ulm university universität ulm

Universität Ulm | 89069 Ulm | Germany

Fakultät für Mathematik und Wirtschaftswissenschaften

Prof. Dr. Werner Smolny
Institut für Wirtschaftspolitik

Ludwig-Erhard-Stiftungsprofessur
Institutsdirektor

Sommersemester 2016

Das lineare Regressionsmodell

Das ökonomische Modell

$$y = \beta_0 + \beta_1 \cdot x$$

y:	*die endogene Variable*
x:	*die erklärende Variable*
β_0, β_1:	*die Parameter (Koeffizienten) des ökonomischen Modells*

Eine Verhaltensgleichung

- *y ist linear abhängig von x*
- *β_0 ist der y-Achsenabschnitt*
  (also wie hoch ist y, wenn x den Wert 0 annimmt)
- *β_1 ist der marginale Effekt von x auf y $(\beta_1 = \partial y / \partial x)$*

z.B. eine Keynessche Konsumfunktion

$$C = \overline{c} + c' \cdot Y^v$$

Der aggregierte Konsum C
ist abhängig vom aggregierten verfügbaren Einkommen Y^v

$\overline{c}$ ist der autonome Konsum, c' ist die marginale Konsumneigung

Helmholtzstr. 20, Raum E 05
Tel. 0731 50-24260, Fax -24262

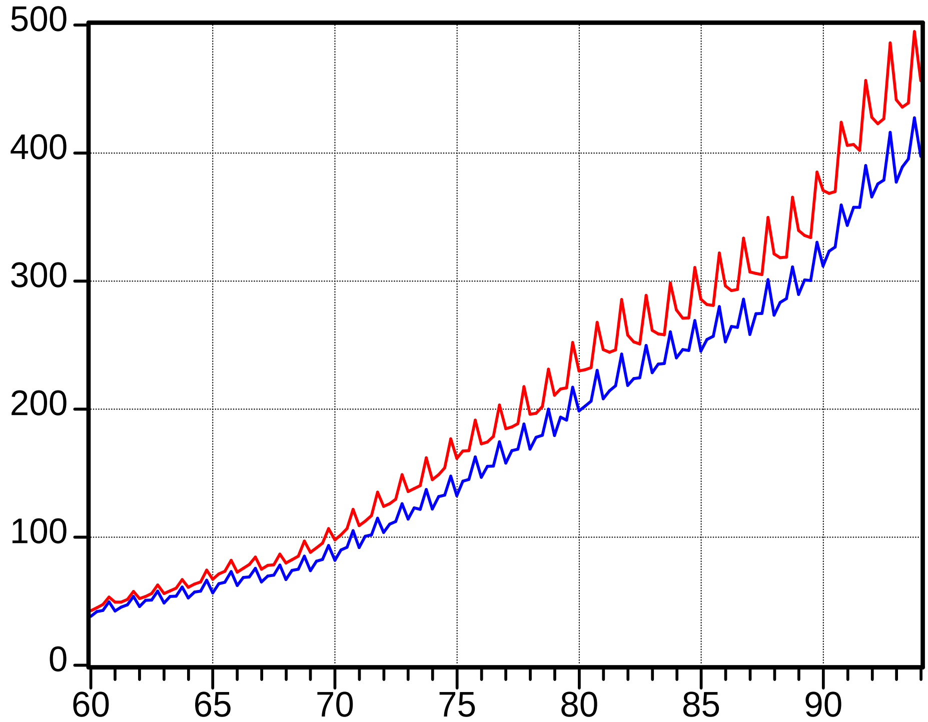
http://www.uni-ulm.de/mawi/wirtschaftspolitik
Werner.Smolny@uni-ulm.de

Die Daten $\rightarrow$ Eviews Workfile ew21

Schätzung anhand der Daten der vierteljährlichen volkswirtschaftlichen Gesamtrechnung

Westdeutschland, 1960.1 bis 1994.4

PLOT (Zeitreihenschaubild)



