5. Autokorrelation und Heteroskedastizität

(Maddala Kapitel 5 und 6)

Datengenerierender Prozess:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + u_i \qquad u_i \sim IN(0, \sigma^2)$$

Geschätzte Regressionsgleichung:

$$y_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{u}_i$$

Schätzung der Parameter $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \ldots$ durch Minimierung von

$$RSS = \sum_{i=1}^{T} \hat{u}_i^2$$

Annahmen im bezüglich des Störterms:

1) $E(u_i) = 0$

2) $\text{var}(u_i) = E(u_i^2) = \sigma^2$ ist konstant.

3) u_i und u_j sind unabhängig für alle $i \neq j$.

4) u_i und x_j sind unabhängig für alle i und j.

5) u_i ist normalverteilt.

2) ist die Homoskedastizitätsannahme. Falls 2) verletzt ist, spricht man von ***Heteroskedastizität.***

3) ist die Annahme zeitlich unabhängiger Störterme. Falls 3) verletzt ist, spricht man von ***Autokorrelation.***

Falls 4) verletzt ist, handelt es sich bei einzelnen oder mehreren *x*-Variablen um ***endogene Variablen.***

Nur wenn alle Annahmen 1) bis 5) erfüllt sind, sind die Parameterschätzungen, die man durch Minimierung von RSS erhält, *unverzerrt* (= *erwartungstreu*), *konsistent* und *effizient.*

Auswirkungen von Heteroskedastizität und Autokorrelation auf die Parameterschätzungen. Tests und modifizierte Schätzmethoden.

$$y_t = \alpha + \beta_1 x_{1_t} + \beta_2 x_{2_t} + \ldots + \gamma y_{t-1} + u_t$$

	Heteroskedastizität	*Autokorrelation*	
		$\gamma = 0$	$\gamma \neq 0$
Parameterschätzung	unverzerrt nicht effizient	unverzerrt	verzerrt
		nicht effizient	
Standardfehler *t-Werte*	verzerrt verzerrt	unterschätzt überschätzt	
Testverfahren	Goldfeld/Quandt White	Durbin/Watson	Durbin-h
Angepasste Schätzmethode	"Weighted Least Squares"	Cochrane/Orcutt	

Heteroskedastizität

Test von Goldfeld und Quandt:

Man unterteilt die Stichprobe mit n Beobachtungen in zwei Teile 1 und 2 mit n_1 und n_2 Beobachtungen (aufgrund einer grafischen Inspektion der Residuen) und schätzt dafür separate Regressionen. Unter der Null-Hypothese einer konstanten Störterm-Varianz (Homoskedastizität) ist das Verhältnis der geschätzten Störterm-Varianzen (grössere im Zähler) F-verteilt mit (n_2 - k) und (n_1 - k) Freiheitsgraden (es ist angenommen, $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$):

$$F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}$$

Falls der berechnete F-Wert grösser ist als der kritische Wert gemäss F-Tabelle, ist die Nullhypothese (Homoskedastizität) zugunsten der Alternativhypothese (Heteroskedastizität) zu verwerfen.

Test von White:

Man quadriert die berechneten Residuen (als Mass für die Störterm-Varianz) und regressiert sie auf die erklärenden Variablen, deren Quadrate und möglicherweise deren Kreuzprodukte, z.B.:

$$\hat{u}^2 = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_2 + a_4 x_2^2 + a_5 x_1 x_2$$

Man testet die Nullhypothese

$$a_1 = a_2 = a_3 = a_4 = a_5 = 0$$

(kein Zusammenhang zwischen der Störterm-Varianz und den erklärenden Variablen)

mit einem F-Test. Fällt der berechnete F-Wert signifikant aus, so ist die Nullhypothese (Homoskedastizität) zugunsten der Alternativhypothese (Heteroskedastizität) zu verwerfen.

Deuten die Tests von Goldfeld/Quandt oder White auf heteroskedastische Störterme hin, so ist eine OLS-Schätzung ineffizient. Grund: Bei der Berechnung der Regressionsparameter sollten Beobachtungen mit grosser Varianz weniger stark gewichtet werden als Beobachtungen mit kleiner Varianz. => "Weighted Least Squares".

