

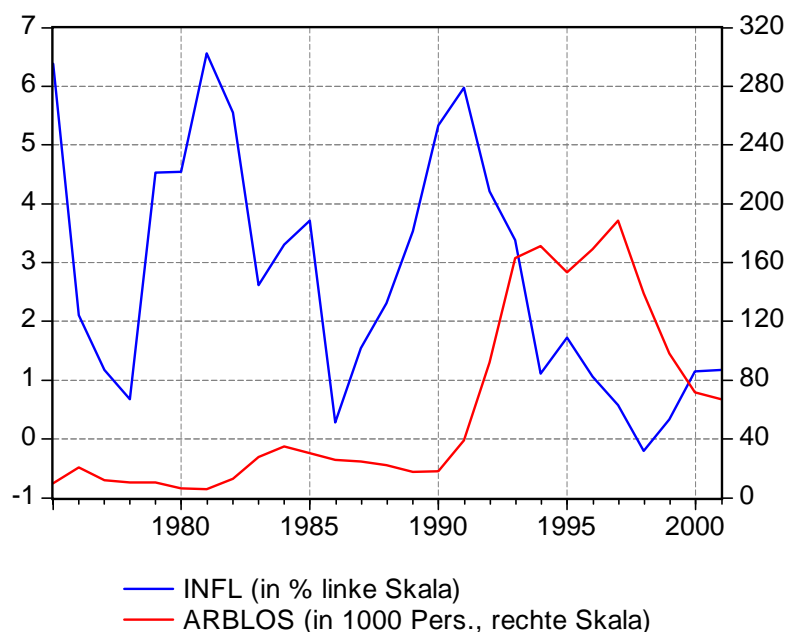
Beispiel für Varianzanalyse in multipler Regression mit zwei erklärenden Variablen

Hypothese: Die Inflation hängt positiv von der Inflation im Vorjahr und negativ von der Arbeitslosigkeit ab. Bei diesem Zusammenhang handelt es sich um die sogenannte *Phillips-Kurve*.

Daten (aus der Tabelle Konsum.xls):

Inflationsrate INFL berechnet als $INFL = 100 * (PKONS/PKONS(-1) - 1)$

Arbeitslosenzahl ARBLOS (in 1000 Personen)



Schätzperiode 1975 - 2001 (weil Arbeitslosenzahl vor 1975 praktisch = 0)

Schätzgleichung:

$$INFL = C(1) + C(2) * ARBLOS(-1) + C(3) * INFL(-1) + RES$$

Fragestellung Varianzanalyse:

1) Aufspaltung der Varianz in INFL (TSS) in einen erklärten Teil (ESS) und in einen unerklärten Rest (Summe der quadrierten Residuen RSS):

$$TSS = ESS + RSS$$

2) Innerhalb ESS: Erklärungsbeiträge von ARBLOS(-1) und INFL(-1)

Vorgehen in Eviews:

"Pseudo-Regression" zur Berechnung von TSS:

INFL= C(1)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.765290	0.394485	7.009876	0.0000
R-squared	0.000000	Mean dependent var	2.765290	
Adjusted R-squared	0.000000	S.D. dependent var	2.049803	
S.E. of regression	2.049803	Akaike info criterion	4.309698	
Sum squared resid	109.2440	Schwarz criterion	4.357692	
Log likelihood	-57.18093	Durbin-Watson stat	0.724624	

⇒ TSS = 109.244

(a) Erklärung der Inflation mit der Arbeitslosenzahl des Vorjahres

INFL= C(1)+C(2)*ARBLOS(-1)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	3.952468	0.434660	9.093237	0.0000
C(2)	-0.020328	0.005125	-3.966238	0.0005
R-squared	0.386218	Mean dependent var	2.765290	
Adjusted R-squared	0.361666	S.D. dependent var	2.049803	
S.E. of regression	1.637707	Akaike info criterion	3.895658	
Sum squared resid	67.05207	Schwarz criterion	3.991646	
Log likelihood	-50.59138	Durbin-Watson stat	1.041395	

⇒ $ESS_{\text{ARBLOS}} = 109.244 - 67.052 = 42.192$

(b) Erklärung der Inflation mit der Inflation des Vorjahres

INFL= C(1)+C(3)*INFL(-1)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.017817	0.484882	2.099101	0.0461
C(3)	0.565680	0.124020	4.561212	0.0001
R-squared	0.454204	Mean dependent var	2.765290	
Adjusted R-squared	0.432372	S.D. dependent var	2.049803	
S.E. of regression	1.544344	Akaike info criterion	3.778263	
Sum squared resid	59.62496	Schwarz criterion	3.874251	
Log likelihood	-49.00655	Durbin-Watson stat	1.486056	