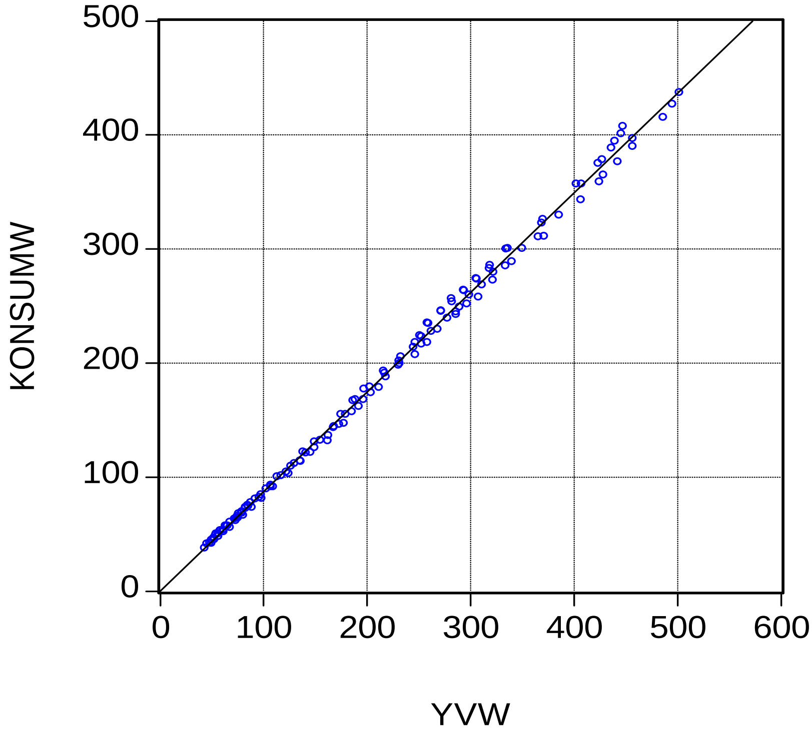
Der Konsum und das verfügbare Einkommen stiegen über die Zeit an

Ursache: Einerseits Preisanstieg, andererseits Anstieg der Realeinkommen

Konsum und verfügbares Einkommen wiesen auch
starke saisonale Schwankungen auf

Ursache: Zahlung eines 13. Monatsgehalts im 4. Quartal (Weihnachtsgeld), Ausgaben für Weihnachten

SCAT ($x - y$ Diagramm)



Konsum und verfügbares Einkommen sind positiv korreliert

Das empirische Modell

1. *Die Schätzung der Parameter des Modells kann anhand aggregierter Daten für den Konsum und das verfügbare Einkommen erfolgen*
2. *Natürlich ist der Konsum nicht nur vom verfügbaren Einkommen abhängig, sondern auch von anderen Faktoren*
3. *Die Abweichungen vom linearen Zusammenhang werden als Residuen bezeichnet*

Das empirische Modell

$$KONSUMw_t = \beta_0 + \beta_1 \cdot YVw_t + \varepsilon_t$$

t: *Zeitindex,* $t = \{1960.1, \ldots, 1994.4\}$
ε: *Residuen bzw. Störterme bzw. Abweichungen bzw. Fehler*

Schätzung der Koeffizienten β_0, β_1 *anhand der Daten der vierteljährlichen volkswirtschaftlichen Gesamtrechnung für die Zeit von 1960 bis 1994*

Das ökonometrische Modell

Die Schätzung mit der Methode der Kleinsten Quadrate

$$KONSUMw_t = \hat{\beta}_0 + \hat{\beta}_1 \cdot YVw_t + \hat{\varepsilon}_t$$

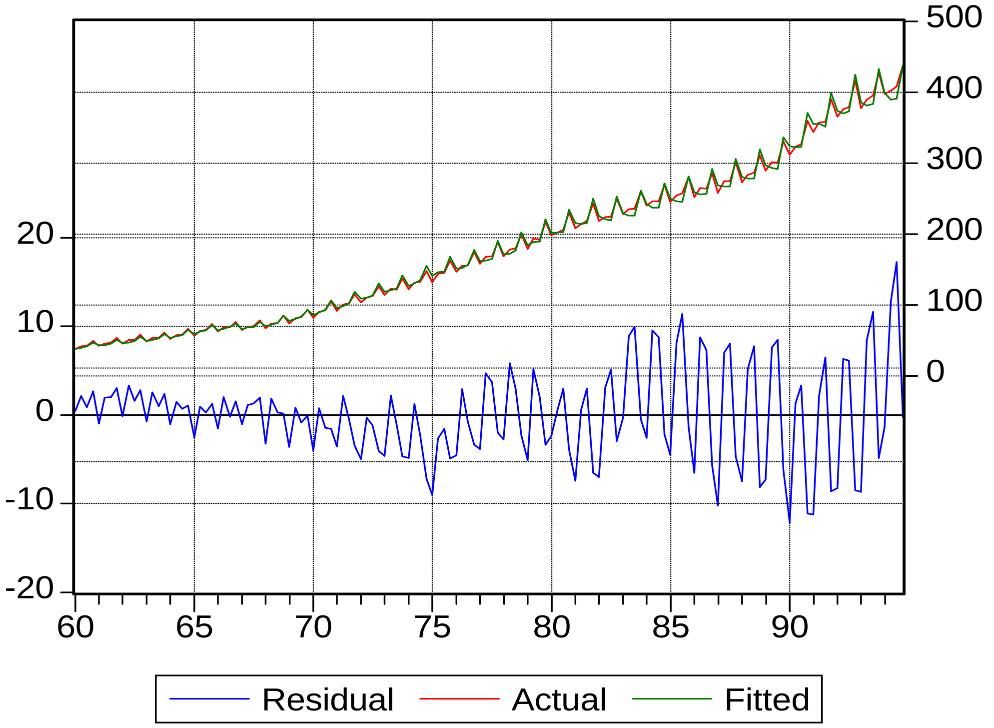
bzw.