Der Gewichtungsfaktor ergibt sich aus der vermuteten Form der Heteroskedastizität.

Zwei einfache Fälle:

1) Die Störterm-Varianz in der Gleichung

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

ist proportional zum Quadrat der erklärenden Variablen, d.h.

$$\sigma_t^2 = k\, x_t^2$$

In diesem Fall ist der Gewichtungsfaktor durch x_t gegeben. Die entsprechend transformierte Gleichung lautet:

$$\frac{y_t}{x_t} = \alpha \frac{1}{x_t} + \beta + v_t$$

wobei $v_t = u_t / x_t$ eine konstante Varianz k ausweist.

2) Die Störterm-Varianz in der Gleichung

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

ist proportional zum Quadrat einer Variablen z_t, d.h.

$$\sigma_t^2 = k\, z_t^2$$

In diesem Fall ist der Gewichtungsfaktor durch z_t gegeben. Die entsprechend transformierte Gleichung lautet:

$$\frac{y_t}{z_t} = \alpha \frac{1}{z_t} + \beta \frac{x_t}{z_t} + v_t$$

wobei $v_t = u_t / z_t$ eine konstante Varianz k ausweist.

Beispiel: Schätzung einer Konsumfunktion für die Schweiz mit Quartalsdaten 1970 – 1995:

CONSR: Reale Konsumausgaben der Haushalte

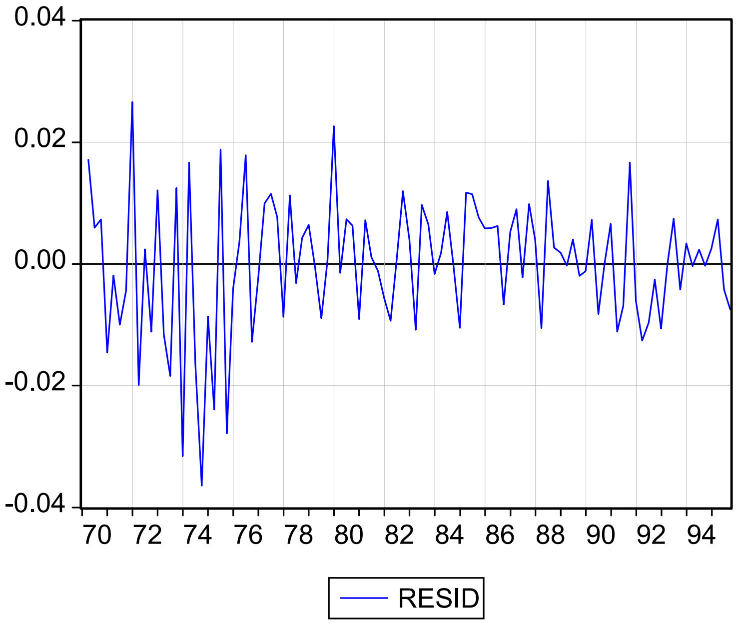
YDISPR: Real verfügbares Haushaltseinkommen

Sample(adjusted): 1970:2 1995:4
Included observations: 103 after adjusting endpoints

LOG(CONSR) = C(1)+C(2)*LOG(YDISPR)+C(3)*LOG(CONSR(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.360179	0.107498	3.350570	0.0011
C(2)	0.103231	0.035076	2.943079	0.0040
C(3)	0.861032	0.040037	21.50574	0.0000

R-squared	0.989715	Adjusted R-squared	0.989509
S.D. dependent var	0.108183	S.E. of regression	0.011080
Sum squared resid	0.012278	F-statistic	4811.500
Durbin-Watson stat	2.309926	Prob(F-statistic)	0.000000



White Heteroskedasticity Test:

