsilon'M_1X_2\beta_2 + E\varepsilon'M_1\varepsilon$$
$$= \underbrace{\beta_2'X_2'M_1X_2\beta_2}_{\geq 0,\text{ d.h. Verzerrung nach oben}} + (T - K_1)\sigma^2$$

Schlussfolgerungen:

- Die Schätzungen von β_1 und σ^2 sind verzerrt.
- Die geschätzte Varianz von $\hat{\beta}_1$ kann geringer sein als jene von $\hat{\beta}_{1.2}$.
- Falls $X_1'X_2 = 0$, ist $\hat{\beta}_1$ unverzerrt. Allerdings bleibt s^2 weiterhin verzerrt.

4.3 Zu viele Regressoren

Literatur: Wooldridge (2003), Kapitel 6.3 oder Greene (1997, 2000), Kapitel 8.4.3.

Das "wahre" Modell sei: $y = X_1\beta_1 + \varepsilon$

Geschätzt wird aber das erweiterte Modell: $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

Dieses Modell ist im Prinzip nicht falsch (!), es wurde lediglich die Information $\beta_2 = 0$ nicht berücksichtigt. Der OLS-Schätzer ist nach wie vor erwartungstreu: $E\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \begin{pmatrix}\beta_1\\ 0\end{pmatrix}$

Das selbe gilt für den Schätzer der Varianz: $E\dfrac{e'e}{T - K_1 - K_2} = \sigma^2$

Es schaut so aus, als ob es immer günstiger ist zu viele Regressoren als zu wenige zu verwenden. Das Problem mit dieser Ansicht ist, dass die Präzision der Schätzung mit der Zahl der Regressoren abnimmt; die Varianz der Koeffizienten ist bei der Regression mit weniger Variablen kleiner.

5 Der restringierte OLS-Schätzer

Literatur: Wooldridge (2003), Kapitel 4.5 oder Greene (1997, 2000), Kapitel 7.3, 7.4, 7.6.

5.1 Schätzung und Eigenschaften

Modell: $y = X\beta + \varepsilon$

Minimierungsproblem:

$$S(\tilde{\beta}) = (y - X\tilde{\beta})'(y - X\tilde{\beta}) \rightarrow \min_{\tilde{\beta}}$$

$$\text{NB}: \; R\tilde{\beta} = q$$

Die Nebenbedingung (NB, Restriktion) enthält zusätzliche Information!

Die Lagrange-Funktion L mit dem Vektor der Lagrange-Multiplikatoren λ zur Lösung des Minimierungsproblems unter der Nebenbedingung lautet:

$$L = (y - X\tilde{\beta})'(y - X\tilde{\beta}) + 2\lambda'(R\tilde{\beta} - q)$$

Bedingungen 1. Ordnung: $$\frac{\partial L}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) + 2R'\lambda \overset{!}{=} 0$$

$$\frac{\partial L}{\partial \lambda} = 2(R\tilde{\beta} - q) \overset{!}{=} 0$$

$\Rightarrow$ Gleichungssystem in zwei Unbekannten: $\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}\begin{pmatrix} \tilde{\beta} \\ \lambda \end{pmatrix} = \begin{pmatrix} X'y \\ q \end{pmatrix}$

Lösung: $$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta} - q\right)$$

$$\lambda = \left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta} - q\right)$$

$$V\tilde{\beta} = \underbrace{\sigma^2(X'X)^{-1}}_{V\hat{\beta}} - \underbrace{\sigma^2(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}}_{\text{positiv definite Matrix}}$$

Verglichen mit dem unrestringierten Fall nimmt die Varianz ab, dies entspricht dem Wert der zusätzlich in der Restriktion enthaltenen Information.

Die Anpassung an die Daten ist allerdings schlechter.

5.2 Test Idee

Vergleich der Anpassung an die Daten des Modells mit und ohne Restriktion

$$\tilde{e} = y - X\tilde{\beta} = y - X\hat{\beta} - X\left(\tilde{\beta} - \hat{\beta}\right) = e - X\left(\tilde{\beta} - \hat{\beta}\right)$$

$$\tilde{e}'\tilde{e} = e'e + \left(\tilde{\beta} - \hat{\beta}\right)' X'X\left(\tilde{\beta} - \hat{\beta}\right) \geq e'e$$

$$\Rightarrow \quad \tilde{e}'\tilde{e} - e'e = \left(R\hat{\beta} - q\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta} - q\right)$$

$$\boxed{F(J, T-K) = \frac{(\tilde{e}'\tilde{e} - e'e)/J}{e'e/(T-K)} = \frac{\left(R^2 - \tilde{R}^2\right)/J}{\left(1 - R^2\right)/(T-K)}}$$

5.3 Beispiel: Signifikanz der Regression

Literatur: Wooldridge (2003), Kapitel 4.5 oder Greene (1997, 2000), Kapitel 6.6.7.

Hypothese: alle Koeffizienten bis auf die Konstante sind null,

$$\text{d.h. } \tilde{R}^2 = 0,\ J = K-1. \qquad \Rightarrow F(K-1, T-K) = \frac{R^2/(K-1)}{(1-R^2)/(T-K)}$$

5.4 Beispiel: Test auf Strukturbruch (Chow-Test)

Literatur: Wooldridge (2003), Kapitel 13.1 oder Greene (1997, 2000), Kapitel 7.6.

Zwei oder mehrere Zeitperioden oder auch zwei oder mehrere Gruppen! Es soll getestet werden, ob die Koeffizienten der einen Zeitperiode/Gruppe verschieden von denjenigen der anderen sind oder nicht.