$$KON\hat{S}UMw_t = \hat{\beta}_0 + \hat{\beta}_1 \cdot YVw_t$$

$\hat{\beta}_0$: *geschätzter autonomer Konsum*
$\hat{\beta}_1$: *geschätzte marginale Konsumneigung*
$\hat{\varepsilon}_t$: *geschätzte Residuen*
$KON\hat{S}UMw_t$: *geschätzte Konsumausgaben*

```
ls konsumw c yvw

============================================================
Dependent Variable: KONSUMW          Method: Least Squares
Sample: 1960:1 1994:4
Included observations: 140
============================================================
     Variable      CoefficientStd. Errort-Statistic  Prob.
============================================================
         C           0.407818    0.866981    0.470389    0.6388
        YVW          0.872722    0.003517    248.1528    0.0000
============================================================
R-squared            0.997764    Mean dependent var 184.6276
Adjusted R-squared   0.997748    S.D. dependent var 111.6542
S.E. of regression   5.298797    Akaike info criteri6.187019
Sum squared resid    3874.661    Schwarz criterion  6.229043
Log likelihood      -431.0914    F-statistic        61579.81
Durbin-Watson stat   1.654655    Prob(F-statistic)  0.000000
============================================================
```



$\boxed{\text{Die Schätzung der Parameter}}$

Wähle ("schätze" bzw. berechne) die Koeffizienten so,
dass die Summe der quadrierten Abweichungen der Schätzgleichung
für die Beobachtungsperiode minimiert wird

$$\min_{\rightarrow \hat{\beta}_0, \hat{\beta}_1} \quad \sum_{t=1960.1}^{1994.4} \quad \hat{\varepsilon}_t^2$$

$\boxed{\text{Die Berechnung der Parameter}}$

$$\sum^t \hat{\varepsilon}_t^2 = \sum^t (KONSUMw_t - KONS\hat{U}Mw_t)^2$$

$$= \sum^t (KONSUMw_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot YVw_t)^2$$

Die Summe der quadrierten Abweichungen wird minimiert, wenn die partiellen Ableitungen in Bezug auf die Parameter gleich 0 gesetzt werden

$$\frac{\partial \sum^t \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} = \sum^t 2 \cdot (KONSUMw_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot YVw_t) \cdot (-1) = 0$$

$$\frac{\partial \sum^t \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} = \sum^t 2 \cdot (KONSUMw_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot YVw_t) \cdot (-YVw_t) = 0$$

Die Normalgleichungen

$$\sum^{t} KONSUMw_t - \hat{\beta}_0 \cdot T - \hat{\beta}_1 \cdot \sum^{t} YVw_t = 0$$

Der Mittelwert der geschätzten Residuen ist gleich 0

$$\sum^{t} KONSUMw_t \cdot YVw_t - \hat{\beta}_0 \cdot \sum^{t} YVw_t - \hat{\beta}_1 \sum^{t} YVw_t^2 = 0$$

Die Residuen sind unkorreliert mit der (den) erklärenden Variablen

Die Beziehungen zwischen den Mittelwerten

$$\overline{KONSUMw} - \hat{\beta}_0 - \hat{\beta}_1 \cdot \overline{YVw} = 0$$

$$\overline{KONSUMw \cdot YVw} - \hat{\beta}_0 \cdot \overline{YVw} - \hat{\beta}_1 \overline{YVw^2} = 0$$