F-statistic 4.443318 Probability 0.002420

Test Equation:
LS // Dependent Variable is RESID^2
Sample: 1970:2 1995:4
Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001722	0.002529	-0.681016	0.4975
YDISPR	-2.67E-07	2.31E-07	-1.157956	0.2497
YDISPR^2	5.32E-12	3.82E-12	1.393549	0.1666
CONSR(-1)	4.65E-07	3.14E-07	1.482004	0.1415
CONSR(-1)^2	-1.02E-11	6.06E-12	-1.679379	0.0963

R-squared	0.153518	Adjusted R-squared	0.11896
S.D. dependent var	0.000205	S.E. of regression	0.000193
Sum squared resid	3.64E-06	F-statistic	4.443318
Durbin-Watson stat	1.971277	Prob(F-statistic)	0.002420

Testergebnis: Die Nullhypothese (Homoskedastische Störterme) wird zugunsten der Alternativhypothese (Heteroskedastizität) verworfen. Beobachtetes Signifikanzniveau P = 0.24%.

Test von Goldfeld und Quandt:

Die Regression wird separat für die Perioden 1970:2 bis 1979:4 und 1980:1 bis 1995:4 geschätzt:

Sample: 1970:2 1979:4
Included observations: 39 after adjusting endpoints
LOG(CONSR) = C(1)+C(2)*LOG(YDISPR)+C(3)*LOG(CONSR(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.220188	0.680052	1.794255	0.0812
C(2)	0.057230	0.099485	0.575270	0.5687
C(3)	0.822164	0.070863	11.60215	0.0000

R-squared	0.899812	Adjusted R-squared	0.89424
S.D. dependent var	0.045592	**S.E. of regression**	**0.014826**
Sum squared resid	0.007914	F-statistic	161.6626
Durbin-Watson stat	2.508804	Prob(F-statistic)	0.000000

Sample: 1980:1 1995:4
Included observations: 64
LOG(CONSR) = C(1)+C(2)*LOG(YDISPR)+C(3)*LOG(CONSR(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.188883	0.327838	3.626437	0.0006
C(2)	0.221355	0.069119	3.202502	0.0022
C(3)	0.661485	0.099636	6.639042	0.0000

R-squared	0.989215	Adjusted R-squared	0.988862
S.D. dependent var	0.068572	**S.E. of regression**	**0.007237**
Sum squared resid	0.003195	F-statistic	2797.610
Durbin-Watson stat	1.824550	Prob(F-statistic)	0.000000

Der berechnete F-Wert ist $F = 0.014826^2 / 0.007237^2 = 4.197$.

Der kritische F-Wert (5%-Niveau, 36 und 61 Freiheitsgrade) ist ungefähr 1.6. Die Nullhypothese einer konstanten Störterm-Varianz ist somit ebenfalls abzulehnen.

Autokorrelation

Durbin-Watson-Test

Die Durbin-Watson-Teststatistik wird auf Basis der Residuen einer geschätzten Gleichung nach folgender Formel berechnet:

$$DW = \frac{\sum_{i=2}^{n} \left(\hat{u}_i - \hat{u}_{i-1}\right)^2}{\sum_{i=1}^{n} \hat{u}_i^2}$$

Approximativ gilt folgende Beziehung zwischen der *DW*-Statistik und *r*(1), dem Autokorrelationskoeffizienten erster Ordnung der Residuen:

$$DW = \frac{\sum_{i=2}^{n} \hat{u}_i^2 + \sum_{i=2}^{n} \hat{u}_{i-1}^2 - 2\sum_{i=2}^{n} \hat{u}_i \hat{u}_{i-1}}{\sum_{i=1}^{n} \hat{u}_i^2}$$

$$\cong 2 - 2\,\frac{\sum_{i=2}^{n} \hat{u}_i \hat{u}_{i-1}}{\sum_{i=1}^{n} \hat{u}_i^2} \quad = 2 - 2r(1)$$

oder: $r(1) \cong 1 - DW/2$

Eigenschaften:

1) $0 \leq DW \leq 4$ $\quad r(1) \rightarrow 1 \Longleftrightarrow DW \rightarrow 0 \quad r(1) \rightarrow -1 \Longleftrightarrow DW \rightarrow 4$