Im Fall von zwei Perioden/Gruppen ergeben sich die folgenden Ansätze:

- Unrestringiertes Modell: Strukturbruch

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$y = X\beta + \varepsilon$$

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1'y_1 \\ X_2'y_2 \end{pmatrix}$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y_1 \qquad \hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y_2$$

$$\Rightarrow e'e = e_1'e_1 + e_2'e_2$$

Hypothese: $\beta_1 = \beta_2$; R=(I, -I) q=0

Entweder verwendet man die vorigen Resultate <u>oder</u> schätzt folgende zusätzliche Regression:

- Restringiertes Modell: kein Strukturbruch

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$$\Rightarrow \begin{matrix} \tilde{e}'\tilde{e} \text{ restringierte} \\ \text{Fehlerquadratsumme} \end{matrix}$$

Daraus lässt sich der Chow-Test für den allgemeinen Fall ableiten:

$$\boxed{F\left(\underbrace{J}_{\substack{\text{Anzahl der} \\ \text{Restriktionen}}}, \underbrace{T-2K}_{\substack{\text{Freiheitsgrade im} \\ \text{unrestringierten Modell}}}\right) = \frac{(\tilde{e}'\tilde{e} - e_1'e_1 - e_2'e_2)/J}{(e_1'e_1 + e_2'e_2)/(T-2K)} \sim F_{J,\,T\text{-}2K}}$$

Wobei: J = Anzahl der Restriktionen (hier K)

2K = Anzahl Regressoren im unrestringierten Modell

T = Gesamtzahl der Beobachtungen

Variante: Die Konstanten bleiben unrestringiert, Änderungen sind nur in der Steigung möglich. Die restringierte Regressormatrix ist daher $X = \begin{pmatrix} i & 0 & W \\ 0 & i & W \end{pmatrix}$. Die Regressoren (i´ 0´)´ bzw. (0´ i´)´ werden als Dummy-Variablen bezeichnet.

Falls nicht genügend Freiheitsgrade vorhanden sind, kann man auch einen alternativen Chow-Test (predictive Chow-test) durchführen. Dieser besteht aus folgenden Schritten:

1. Unter Verwendung des vollständigen Datensatzes wird die restringierte Schätzung durchgeführt. Diese liefert die Residuen $\tilde{e}$.
2. Unter Verwendung der grösseren der beiden Stichproben wird die unrestringierte Schätzung durchgeführt. Diese liefert die Residuen e.
3. $F(T_1, T_2 - K) = \dfrac{(\tilde{e}'\tilde{e} - e'e)/T_1}{e'e/(T_2 - K)}$, wobei $T_1<K$.

Im Fall von n verschiedenen Perioden bzw. Gruppen folgt aus ähnlichen Überlegungen folgende Teststatistik für den Chow-Test:

$$\boxed{F\left(\underbrace{J}_{\substack{\text{Anzahl der}\\ \text{Restriktionen}}}, \underbrace{T - nK}_{\substack{\text{Freiheitsgrade im}\\ \text{unrestringierten Modell}}}\right) = \frac{\left(\tilde{e}'\tilde{e} - \sum_{i=1}^{n} e_i' e_i\right) \Big/ J}{\sum_{i=1}^{n} e_i' e_i \Big/ (T - nK)} \sim F_{J,\, T-nK}}$$

Wichtig: Alle Varianten des Chow-Tests nehmen eine über alle Perioden bzw. Gruppen konstante Varianz an.

6 Prognose

Literatur: Wooldridge (2003), Kapitel 6.4 oder Greene (1997, 2000), Kapitel 7.11.

Gegeben $x^0 = \left(x_1^0, \cdots, x_K^0\right)$ wollen wir eine Prognose für y^0 unter Verwendung des Modells $y^0 = x^0\beta + \varepsilon^0$ erstellen. Aus dem Gauss-Markov-Theorem folgt, dass $\hat{y}^0 = x^0\hat{\beta}$ der beste lineare erwartungstreue Schätzer von Ey^0 ist.

Prognosefehler:

$$e^0 = y^0 - \hat{y}^0 = x^0\left(\beta - \hat{\beta}\right) + \varepsilon^0 \quad \Rightarrow \quad Ee^0 = 0$$

Varianz des Prognosefehlers:

$$Ve^0 = \sigma^2 + V\left(x^0\left(\beta - \hat{\beta}\right)\right) = \sigma^2 + \sigma^2 x^0 (X'X)^{-1} x^{0\prime}$$

Es kann gezeigt werden, dass die Prognosegenauigkeit abnimmt, je weiter x^0 vom "Zentrum" der Daten $(\Rightarrow \bar{x})$ entfernt ist. Das Konfidenzintervall für die Prognose ist gegeben durch: $\hat{y}^0 \pm t_{\alpha/2} se\left(e^0\right)$; se = Standardfehler

Im Spezialfall mit zwei Regressoren (K = 2)

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ & \vdots \\ 1 & x_T \end{pmatrix} = \begin{pmatrix} 1 & x \end{pmatrix}, \qquad (X'X)^{-1} = \frac{1/T}{Vx} \begin{pmatrix} \frac{1}{T}\sum x_t^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},$$

wobei einer der Regressoren eine Konstante ist (X enthält einen Spaltenvektor aus lauter Einsen), resultiert:

$$Ve^0 = \sigma^2 \left[1 + \frac{1}{T} + \frac{\left(x^0 - \bar{x}\right)^2}{\sum \left(x_t - \bar{x}\right)^2} \right] \; ; \quad \text{mit} \; \sum \left(x_t - \bar{x}\right)^2 = (T-1) \cdot Vx$$

6.1 Messung der Prognosegenauigkeit

T^0 = Anzahl durchgeführte Prognosen (Anzahl der Prognoseperioden)

- RMSE ("root mean-squared error"): $\sqrt{\left(1/T^0\right) \sum \left(y_t - \hat{y}_t\right)^2}$

  Der RMSE ist das häufigste Prognosemass

- MAE ("mean absolute error"): $\frac{1}{T^0} \sum \left| y_t - \hat{y}_t \right|$

- Theil'sches U: $\sqrt{\dfrac{\left(\frac{1}{T^0}\right) \sum \left(y_t - \hat{y}_t\right)^2}{\left(\frac{1}{T^0}\right) \sum y_t^2}}$

  Vorteil des Theil'schen U ist dessen Invarianz gegenüber Masseinheiten.

Einschub: Konsistenz und asymptotische Normalität

Literatur: Wooldridge (2003), Anhang C.3 oder Greene (1997, 2000), Kapitel 4.4.1, (4.4.2-4.4.5).

Was geschieht, wenn die Grösse der Stichprobe gegen unendlich geht?

DEFINITION: Eine Folge von Zufallsvariablen $\{X_t\}$ <u>konvergiert in Wahrscheinlichkeit</u> gegen eine Konstante c genau dann, wenn der Grenzwert der Wahrscheinlichkeit, dass der Betrag $|X_t - c|$ grösser als jedes feste $\varepsilon > 0$ ist, für $t \to \infty$ gleich Null ist:

$$\lim_{t\to\infty} P\left[|X_t - c| > \varepsilon\right] = 0 \text{ für alle } \varepsilon > 0$$

Schreibweise: plim $X_t = c$ (plim = Wahrscheinlichkeitslimes)

DEFINITION: Ein Schätzer heisst <u>konsistent</u>, falls $\operatorname{plim}\hat{\beta}_T = \beta$.