⇒ $ESS_{\text{INFL}} = 109.244 - 59.625 = 49.619$

(c) Erklärung der Inflation mit der Arbeitslosenzahl und der Inflation des Vorjahres

$$\text{INFL} = C(1) + C(2) \cdot \text{ARBLOS}(-1) + C(3) \cdot \text{INFL}(-1)$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.234119	0.668560	3.341686	0.0027
C(2)	-0.012404	0.005103	-2.430470	0.0229
C(3)	0.406442	0.130957	3.103632	0.0048
R-squared	0.562008		Mean dependent var	2.765290
Adjusted R-squared	0.525509		S.D. dependent var	2.049803
S.E. of regression	1.411973		Akaike info criterion	3.632292
Sum squared resid	47.84800		Schwarz criterion	3.776274
Log likelihood	-46.03594		Durbin-Watson stat	1.586869

$$\Rightarrow \text{ESS} = 109.244 - 47.848 = 61.396$$

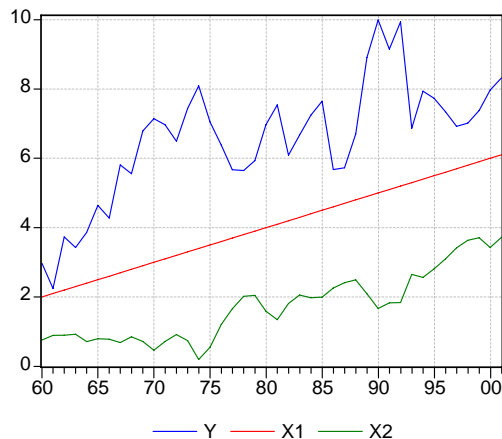
Varianzanalyse: Multiples, einfaches und partielles R^2

INFL	(a) ARBLOS(-1)		(b) INFL(-1)		(c) ARBLOS(-1)+ INFL(-1)		partieller Erklärungsbeitrag v. INFL(-1) = ESS(c)-ESS(a) = RSS(a)-RSS(c)	partieller Erklärungsbeitrag v. ARBLOS(-1) = ESS(c)-ESS(b) = RSS(b)-RSS(c)
TSS	ESS	RSS	ESS	RSS	ESS	RSS		
109.24	42.19	67.05	49.62	59.62	61.40	47.85	19.20	11.78
	einfaches R^2 = ESS/TSS = 1-RSS/TSS		einfaches R^2 = ESS/TSS = 1-RSS/TSS		multiples R^2 = ESS/TSS = 1-RSS/TSS		partielles R^2 = 1 - RSS(c)/RSS(a)	partielles R^2 = 1 - RSS(c)/RSS(b)
	0.386		0.454		0.562		0.286	0.198
	(1 - multiples R^2)		= (1 - einfaches R^2)		X (1 - partielles R^2)			
	1- 0.562 0.438		1- 0.386 0.614		1- 0.286 :0.714		zuerst ARBLOS, dann INFL(-1)	
	1- 0.562 0.438		1- 0.454 0.546		1- 0.198 :0.802		zuerst INFL(-1), dann ARBLOS	

Ergebnis:

Die Inflation in Periode t ist etwas stärker von der Inflation in t-1 als von der Arbeitslosigkeit in t-1 geprägt. Das einfache R^2 ist grösser (0.454 vs. 0.386), und auch das partielle R^2 ist grösser (0.286 vs. 0.198).

Bemerkung zur einfacher und partieller Korrelation



Einfache und partielle Korrelation können unterschiedliche Vorzeichen haben.

Beispiel mit hypothetischen Daten

$R_{y x1}$ ist positiv
 $R_{y x2}$ ist positiv
 $R_{y x2.x1}$ ist negativ

Y hat ein TSS von **123.0146**

Sample: 1960 2001

$Y = C(1) + C(2) * X1$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.422624	0.650198	3.725977	0.0006
C(2)	1.023418	0.153802	6.654106	0.0000

R-squared	0.525375	Adjusted R-squared	0.513510
S.E. of regression	1.208157	Sum squared resid	58.38576
Durbin-Watson stat	0.667430	ESS =	64.62884

Sample: 1960 2001

$Y = C(1) + C(3) * X2$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	5.375699	0.503227	10.68246	0.0000
C(3)	0.686070	0.251899	2.723593	0.0095

R-squared	0.156438	Adjusted R-squared	0.135349
S.E. of regression	1.610671	Sum squared resid	103.7705
Durbin-Watson stat	0.508466	ESS =	19.2441

Sample: 1960 2001

$Y = C(1) + C(2) * X1 + C(3) * X2$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.711284	0.391694	-1.815919	0.0771
C(2)	2.874339	0.165972	17.31826	0.0000
C(3)	-2.511278	0.203898	-12.31634	0.0000

R-squared	0.902931	Adjusted R-squared	0.897953
S.E. of regression	0.553334	Sum squared resid	11.94094
Durbin-Watson stat	1.411960	ESS =	111.07366

Kommentar

Regressiert man Y auf X_2 , so ergibt sich – weil beide Variablen trendmässig ansteigen – ein positiver Einfluss $C(3)$. Der Erklärungskraft von X_2 ist jedoch klein: Das TSS von Y beträgt 123.0146, davon wird von X_2 nur ein ESS von 19.2441 erklärt.

Fügt man hingegen X_2 in eine Regression ein, in der X_1 bereits enthalten ist, so wird der trendmässige Anstieg von Y durch X_1 aufgefangen und X_2 übernimmt die Aufgabe, die kürzerfristigen Schwankungen von Y zu erklären. Da diese gegenläufig zu X_2 sind, wird der Einfluss von X_2 auf Y negativ. Der partielle Erklärungsbeitrag von X_2 ist zudem viel grösser als in der einfachen Regression von Y auf X_2 , denn ESS steigt durch den Einbezug von X_2 von 64.62884 auf 111.07366, also um 46.44482.

Mit anderen Worten eignet sich X_2 allein - mit positivem Vorzeichen - nicht gut zur Erklärung von Y , zusammen mit X_1 aber - mit negativem Vorzeichen - viel besser.