2) Unter der Nullhypothese, dass u_i zeitlich unabhängig verteilt (also nicht autokorreliert) ist, gilt (in grossen Stichproben):

$$E(DW) = 2$$

3) $DW < 2$ deutet auf positive Autokorrelation hin, $DW > 2$ deutet auf negative Autokorrelation hin. Kritische Werte sind tabelliert. Sie hängen vom Stichprobenumfang n und der Anzahl erklärender Variablen k ab. Beispiel: Für $n = 65$ und $k = 3$ sind die kritischen Werte $d_L = 1.24$ und $d_U = 1.65$. Falls $DW < d_L$, ist die Nullhypothese "keine Autokorrelation" zu verwerfen. Falls $DW > d_U$ ist die Null-hypothese zu akzeptieren. Fall $d_L < DW < d_U$ ist der Test unschlüssig.

Autokorrelations-Funktion in EViews

Den Autokorrelationskoeffizienten 1. Ordnung kann man aus dem Korrelogram der Residuen entnehmen. Ein Korrelogram stellt die Autokorrelation (AC) *r(i)* und die partielle Autokorrelation (PAC) *q(i)* in Abhängigkeit des "lags" i (Zeitverzögerung) dar. Die *r(i)* und *q(i)* sind anhand der folgenden Regressionen berechnet ($\hat{u}_t$: Residuen):

$$\hat{u}_t = r(1)\hat{u}_{t-1} + v_t \qquad \hat{u}_t = q(1)\hat{u}_{t-1} + v_t$$

$$\hat{u}_t = r(2)\hat{u}_{t-2} + v_t \qquad \hat{u}_t = c(1)\hat{u}_{t-1} + q(2)\hat{u}_{t-2} + v_t$$

$$\hat{u}_t = r(3)\hat{u}_{t-3} + v_t \qquad \hat{u}_t = c(1)\hat{u}_{t-1} + c(2)\hat{u}_{t-2} + q(3)\hat{u}_{t-3} + v_t$$

Folgt der Störterm einem autoregressiven Prozess erster Ordnung

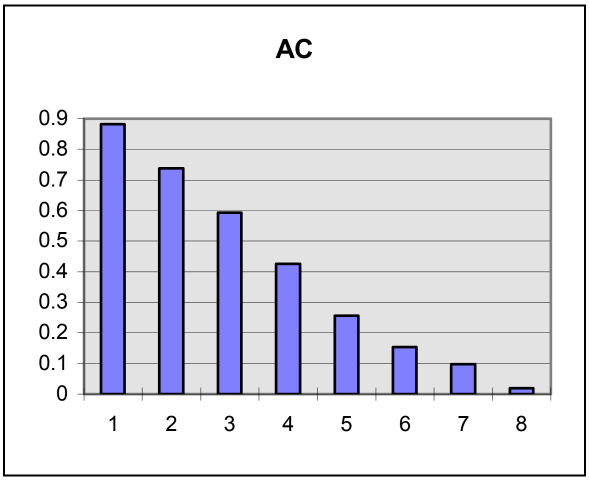
$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \text{AR(1)-Prozess}$$

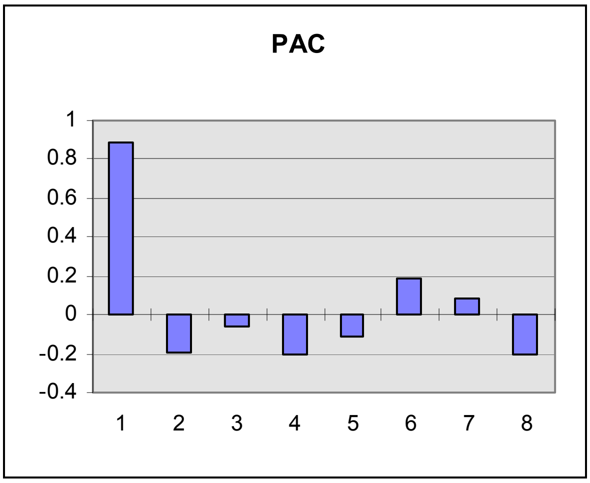
so ist eine Korrelogram zu erwarten, bei dem die Autokorrelationsfunktion *r(i)* mit zunehmendem *i* kontinuierlich gegen Null tendiert und die partielle Autokorrelationsfunktion *q(i)* nach *i* = 1 sofort abbricht.