Insbesondere gilt: $\left.\begin{array}{l}\lim\limits_{t\to\infty} EX_t = c \\ \lim\limits_{t\to\infty} VX_t = 0\end{array}\right\} \Rightarrow \quad \operatorname{plim} X_t = c$

Neben der Konsistenz stellt sich die Frage nach der Verteilung eines Schätzers, wenn die Anzahl der Beobachtungen gegen unendlich geht. Dies ist deshalb wichtig, da die Verteilung bei kleinen Stichproben im allgemeinen nicht bekannt ist. Unter bestimmten Umständen kann diese aber durch die asymptotische Verteilung (Verteilung bei unendlicher Stichprobe) approximiert werden. Aussagen über die asymptotische Verteilung werden unter dem Stichwort zentraler Grenzwertsatz ("central limit theorem") zusammengefasst. Eine einfache Version eines zentralen Grenzwertsatzes ist der Satz von Lindeberg-Lévy.

SATZ: Falls $\{X_t\}$ eine Folge von identisch und unabhängig verteilten Zufallsvariablen mit $EX_t = \mu$ und $VX_t = \sigma^2$ ist, dann gilt:

$$Z_T = \frac{\overline{X}_T - E\overline{X}_T}{\sqrt{V\overline{X}_T}} \to N(0,1) \ , \qquad \text{wobei } \overline{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t \ .$$

Für endliche Stichproben kann deshalb die Verteilung von $\overline{X}_T$ angenähert werden:

$$\overline{X}_T \overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{T}\right).$$

7 Asymptotische Eigenschaften des OLS-Schätzers

Literatur: Wooldridge (2003), Kapitel 5 und Anhang C.3 oder Greene (1997), Kapitel 6.7, 6.7.1-6.7.3, 6.7.6, Greene (2000), Kapitel 9, 9.2, 9.3.1, 9.3.2, 9.3.5.

Unter der Annahme, dass $\lim_{T\to\infty} \frac{1}{T} X'X = Q$ eine positiv definite Matrix ist, und daher Q^{-1} existiert, gilt:

$$E\hat{\beta} = \beta$$

$$V\hat{\beta} = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{T}\left(\frac{X'X}{T}\right)^{-1} \to 0$$

Der KQ-Schätzer ist demnach konsistent.

Mit dem zentralen Grenzwertsatz kann gezeigt werden, dass unter den beiden folgenden Annahmen:

- $E\varepsilon_t = 0$ und $V\varepsilon_t = \sigma^2$
- $|x_{kt}|$ sind beschränkt, d.h. die Elemente der X-Matrix sind endlich,

der OLS-Schätzer von β asymptotisch normal verteilt ist, selbst wenn der Störterm nicht normalverteilt sein sollte:

$$\sqrt{T}\left(\hat{\beta} - \beta\right) \to N\left(0, \sigma^2 Q^{-1}\right).$$

In endlichen Stichproben gilt somit:

$$\hat{\beta} \overset{A}{\sim} N\left(\beta, \frac{\sigma^2}{T} Q^{-1}\right).$$

In der Praxis wird σ^2 durch e'e/(T-K) und Q^{-1}/T durch $(X'X)^{-1}$ geschätzt.

Die t-Statistik für den Test einer linearen Restriktion folgt asymptotisch einer Normalverteilung:

$$t_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{s^2 (X'X)^{-1}_{kk}}} \rightarrow N(0,1).$$

Die F-Statistik für den Test mehrerer linearer Restriktionen ist nicht mehr F-verteilt, aber die Wald Statistik

$$W = JF = \frac{\tilde{e}'\tilde{e} - e'e}{e'e/(T-K)} = \left(R\hat{\beta} - q\right)' \left[Rs^2 (X'X)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right) \overset{A}{\sim} \chi^2_J$$

ist asymptotisch χ^2-verteilt mit J Freiheitsgraden.

8 Generalized Least Squares

Literatur: Wooldridge (2003), Kapitel 8.4 (und Kapitel 12) oder Greene (1997, 2000), Kapitel 11.1.

Verallgemeinerte Methode der kleinsten Quadrate, die Annahme 3 (Homoskedastizität und keine Korrelation zwischen den Beobachtungen) sowie die Annahme 6 (Normalität) werden nun weggelassen.

Modell: $$y = X\beta + \varepsilon,$$

$$E\varepsilon = 0, \quad V\varepsilon = \sigma^2\Omega.$$

Die Varianz-Kovarianz Matrix Ω ist eine symmetrisch positiv definite Matrix.

Zwei Beispiele, die typischerweise in der Praxis auftreten:

1. Heteroskedastizität: $$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix}$$

In diesem Fall ist die Varianz des Störterms nicht für alle Beobachtungen die selbe. Dieses Phänomen tritt typischerweise in Querschnittsuntersuchungen auf. Es gilt allerdings weiterhin, dass $E\varepsilon_t\varepsilon_s = 0$ für $t \neq s$:

2. Autokorrelation: $$\Omega = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{pmatrix}$$

Dieser Fall tritt häufig bei Zeitreihen auf. In diesem Fall tritt Autokorrelation zwischen den Residuen auf: $E\varepsilon_t\varepsilon_s \neq 0$ für manche $t \neq s$.

8.1 Schätzung mit OLS

Literatur: Greene (1997, 2000), Kapitel 11.2-11.2.2.

- Der OLS-Schätzer bleibt erwartungstreu

$$\hat{\beta} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'\varepsilon \quad \Rightarrow E\hat{\beta} = \beta$$

- Die Varianz des OLS-Schätzers ist:

$$V\hat{\beta} = (X'X)^{-1} X'(E\varepsilon\varepsilon')X(X'X)^{-1} = \sigma^2 (X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$

$$= \frac{\sigma^2}{T}\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'\Omega X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}$$

- Falls $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$, dann ist

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}(X'\Omega X)(X'X)^{-1}\right).$$

Die Standardtestverfahren sind nicht mehr anwendbar, da die Varianz des Schätzers nicht mehr durch $\sigma^2(X'X)^{-1}$ gegeben ist. Ausserdem ist s^2 unter Umständen ein verzerrter Schätzer von σ^2. $\sigma^2(X'X)^{-1}$ kann grösser oder kleiner als die wahre Varianz sein.