Lösen der beiden Gleichungen für die 2 Unbekannten ($\hat{\beta}_0, \hat{\beta}_1$)
liefert die geschätzten Parameter

Die marginale Konsumneigung beträgt etwa 0.87,
der autonome Konsum ist gering (nicht signifikant von 0 verschieden)

Unterschiedliche empirische Spezifikationen des Zusammenhangs von Konsum und Einkommen

a) Die Schätzung in Logarithmen

```
============================================================
Dependent Variable: LOG(KONSUMW)       Method: Least Squares
Sample: 1960:1 1994:4
Included observations: 140
============================================================
     Variable      CoefficientStd. Errort-Statistic  Prob.
============================================================
         C          -0.078648   0.017249  -4.559593   0.0000
     LOG(YVW)        0.989736   0.003328   297.3719   0.0000
============================================================
R-squared             0.998442    Mean dependent var 5.004162
Adjusted R-squared    0.998431    S.D. dependent var 0.692430
S.E. of regression    0.027431    Akaike info criter-4.340089
Sum squared resid     0.103841    Schwarz criterion -4.298066
Log likelihood        305.8062    F-statistic         88430.05
Durbin-Watson stat    1.689312    Prob(F-statistic)   0.000000
============================================================
```

Die Schätzung der logarithmierten Werte liefert nicht marginale Effekte von x auf y, sondern die marginalen Effekte von ln x auf ln y, also Elastizitäten

$$\beta_1 = \frac{\partial \ln y}{\partial \ln x} = \frac{\partial y}{\partial x} \cdot \frac{x}{y}$$

Interpretation:

Der marginale Effekt einer relativen Einkommensänderung
auf die relative Änderung des Konsums
$\rightarrow$ *die Elastizität des Konsums in Bezug auf das verfügbare Einkommen*

Achtung: Das ist ein anderes Modell !

$$\ln KONSUMw_t = \beta_0 + \beta_1 \cdot \ln YVw_t + \varepsilon_t$$

bzw.

$$KONSUMw_t = \exp(\beta_0) \cdot YVw_t^{\beta_1} \cdot \exp(\varepsilon_t)$$

- *Das verfügbare Einkommen wirkt proportional auf den Konsum, der Störterm wirkt multiplikativ auf die Gleichung (den Konsum)*
- *Die geschätzte Elastizität des Konsums in Bezug auf das verfügbare Einkommen beträgt etwa 1, d.h. ein um ein Prozent höheres Einkommen bedeutet einen um etwa 1 Prozent höheren Konsum*
- *Die Varianz der Residuen erscheint weitgehend konstant über die Beobachtungsperiode, die Residuen weisen immer noch positive Autokorrelation 4. Ordnung auf (Saison)*

b) Die Schätzung mit Saisondummies

Saisondummies weisen für jeweils ein Quartal (jedes Jahr) den Wert 1 auf, sonst 0

```
show @SEAS(1) @SEAS(2) @SEAS(3) @SEAS(4)
```