So spricht z.B. das Korrelogram für die Residuen der Gleichung

LOG(KNARK) = c(1) + c(2) LOG(EINRK)

ganz klar für einen AR(1)-Prozess:





"Chow Breakpoint Test" auf Parameterstabilität

Wir betrachten als Beispiel das Anpassungsmodell

LOG(KNARK) =
C(4)*C(1)+C(4)*C(2)*LOG(EINRK)+C(4)*C(3)*LOG(PREL)
+(1-C(4))*LOG(KNARK(-1))

Wir schätzen dieses Modell für die Subperioden 1961-1984 und 1985-1995 sowie über den gesamten Zeitraum 1961-1995 (altes Datenset):

Schätz-periode	Koeffizienten und *Standardfehler* C(1)	C(2)	C(3)	C(4)	R-Quadrat	RSS
(a) 1961-84	2.902	0.534	-0.925	0.667	0.994	**0.001934**
	0.438	*0.045*	*0.111*	*0.080*		
(b) 1985-95	6.732	0.143	0.414	0.532	0.873	**0.000133**
	1.171	*0.118*	*0.118*	*0.332*		
(c) 1961-95	5.792	0.239	-0.783	0.186	0.982	**0.008331**
	1.931	*0.196*	*0.440*	*0.440*		

Die geschätzten Koeffizienten differieren beträchtlich zwischen den beiden Subperioden (a) und (b). Die Einkommenselastizität C(2) z.B. fällt von $C(2)^a = 0.53$ auf $C(2)^b = 0.14$. Schätzt man die Gleichung über den gesamten Zeitraum (c), so zwingt man die Einkommenselastizität auf einen einheitlichen Wert. Dies entspricht der Restriktion $C(2)^a = C(2)^b$. Insgesamt gibt es 4 solche Restriktionen. Man hat somit ein *unrestringiertes* Modell, in welchem Parameteränderungen zwischen den beiden Subperioden zugelassen sind, und ein *restringiertes* Modell, in welchem *stabile Parameter* unterstellt sind.

$URSS = RSS^a + RSS^b = 0.002067$ unrestringerte Schätzung
$RRSS = RSS^c = 0.008331$ restringierte Schätzung

$$F = \frac{(RRSS - URSS)/4}{URSS/(35-8)} = 20.45$$

Bei einem kritischen F(4,27)-Wert von 2.7 (5%-Niveau) ist die Null-Hypothese stabiler Parameter klar abzulehnen.

In EViews geht man nach der Schätzung einer Gleichung über View / Stability Test zum Chow Breakpoint Test und gibt dann den gewünschten Breakpoint ein.

Breakpoint 1975 z.B. bedeutet: Teilstichproben 1961-1974 und 1975-1995. Man kann den Breakpoint auch verschieben und nach dem höchsten F-Wert bzw. dem tiefsten Signifikanzniveau suchen:

*Breakpoint**	*F-Wert*	*Signifikanzniveau*
1979	11.07198	0.000019
1980	12.03710	0.000010
1981	13.91687	0.000003
1982	15.47752	0.000001
1983	26.08703	0.000000
1984	23.50646	0.000000
1985	20.45335	0.000000
1986	18.94962	0.000000
1987	16.43390	0.000001

Gemäss dieser Analyse kommt die Parameterinstabilität bzw. die ihr zugrundeliegende Verhaltensänderung am klarsten zum Ausdruck, wenn man den Breakpoint auf 1983 legt.

Parameterinstabilität schlägt sich häufig in autokorrelierten Residuen nieder, weil in der Regressionsgleichung fälschlicherweise zeitstabile Parameter unterstellt werden.

Die Übung "Autokorrelation" dient der Vertiefung dieser Punkte.