Falls sowohl $(X'X)/T$ als auch $(X'\Omega X)/T$ gegen positiv definite Matrizen konvergieren, dann ist der OLS-Schätzer weiterhin konsistent. Ausserdem bleibt der OLS-Schätzer asymptotisch normal verteilt, solange die Diagonalelemente von Ω endlich sind.

Der OLS-Schätzer ist allerdings nicht mehr effizient. Ein Problem besteht darin, dass Ω unbekannt ist $\Rightarrow$ Zwei Schritte:

8.2 Schätzung mit GLS bei bekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 11.3.1.

Da Ω eine symmetrische positiv definite Matrix ist, gilt:

$$\Omega = C\Lambda C' = C\Lambda^{\frac{1}{2}}\,\Lambda^{\frac{1}{2}}C' \text{ und daher ist}$$
$$\Omega^{-1} = C\Lambda^{-1}C' = \underbrace{C\Lambda^{\frac{-1}{2}}}_{P'}\;\underbrace{\Lambda^{\frac{-1}{2}}C'}_{P} = P'P$$

wobei: C = orthogonale Matrix, d.h. $C'C = CC' = I_T$, d.h. $C' = C^{-1}$

Λ = Diagonalmatrix mit positiven Eigenwerte, d.h. $\Lambda = \Lambda'$

Wir bilden nun die Transformation:

$$\tilde{y} = Py = P(X\beta + \varepsilon) = PX\beta + P\varepsilon = \tilde{X}\beta + \tilde{\varepsilon}$$

Dieses transformierte Regressionsmodell erfüllt alle Voraussetzungen des klassischen linearen Regressionsmodells, die Annahmen 1 bis 6 (Seite 21) sind wieder gültig. Insbesondere gilt:

$$V\tilde{\varepsilon} = E\tilde{\varepsilon}\tilde{\varepsilon}' = EP\varepsilon\varepsilon'P' = \sigma^2 P\Omega P' = \sigma^2 P(P'P)^{-1}P' = \sigma^2 I_T$$

Der verallgemeinerte OLS-Schätzer (GLS: "generalized least squares") oder Aitken-Schätzer lautet:

$$\boxed{\beta_{GLS} = \hat{\beta} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{y} = (X'P'PX)^{-1}X'P'Py = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y}$$

Bei Verwendung dieses Schätzers können alle bisher behandelten Testverfahren angewandt werden.

Unter der Annahme $\frac{\tilde{X}'\tilde{X}}{T} = \frac{X'\Omega^{-1}X}{T} \xrightarrow{T\to\infty} Q$, wobei Q eine endliche, symmetrische und positiv definite Matrix ist, gelten die folgenden Eigenschaften für den GLS-Schätzer.

- Der GLS-Schätzer ist konsistent:

$$\operatorname{plim}\hat{\beta} = \beta .$$

- Der GLS-Schätzer ist asymptotisch normal verteilt mit $E\hat{\beta} = \beta$ und $V\hat{\beta} = \sigma^2\left(\tilde{X}'\tilde{X}\right)^{-1} = \sigma^2\left(X'\Omega^{-1}X\right)^{-1}$:

$$\beta \to N[\beta, \sigma^2 (X'\Omega^{-1}X)^{-1}].$$

- Der GLS-Schätzer ist der beste erwartungstreue lineare Schätzer von β (Aitken 1935).

- Die üblichen Teststatistiken bleiben gültig; einzig das R^2 lässt sich nicht übertragen.

8.3 Schätzung mit FGLS bei unbekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 11.4.1.

Wegen der Symmetrie hat Ω im allgemeinen T(T+1)/2 unterschiedliche Matrixelemente. Typischerweise hängt Ω jedoch nur von ein paar wenigen Parametern θ ab:

$$\Omega = \Omega(\theta).$$

- Im Fall der Autokorrelation erster Ordnung z.B. ist θ = ρ und Ω ist

$$\Omega = \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}$$

- Liegt Heteroskedastizität z.B. der Form $\sigma_t^2 = \sigma^2 z_t^\alpha$ vor, wobei z_t eine bestimmte beobachtbare Variable ist, so ist θ = α und

$$\sigma^2\Omega = \sigma^2 \begin{pmatrix} z_1^\alpha & 0 & \cdots & 0 \\ 0 & z_2^\alpha & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_T^\alpha \end{pmatrix}$$

Angenommen wir hätten einen konsistenten Schätzer für θ, dann ersetzen wir Ω durch $\hat{\Omega} = \Omega(\hat{\theta})$ und bilden den "feasible GLS":

$$\hat{\theta} \xrightarrow{T \to \infty} \theta \;; \qquad \Rightarrow \; \hat{\Omega} \xrightarrow{T \to \infty} \Omega$$

$$\beta_{FGLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}y$$

β_{FGLS} ist asymptotisch äquivalent zu β_{GLS} falls

$$\operatorname{plim} \frac{X'\hat{\Omega}^{-1}X}{T} = \operatorname{plim} \frac{X'\Omega^{-1}X}{T}$$

und

$$\operatorname{plim} \frac{1}{\sqrt{T}} X'\hat{\Omega}^{-1}\varepsilon = \operatorname{plim} \frac{1}{\sqrt{T}} X'\Omega^{-1}\varepsilon .$$

Der FGLS hat die selben Eigenschaften wie der GLS; man benötigt lediglich einen konsistenten (nicht jedoch effizienten Schätzer) von θ. Alle Eigenschaften des GLS bleiben erhalten.

9 Heteroskedastizität

Literatur: Wooldridge (2003), Kapitel 12 oder Greene (1997, 2000), Kapitel 12.1.

Bei Querschnittsdaten tritt oft das Problem heteroskedastischer Störungen auf.

Modell: $$y = X\beta + \varepsilon,$$

$$E\varepsilon = 0,$$

$$E\varepsilon_t\varepsilon_t = V\varepsilon_t = \sigma_t^2 = \sigma^2\omega_t,$$

$$E\varepsilon_t\varepsilon_s = Cov(\varepsilon_t, \varepsilon_s) = 0 \text{ für } t \neq s,$$

$$V\varepsilon = E\varepsilon\varepsilon' = \sigma^2\Omega = \sigma^2 \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_T \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{pmatrix}$$

9.1 Schätzung mit OLS

Literatur: Greene (1997, 2000), Kapitel 12.2.