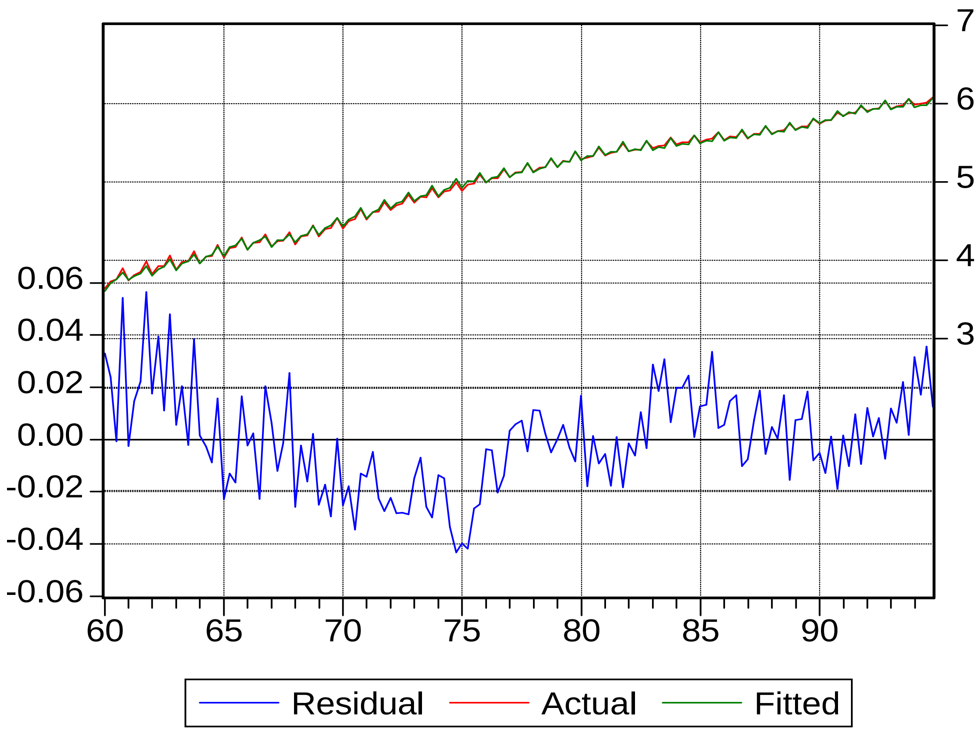
obs	@SEAS(1)	@SEAS(2)	@SEAS(3)	@SEAS(4)
1960:1	1.000000	0.000000	0.000000	0.000000
1960:2	0.000000	1.000000	0.000000	0.000000
1960:3	0.000000	0.000000	1.000000	0.000000
1960:4	0.000000	0.000000	0.000000	1.000000
1961:1	1.000000	0.000000	0.000000	0.000000
.	.	.	.	.

Die Schätzung mit Saisondummies impliziert eine einfache (log)-lineare Form der Saisonbereinigung (konstante Saisonfigur, konstante Saisonfaktoren)

Dependent Variable: LOG(KONSUMW) Method: Least Squares
Sample: 1960:1 1994:4
Included observations: 140

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.085586	0.012934	-6.617335	0.0000
LOG(YVW)	0.989902	0.002384	415.1491	0.0000
@SEAS(1)	-0.022565	0.004693	-4.807790	0.0000
@SEAS(2)	0.027245	0.004691	5.807518	0.0000
@SEAS(3)	0.019648	0.004688	4.190742	0.0000

R-squared	0.999224	Mean dependent var	5.004162
Adjusted R-squared	0.999201	S.D. dependent var	0.692430
S.E. of regression	0.019571	Akaike info criter	-4.994474
Sum squared resid	0.051708	Schwarz criterion	-4.889415
Log likelihood	354.6132	F-statistic	43465.34
Durbin-Watson stat	1.060116	Prob(F-statistic)	0.000000



c) Schätzung in Differenzen der Logarithmen (Wachstumsraten)

Die Schätzung in Differenzen der Logarithmen, also Wachstumsraten, liefert den kurzfristigen Effekt von Einkommensänderungen auf Konsumänderungen, bzw. die kurzfristige Elastizität von Konsumänderungen in Bezug auf Einkommensänderungen

Unterscheidung von 1. Differenzen (Differenzen zum Vorquartal) und 4. Differenzen (Veränderungen zum Vorjahresquartal)

Ergebnis:

Die Schätzung in 4. Differenzen impliziert eine weitere Art der Saisonbereinigung, die Saisondummies haben keinen signifikanten Einfluss in der geschätzten Gleichung

Die Bildung von 4. Differenzen eliminiert die Saisonfigur für die Daten des Konsums und des verfügbaren Einkommens

```
============================================================
Dependent Variable: D(LOG(KONSUMW),0,4) Method: Least Squares
Sample(adjusted): 1961:1 1994:4
Included observations: 136 after adjusting endpoints
============================================================
     Variable      CoefficientStd. Errort-Statistic  Prob.
============================================================
         C           0.014034   0.002641   5.312863   0.0000
  D(LOG(YVW),0,4)    0.779952   0.035928   21.70874   0.0000
============================================================
R-squared            0.778611    Mean dependent var 0.066539
Adjusted R-squared   0.776959    S.D. dependent var 0.026221
S.E. of regression   0.012383    Akaike info criter-5.930314
Sum squared resid    0.020549    Schwarz criterion -5.887480
Log likelihood       405.2613    F-statistic        471.2696
Durbin-Watson stat   1.125757    Prob(F-statistic)  0.000000
============================================================
```