Vgl. auch Abschnitt 8.1. Der OLS-Schätzer ist auch bei heteroskedastischem Störterm erwartungstreu, jedoch nicht mehr effizient. Der OLS-Schätzer bleibt konsistent und (asymptotisch) normalverteilt. Die Standardtestverfahren sind nicht mehr anwendbar, da die Varianz des Schätzers nicht mehr durch $\sigma^2(X'X)^{-1}$ gegeben ist.

9.2 Schätzung der Varianz des OLS-Schätzers

Literatur: Greene (1997, 2000), Kapitel 12.2.3.

Anders als bei der Standard OLS-Schätzung ist die Varianz des Schätzers gegeben durch

$$V\hat{\beta}_{OLS} = (X'X)^{-1}\left[X'\sigma^2\Omega X\right](X'X)^{-1}.$$

Unter sehr allgemeinen Bedingungen, und ohne dass die tatsächliche Form der Heteroskedastizität bekannt sein muss, kann die Matrix

$$\Sigma = \frac{X'\sigma^2\Omega X}{T} = \frac{1}{T}\sum_t \sigma_t^2 x_t x_t'$$

durch $S_0 = \frac{1}{T}\sum_t e_t^2 x_t x_t'$ geschätzt werden. Der sogenannte <u>White-Schätzer</u> für die Varianz von $\hat{\beta}_{OLS}$ ist dann

$$\hat{V}\hat{\beta}_{OLS} = T(X'X)^{-1}S_0(X'X)^{-1}.$$

Aus der so geschätzten Varianz lassen sich heteroskedastizitäts-robuste Standardabweichungen für die Koeffizienten berechnen. Wie in Abschnitt 7 beschrieben, sind die t- und F-Statistiken asymptotisch zu interpretieren.

9.3 Test auf Heteroskedastizität

Literatur: Greene (1997, 2000), Kapitel 12.3-12.3.1.

White's Test auf Heteroskedastizität beruht auf dem Vergleich von

$$\underbrace{s^2(X'X)^{-1}}_{\substack{\text{Varianz-Kovarianz-Matrix bei}\\ \text{Fehlen von Heteroskedastizität}}} \qquad \text{und} \qquad \underbrace{(X'X)^{-1}\left(\sum_t e_t^2 x_t x_t'\right)(X'X)^{-1}}_{\substack{\text{Varianz-Kovarianz-Matrix bei Vorhandensein}\\ \text{von Heteroskedastizität}}}$$

Die Nullhypothese lautet: es besteht keine Heteroskedastizität. Implementiert wird der Test, indem man die quadrierten OLS-Residuen auf alle Kreuzprodukte der Regressoren regressiert. Doppelte Regressoren werden aus Gründen der Multikollinearität nur einmal berücksichtigt. Die Grösse TR^2 (Anzahl der Beobachtungen multipliziert mit dem Bestimmtheitsmass der White-Regression) ist χ^2 mit L Freiheitsgraden verteilt. L bezeichnet dabei die Anzahl der Regressoren ohne Berücksichtigung der Konstanten. Die Nullhypothese wird abgelehnt, wenn $TR^2 > \chi^2(L)$ ist.

Der Test von White gibt allerdings keine Auskunft über die Form der Heteroskedastizität.

9.4 Schätzung mit GLS bei bekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 12.4.

Der GLS-Schätzer ist nach Abschnitt 8.2:

$$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y,$$
$$V\hat{\beta}_{GLS} = \sigma^2\left(X'\Omega^{-1}X\right)^{-1},$$

wobei hier

$$\Omega^{-1} = \begin{pmatrix} 1/\omega_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\omega_T \end{pmatrix} = \operatorname{diag}\left(\frac{1}{\omega_1}, \frac{1}{\omega_2}, \cdots, \frac{1}{\omega_T}\right) = P'P,$$

$$P = P' = \operatorname{diag}\left(\frac{1}{\sqrt{\omega_1}}, \frac{1}{\sqrt{\omega_2}}, \cdots, \frac{1}{\sqrt{\omega_T}}\right).$$

Die Gewichte ω_t und die Varianz σ^2 sind nicht eindeutig, sondern nur bis auf einen Proportionalitätsfaktor bestimmt.

Auch dieser GLS-Schätzer kann als OLS-Schätzer nach einer Transformation des ursprünglichen Modells gesehen werden:

$$\tilde{y} = Py = \begin{pmatrix} y_1/\sqrt{\omega_1} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{pmatrix}, \tilde{X} = PX = \begin{pmatrix} x_1'/\sqrt{\omega_1} \\ \vdots \\ x_T'/\sqrt{\omega_T} \end{pmatrix} = \begin{pmatrix} x_{11}/\sqrt{\omega_1} & \cdots & x_{1K}/\sqrt{\omega_1} \\ \vdots & \ddots & \vdots \\ x_{T1}/\sqrt{\omega_T} & \cdots & x_{TK}/\sqrt{\omega_T} \end{pmatrix},$$

wobei der Zeilenvektor x_t' die t-te Zeile in der Matrix X ist. Der OLS-Schätzer der transformierten Variablen, $\tilde{y}$ und $\tilde{X}$, ist gleich dem obigen GLS-Schätzer:

$$\hat{\beta}_{GLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = (X'P'PX)^{-1}X'P'Py = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$
$$= \left(\sum_t \frac{1}{\omega_t} x_t x_t'\right)^{-1}\left(\sum_t \frac{1}{\omega_t} x_t y_t\right) = \left(\sum_t w_t x_t x_t'\right)^{-1}\left(\sum_t w_t x_t y_t\right)$$

mit den Gewichten $w_t = 1/\omega_t$. Grosse Varianzen ω_t haben kleine Gewichte w_t. Dieser Schätzer wird gewichteter OLS-Schätzer genannt (WLS: weighted least squares).

9.5 Schätzung mit FGLS bei unbekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 12.5-12.5.1.

Der GLS-Schätzer bei heteroskedastischem Störterm ist wiederum:

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \left(\sum_t \frac{1}{\sigma_t^2} x_t x_t'\right)^{-1}\left(\sum_t \frac{1}{\sigma_t^2} x_t y_t\right).$$

mit x_t' Zeilenvektor der t-ten Beobachtung. Aber σ_t^2 ist unbekannt.