Die Interpretation der geschätzten Koeffizienten als Mittelwerte

Die Schätzung der Konsumquote auf eine Konstante

```
============================================================
Dependent Variable: KONSUMW/YVW        Method: Least Squares
Sample: 1960:1 1994:4
Included observations: 140
============================================================
     Variable      CoefficientStd. Errort-Statistic  Prob.
============================================================
         C           0.877250   0.002099   417.8919   0.0000
============================================================
R-squared            0.000000    Mean dependent var 0.877250
Adjusted R-squared   0.000000    S.D. dependent var 0.024838
S.E. of regression   0.024838    Akaike info criter-4.545736
Sum squared resid    0.085755    Schwarz criterion -4.524724
Log likelihood       319.2015    Durbin-Watson stat 1.581111
============================================================
```

Die Schätzung der Konsumquote auf die Konstante liefert als Koeffizienten den Mittelwert der Variable über den Beobachtungszeitraum, die Residuen sind die Abweichungen der Konsumquote (jede Beobachtung) vom Mittelwert für die Konsumquote

Der Standardfehler der Gleichung ist die durchschnittliche (Wurzel aus der mittleren quadratischen) Abweichung der einzelnen Beobachtungen vom Mittelwert

$\updownarrow$

Der Standardfehler des geschätzten Koeffizienten kann interpretiert werden als durchschnittliche Abweichung des Mittelwerts aus 140 Beobachtungen (Stichprobe: 1960.1 bis 1994.4)

Die Schätzung der Konsumquote auf 4 Saisondummies liefert als Koeffizienten den Mittelwert der Konsumquote für jedes Quartal

Die geschätzten Werte dieser Gleichung können interpretiert werden als bedingter Mittelwert (Erwartungswert) der Konsumquote, gegeben die Jahreszeit (das Quartal)

Ebenso können die geschätzten Werte der Konsumfunktion (siehe oben) interpretiert werden als bedingter Erwartungswert für den Konsum, gegeben das Einkommen

Der geschätzte Koeffizient β_1 kann interpretiert werden als mittlerer Effekt des Einkommens auf den Konsum, der Standardfehler der Gleichung entspricht der durchschnittlichen Abweichung des Konsums (jede Beobachtung) von den geschätzten Werten aus der Konsumfunktion

$$\updownarrow$$

Die Standardabweichung des Koeffizienten ist die durchschnittliche Abweichung des aus 140 Beobachtungen geschätzten (berechneten) Koeffizienten (bzw. Mittelwerts, siehe oben)

Literatur: Winker, Kapitel 3, 6 und 7 (auszugsweise).

Einige der von EViews ausgewiesenen Test-Statistiken für das Kleinste Quadrate Modell

SMPL	Sample	Schätzzeitraum
COEFFICIENT	$\hat{\beta}_i$	die geschätzten Koeffizienten, $\hat{\beta}_i = \frac{\partial y}{\partial x_i}$ mit y: endogene Variable, x_i: erklärende Variable(n)
STD.ERROR	$s(\hat{\beta}_i)$	Geschätzte Standardabweichung der Koeffizienten
T-STAT.	$t-$Wert	$= \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$, die $t-$Statistik des geschätzten Koeffizienten. Für $t-$Wert $\approx$ 2 schließt das 95% Konfidenzintervall die Null gerade (nicht) ein
2-TAIL SIG		Konfidenzniveau des Koeffizienten, mit $(1-p)$-prozentiger Wahrscheinlichkeit liegt der wahre Koeffizient innerhalb $0 < \hat{\beta}_i < 2 \cdot \hat{\beta}_i$
R-squared	R^2	das Bestimmtheitsmaß der Schätzung, Anteil der erklärten Varianz an der Gesamtvarianz
Adjusted R-squared	$\overline{R}^2$	bereinigtes Bestimmtheitsmaß (um die Zahl der Freiheitsgrade)
S.E. of regression	s	Geschätzte Standardabweichung der Residuen, durchschnittlicher Fehler der Schätzung
Durbin-Watson stat.	DW	Teststatistik für Autokorrelation erster Ordnung, DW=2: keine Autokorrelation, DW<2: positive Autokorrelation, DW>2: negative Autokorrelation
F-statistic	$F_{k-1,N-k}$	Test auf Signifikanz des R^2, also Test ob das R^2 von 0 verschieden ist
Prob(F-statistic)		das Signifikanzniveau für den Test $R^2 = 0$

vertiefende Literatur: Pindyck, Rubinfeld,
Econometric models and Economic Forecasts, chapter 3 + 4.