Beispiele für Modelle der Heteroskedastizität:

$$\sigma_t^2(\alpha, z_t) = \sigma^2 z_t^\alpha \quad \Rightarrow \ln\sigma_t^2 = \ln\sigma^2 + \alpha \ln z_t$$

$$\sigma_t^2(\alpha, z_t) = \sigma^2(\alpha' z_t)$$

$$\sigma_t^2(\alpha, z_t) = \sigma^2(\alpha' z_t)^2$$

Zwei-stufiges Schätzverfahren:

In der ersten Stufe wendet man OLS an. Aus den OLS-Residuen wird dann auf die Form der Heteroskedastizität geschlossen oder ein Modell für die Heteroskedastizität geschätzt. Dies liefert eine Schätzung für Ω, welche in der zweiten Stufe für die Berechnung des

FGLS-Schätzers verwendet wird. Dies stellt ein zulässiges Verfahren dar, da OLS nach wie vor konsistent ist, so dass die OLS-Residuen

$$e_t = y_t - x_t'\hat{\beta}_{OLS} = y_t - x_t'\beta + x_t'\left(\beta - \hat{\beta}_{OLS}\right) = \varepsilon_t + x_t'\left(\beta - \hat{\beta}_{OLS}\right)$$

die selbe asymptotische Verteilung wie die wahren Residuen ε_t haben. Insbesondere gilt:

$$e_t^2 = \varepsilon_t^2 + \left(x_t'\left(\beta - \hat{\beta}_{OLS}\right)\right)^2 + 2\varepsilon_t x_t'\left(\beta - \hat{\beta}_{OLS}\right).$$

Da $E\varepsilon_t^2 = \sigma_t^2$, kann z. B. das Regressionsmodell

$$e_t^2 = \sigma_t^2 + v_t = \alpha' z_t + v_t$$

verwendet werden, um α und somit auch Ω zu schätzen. Dies entspricht einer Regression der quadrierten OLS-Residuen auf z_t. Der Störterm

$$v_t = E\left(x_t'\left(\beta - \hat{\beta}_{OLS}\right)\right)^2 + E\left(2\varepsilon_t x_t'\left(\beta - \hat{\beta}_{OLS}\right)\right)$$

hat jedoch nicht Mittelwert null (dieses Problem kann durch die Berücksichtigung einer Konstanten behoben werden), ist heteroskedastisch und autokorreliert (da vom selben β_{OLS} abhängig). Da die Autokorrelation asymptotisch verschwindet ist der Schätzer dennoch konsistent.

Nun wird der FGLS-Schätzer mit Hilfe von $\hat{\Omega} = \Omega(\hat{\alpha})$ berechnet:

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Omega}(\hat{\alpha})^{-1}X\right)^{-1} X'\hat{\Omega}(\hat{\alpha})^{-1} y = \left(\sum_t \frac{1}{\hat{\sigma}_t^2(\hat{\alpha})} x_t x_t'\right)^{-1} \left(\sum_t \frac{1}{\hat{\sigma}_t^2(\hat{\alpha})} x_t y_t\right).$$

Einschub: Begriffe der Zeitreihenanalyse

Literatur: Wooldridge (2003), Kapitel 11 oder Greene (1997, 2000), Kapitel 13.2, 13.3.1.

DEFINITION: Ist $\{\varepsilon_t\}$ eine (zeitlich geordnete) Folge von Zufallsvariablen, dann heisst die Funktion $\gamma_{t,s}$ die Autokovarianzfunktion von $\{\varepsilon_t\}$. Sie ist definiert durch:

$$\gamma_{t,s} = \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = E[(\varepsilon_t - E\varepsilon_t)(\varepsilon_s - E\varepsilon_s)]$$

DEFINITION: Eine (zeitlich geordnete) Folge von Zufallsvariablen $\{\varepsilon_t\}$ heisst stationär, falls für alle r,s und t gilt:

(1) $E\varepsilon_t = m$; unabhängig vom Zeitpunkt konstanter Mittelwert

(2) $V\varepsilon_t = V\varepsilon_s = \gamma_0 < \infty$; endliche, konstante Varianz

(3) $\mathrm{Cov}(\varepsilon_t,\varepsilon_s) = \mathrm{Cov}(\varepsilon_{t+r},\varepsilon_{s+r}) = \gamma_{t-s}$

Bedingung (3) impliziert (2).

BEMERKUNG: Falls $\{\varepsilon_t\}$ stationär ist, kann die Autokovarianzfunktion geschrieben werden als

$$\gamma_h = \gamma_{t-s} = \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \gamma_{-h},$$

$|h|$ wird als Verzögerung ("lag") oder Ordnung ("order") bezeichnet.

Häufig wird anstelle der Autokovarianzfunktion die Autokorrelationsfunktion betrachtet:

$$\rho_h = \rho_{t-s} = \frac{\gamma_{t-s}}{\gamma_0} = \mathrm{Corr}(\varepsilon_t, \varepsilon_s) \qquad h = 0, \pm 1, \pm 2 \cdots.$$

10 Autokorrelation

Literatur: Wooldridge (2003), Kapitel 12.1 oder Greene (1997, 2000), Kapitel 13.1.

Bei Zeitreihen haben wir es meist mit homoskedastischen Störungen ($E\varepsilon_t^2 = \sigma^2$) zu tun. Diese sind jedoch oft über die Beobachtungen hinweg korreliert. Wir nehmen im Folgenden an, dass der Störterm eine <u>stationäre</u> Folge von Zufallsvariablen ist.

Modell: $$y = X\beta + \varepsilon,$$

$$E\varepsilon = 0,$$

$$E\varepsilon_t\varepsilon_t = V\varepsilon_t = \sigma^2 = \gamma_0,$$

$$E\varepsilon_t\varepsilon_s = \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \gamma_{t-s} = \gamma_0\rho_{t-s} \neq 0,$$

$$V\varepsilon = E\varepsilon\varepsilon' = \sigma^2\Omega = \gamma_0\Omega = \sigma^2\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{pmatrix}$$

Prototyp: Autoregressiver Prozess erster Ordnung AR(1)

Literatur: Wooldridge (2003), Kapitel 12.2 oder Greene (1997, 2000), Kapitel 13.3.2.

Bei Autokorrelation erster Ordnung hängt ε_t direkt nur von ε_{t-1} ab:

$$\varepsilon_t = \phi\varepsilon_{t-1} + u_t \ ,$$

mit $Eu_t = 0$, $Vu_t = \sigma_u^2 < \infty$, $|\phi| < 1$ und $Eu_t u_s = 0$ für $t \neq s$ und $Eu_t\varepsilon_s$ für $s < t$. Für $|\phi| < 1$ ist $\{\varepsilon_t\}$ stationär.

Bei Homoskedastizität und Stationarität gilt $V\varepsilon_t = V\varepsilon_{t-1}$ und damit

$$V\varepsilon_t = \phi^2 V\varepsilon_{t-1} + Vu_t \Rightarrow V\varepsilon_t = \sigma^2 = \frac{\sigma_u^2}{1-\phi^2} .$$

Herleitung der Kovarianz der Störterme. Multiplikation der autoregressiven Beziehung durch ε_{t-1} und anschliessender Erwartungsbildung ergibt:

$$E\varepsilon_t\varepsilon_{t-1} = \phi E\varepsilon_{t-1}\varepsilon_{t-1} + Eu_t\varepsilon_{t-1} \Rightarrow \gamma_1 = \phi\gamma_0 = \phi \times \frac{\sigma_u^2}{1-\phi^2}$$

Ebenso ergibt die Multiplikation mit ε_{t-2}:

$$\gamma_2 = E\varepsilon_t\varepsilon_{t-2} = \phi E\varepsilon_{t-1}\varepsilon_t + Eu_t\varepsilon_{t-2} = \phi\gamma_1 = \phi^2\gamma_0 = \phi^2 \times \frac{\sigma_u^2}{1-\phi^2}$$

Daher gilt allgemein: $\gamma_h = E\varepsilon_t\varepsilon_{t-h} = \phi^h\gamma_0 = \dfrac{\phi^s}{1-\phi^2}\sigma_u^2$.

Damit ist die Varianz-Kovarianzmatrix des Störterms

$$E\varepsilon\varepsilon' = \sigma^2\Omega = \frac{\sigma_u^2}{1-\phi^2} \underbrace{\begin{pmatrix} 1 & \phi & \cdots & \phi^{T-1} \\ \phi & 1 & \cdots & \phi^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \cdots & 1 \end{pmatrix}}_{\Omega}$$

10.1 Schätzung mit OLS

Literatur: Wooldridge (2003), Kapitel 12 oder Greene (1997, 2000), Kapitel 13.4.

Vgl. auch Abschnitt 8.1. Der OLS-Schätzer bleibt auch bei Autokorrelation der Störterme erwartungstreu (Ausnahme siehe Spezialfall unten). OLS ist aber nicht mehr effizient. Falls $\lim_{T\to\infty} \frac{X'\Omega X}{T} = Q^*$ positiv definit, dann ist OLS konsistent. Die asymptotische Normalität bleibt unter sehr allgemeinen Bedingungen erhalten. Die Standardtestverfahren sind nicht mehr anwendbar, da die Varianz des Schätzers nicht mehr durch $\sigma^2(X'X)^{-1}$ gegeben ist.

Spezialfall: OLS-Schätzung mit verzögerten Variablen

Literatur: Greene (1997, 2000), Kapitel 13.4.1.

Betrachte ein Regressionsmodell mit verzögerter endogener Variable als Regressor und Autokorrelation in den Residuen:

$$y_t = \beta y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = \phi \varepsilon_{t-1} + u_t$$

wobei $|\beta| < 1$ und $|\phi| < 1$. Daraus folgt:

$$
\begin{aligned}
\operatorname{cov}(y_{t-1},\varepsilon_t) &= \operatorname{cov}(y_{t-1},\phi\varepsilon_{t-1}+u_t)=\phi\operatorname{cov}(y_{t-1},\varepsilon_{t-1})=\phi\operatorname{cov}(y_t,\varepsilon_t)\\
&= \phi\beta\operatorname{cov}(y_{t-1},\varepsilon_t)+\phi\underbrace{\operatorname{cov}(\varepsilon_t,\varepsilon_t)}_{=V\varepsilon_t=\sigma_u^2/1-\phi^2}\\
&= \frac{\phi V\varepsilon_t}{1-\phi\beta}=\frac{\phi\sigma_u^2}{(1-\phi\beta)(1-\phi^2)}\neq 0 \quad \text{für } \phi\neq 0
\end{aligned}
$$

d.h. die Orthogonalitätsannahme für OLS ist verletzt. Der OLS-Schätzer ist deshalb weder erwartungstreu noch konsistent:

$$
\operatorname{plim}\hat{\beta}=\beta+\frac{\operatorname{cov}(y_{t-1},\varepsilon_t)}{Vy_{t-1}}\neq\beta .
$$

10.2 Schätzung der Varianz des OLS-Schätzers

Literatur: Greene (1997, 2000), Kapitel 13.4.3.

Anders als bei der Standard OLS-Schätzung ist die Varianz des Schätzer gegeben durch

$$
V\hat{\beta}_{OLS}=(X'X)^{-1}\left[X'\sigma^2\Omega X\right](X'X)^{-1}.
$$

Unter sehr allgemeinen Bedingungen, und ohne dass die tatsächliche Form der Autokorrelation bekannt sein muss, kann die Matrix

$$
\Sigma=\frac{X'\sigma^2\Omega X}{T}=\frac{1}{T}\sum_t\sum_s\gamma_{|t-s|}\left(x_t x_s'+x_s x_t'\right)
$$

durch

$$S = S_0 + \frac{1}{T}\sum_{j=1}^{L}\sum_{t=j+1}^{T} w_j e_t e_{t-j}\left(x_t x'_{t-j} + x_{t-j}x'_t\right) \quad \text{mit } S_0 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t x'_t$$

geschätzt werden. Die Gewichte w_j sind durch $w_j = 1 - j/(L+1)$ gegeben. Sie sind nicht mit den Gewichten in der gewichteten OLS Regression zu verwechseln. Ein Problem stellt die geeignete Wahl von L dar.

Der sogenannte Newey-West-Schätzer oder –Korrektur der Varianz von $\hat{\beta}_{OLS}$ ist dann:

$$\hat{V}\hat{\beta}_{OLS} = T(X'X)^{-1} S (X'X)^{-1}.$$

Aus der so geschätzten Varianz lassen sich autokorrelations- (und heteroskedastizitäts-) robuste Standardabweichungen für die Koeffizienten berechnen. Wie in Abschnitt 7 beschrieben, sind die t- und F-Statistiken asymptotisch zu interpretieren.

10.3 Tests auf Autokorrelation

Durbin-Watson Test

Literatur: Wooldridge (2003), Kapitel 12.2 oder Greene (1997, 2000), Kapitel 13.5.1.

Durbin-Watson Teststatistik:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = 2\left(1 - \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}\right) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2} \approx 2(1 - \rho_1)$$

- $d < 2$: positive Autokorrelation $\Rightarrow$ $0 < \rho_1 \leq 1$
- $d > 2$: negative Autokorrelation $\Rightarrow$ $-1 \leq \rho_1 < 0$

Test auf positive Autokorrelation (positive Autokorrelation ist bei ökonomischen Zeitreihen häufiger als negative):

H_0: $\rho = 0$; keine Autokorrelation

Die Nullhypothese wird abgelehnt, falls $d < d^*$. Die Verteilung der Teststatistik hängt von der Regressormatrix X ab. Deshalb wird nur ein "approximativer" Test verwendet.

H_0 wird abgelehnt, falls $d < d^L$; (d^L = untere Schranke, d = Testwert)

H_0 wird nicht abgelehnt, falls $d > d^U$; (d^U = obere Schranke, d = Testwert)

Für $d^L < d < d^U$ ist keine Entscheidung möglich, weitere Analysen sind hier notwendig.

Der DW-Test ist relativ mächtig, falls die Störungen einem AR(1)-Prozess folgen (plim d = $2(1-\rho_1)$). Liegt jedoch ein AR-Prozess höherer Ordnung vor, so liefern alternative Testverfahren unter Umständen bessere Ergebnisse.

Tests für autoregressive Prozesse höherer Ordnung

Literatur: Greene (1997, 2000), Kapitel 13.5.2.

Breusch-Godfrey (1978) Test:

H_0: keine Autokorrelation

H_1: Autokorrelation p-ter Ordnung

Regressiere e_t auf x_t, $e_{t-1}, \ldots, e_{t-p}$, wobei die Residuen $e_0, e_{-1}, \ldots, e_{-p+1}$ gleich Null gesetzt werden.

$$T\,R^2 \overset{A}{\sim} \chi^2 \text{ mit p Freiheitsgraden}$$

Für $T\,R^2 < \chi^2(p)$ wird H_0 nicht abgelehnt.

Box-Pierce Statistik:

H_0: keine Autokorrelation

Die Teststatistik $Q = T\sum_{j=1}^{L}\hat{\rho}_j^2 \sim \chi_L^2$, wobei $\hat{\rho}_j = \dfrac{\sum_{t=j+1}^{T} e_t e_{t-j}}{\sum_{t=1}^{T} e_t^2}$ der empirische Autokorrelationskoeffizient der Ordnung j ist.

Für $Q < \chi^2(L)$ wird H_0 nicht abgelehnt.

Verfeinerte Version: Ljung-Box-Statistik:

$$Q' = T(T+2)\sum_{j=1}^{L}\frac{\hat{\rho}_j^2}{T-j} \sim \chi_L^2$$

Ein Problem stellt wiederum die Wahl von L dar.

10.4 Schätzung mit GLS bei bekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 13.6.1.

Der GLS-Schätzer ist nach Abschnitt 8.2:

$$\hat{\beta}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y$$

$$V\hat{\beta}_{GLS} = \sigma^2\left(X'\Omega^{-1}X\right)^{-1}$$

Auch hier kann der GLS-Schätzer als OLS-Schätzung des transformierten ursprünglichen Modells gesehen werden. Falls die Störungen einem autoregressiven Prozess erster Ordnung, AR(1), folgen, ist die Transformationsmatrix P gegeben durch:

$$P = \begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{pmatrix}$$

mit $P'P = \Omega^{-1}$. Durch Transformation ergibt sich:

$$\tilde{y} = Py = \begin{pmatrix} \sqrt{1-\phi^2}\, y_1 \\ y_2 - \phi y_1 \\ \vdots \\ y_T - \phi y_{T-1} \end{pmatrix}, \tilde{X} = PX = \begin{pmatrix} \sqrt{1-\phi^2}\, x_1' \\ x_2' - \phi x_1' \\ \vdots \\ x_T' - \phi x_{T-1}' \end{pmatrix}, \tilde{\varepsilon} = P\varepsilon = \begin{pmatrix} \sqrt{1-\phi^2}\, \varepsilon_1 \\ \varepsilon_2 - \phi \varepsilon_1 \\ \vdots \\ \varepsilon_T - \phi \varepsilon_{T-1} \end{pmatrix}$$

Das transformierte Modell $\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}$ kann mittels OLS effizient geschätzt werden.

Bei autoregressiven Modellen höherer Ordnung kann man analog vorgehen.

10.5 Schätzung mit FGLS bei unbekanntem Ω

Literatur: Greene (1997, 2000), Kapitel 13.7.1.

OLS ist nach wie vor konsistent. Man kann daher die OLS-Residuen zur Schätzung der Autokorrelation verwenden.

Zur Schätzung von β wird ein mehrstufiges Verfahren angewandt:

1. Schätzung mittels OLS liefert die Residuen e_t.
2. Mit Hilfe dieser Residuen wird in ein Schätzer für ϕ konstruiert, was $\hat{\phi}$ ergibt.
3. Mittels $\hat{\phi}$ wird $\hat{\Omega} = \Omega(\hat{\phi})$ konstruiert.
4. Aus $\Omega(\hat{\phi})$ wird die Transformationsmatrix P ermittelt und damit das transformierte Modell $\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon}$ generiert.
5. Die OLS-Schätzung des transformierten Modells ergibt den FGLS-Schätzer $\hat{\beta}_{FGLS}$.

Im Fall, dass die Störungen einem autoregressiven Prozess erster Ordnung, AR(1), folgen, werden verschiedene Schätzer für ϕ vorgeschlagen:

- $$\hat{\phi} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e_t^2}$$

- $$\hat{\phi} = \frac{T-K}{T-1} \times \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e_t^2}$$

- $$\hat{\phi} = 1 - \frac{d}{2} \;;\quad \text{mit } d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \approx 2(1 - r_1)$$

- Cochrane-Orcutt: unrestringierte Regression:

  $$y_t = \phi y_{t-1} + \beta' x_t - (\phi\beta') x_{t-1} + u_t$$

Für autoregressive Prozesse höherer Ordnung kann man analog vorgehen.