

Beurteilung alternativer Modelle mit F-Test und J-Test und Analyse des "equation output" bei der Modellwahl

1. Nested models und non-nested models

Geht ein Modell aus einem anderen durch Null-Restriktionen hervor spricht man von nested models. So ist etwa das Modell A $y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_{At}$ "nested" im Modell B, für das gilt $y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + u_{Bt}$; denn A geht aus B hervor mit den Null-Restriktionen $H_0: \beta_4 = \beta_5 = 0$, die mit dem F-Test zu testen ist. Wird H_0 angenommen, so ist dem kleineren Modell A der Vorzug vor dem größeren Modell B (in das es "eingenistet" ist) zu geben. Die Regressoren von A sind eine echte Teilmenge der Regressoren von B. Beim Vergleich zweier Modelle, bei denen eines nicht im anderen "genestet" ist, ist dagegen der J-Test zu verwenden (vgl. unten Abschn. 4)

2. Der F- Test über Irrelevanz zusätzlicher Regressoren

Zur Demonstration des F-Tests bei L Restriktionen mit der Prüfgröße

$$F = \frac{(S_{\hat{u}\hat{u}}^0 - S_{\hat{u}\hat{u}}) / L}{S_{\hat{u}\hat{u}} / (T - K - 1)}$$

benutzen wir Daten aus dem Download A2 "Mehr Ärzte = höhere Lebenserwartung" mit den Variablen (Zeitreihen der Bundesrepublik Deutschland von 1992 bis 2006)¹:

leben = Lebenserwartung in Jahren, **arzt** = Anzahl der Ärzte je 1000 Einwohner, **aus** = Gesundheitsausgaben in Prozent des BIP, **pkaus** = pro Kopf Gesundheitsausgaben, **hsall** = Anzahl der Hochschulen in Deutschland, **hspriv** = darunter: der privaten Hochschulen.²

Der folgende E-Views equation output zeigt zunächst das Ergebnis wenn man "**leben**" auf alle fünf Regressoren "regressiert" (außer **arzt** ist kein Regressor für sich genommen signifikant!):

Dependent Variable: LEBEN				
Method: Least Squares				
Date: 01/06/10 Time: 12:01				
Sample: 1992 2006				
Included observations: 15				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	69.56570	5.672194	12.26434	0.0000
ARZT	4.944828	1.550694	3.188784	0.0110
AUS	-0.911442	0.709270	-1.285042	0.2309
PKAUS	0.000547	0.001974	0.277313	0.7878
HSPRIV	-0.007013	0.029949	-0.234184	0.8201
HSALL	0.010446	0.017919	0.582938	0.5742
R-squared	0.981897	Mean dependent var	80.81333	
Adjusted R-squared	0.971839	S.D. dependent var	0.966486	
S.E. of regression	0.162188	Akaike info criterion	-0.510949	
Sum squared resid	0.236744	Schwarz criterion	-0.227729	
Log likelihood	9.832121	F-statistic	97.62896	
Durbin-Watson stat	1.229789	Prob(F-statistic)	0.000000	

¹ Auf die Zeitvariable **year** (1992, ..., 2006, also T = 15 Jahre) kann hier verzichtet werden.

² Die Daten sind am Ende des downloads A2 vollständig aufgelistet.

Die Frage ist nun ob die letzten beiden Regressoren – die ja sachlich mit der Lebenserwartung wenig zu tun haben - einen signifikanten Erklärungsbeitrag liefern. Wir rechnen deshalb die Regression noch einmal jedoch ohne **hsall** (= x_4) und **hspriv** (= x_5), d.h. wir betrachten ein restringiertes Modell (man beachte, dass auch **pkaus** nicht signifikant ist):

Dependent Variable: LEBEN					
Method: Least Squares					
Date: 01/06/10 Time: 12:22					
Sample: 1992 2006					
Included observations: 15					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
ARZT	5.307830	1.158836	4.580311	0.0008	
AUS	-0.958018	0.306725	-3.123373	0.0097	
PKAUS	0.000624	0.000359	1.735713	0.1105	
C	72.02130	2.227328	32.33530	0.0000	
R-squared	0.980812	Mean dependent var	80.81333		
Adjusted R-squared	0.975579	S.D. dependent var	0.966486		
S.E. of regression	0.151036	Akaike info criterion	-0.719420		
Sum squared resid	0.250930	Schwarz criterion	-0.530606		
Log likelihood	9.395649	F-statistic	187.4231		
Durbin-Watson stat	1.115907	Prob(F-statistic)	0.000000		

Es ist nun leicht, die Prüfgröße F für die Hypothese $H_0: \beta_4 = \beta_5 = 0$ ($L = 2$) zu errechnen. Man erhält für die Prüfgröße den Wert

$$F = \frac{(0.250930 - 0.236744)/2}{0.236744/9} = 0.269645693$$

der mit der F-Verteilungstabelle bei 2 und 9 Freiheitsgraden zu vergleichen ist,

2. Omitted variables Test

In der Praxis ist es natürlich nicht nötig, sich umständlich zwei Regressionen anzusehen und die Prüfgröße F aus den beiden Größen "Sum squared resid" zu berechnen. Man erhält die gleiche Prüfgröße F in EViews ausgehend von der Seite der Schätzergebnisse des underfitted Modells (oben auf dieser Seite) mit **arzt**, **aus** und **pkaus** über

view → coefficient tests

angeboten wird dabei: 1. Confidence Ellipse [für je zwei Regressionskoeffizienten], 2. Wald Coefficient Restrictions, 3. *Omitted Variables Likelihood Ratio Test* (LRT), 4. Redundant Variables LRT).

Wählt man 3 und gibt als potenziell ausgelassene (aber relevante) Variablen an **hsall** und **hspriv**, dann erhält man den Ausdruck auf der nächsten Seite:

"F statistic" ist die oben definierte Größe (Unterschied zwischen 0,269653 und 0,269645.. von oben wegen Rundungsfehler in der obigen Berechnung). Der relativ große Wert des prob values von $F \text{ Prob.F}(2,9) = 0,7696$ zeigt dass die H_0 der Irrelevanz der hinzugekommenen Regressoren Anzahlen von Hochschulen, **hsall** und **hspriv** nicht verworfen werden kann (die beiden Variablen sind also redundant)

Der equation output hier und auf der ersten Seite (das überladene [overfitted] Modell) unterscheidet sich nur durch den oberen Teil. Die Schätzungen der Koeffizienten $\hat{\beta}$ und deren Standardabweichungen $\hat{\sigma}_{\hat{\beta}}$ sowie die Maße der Güte der Anpassung (R^2 , \bar{R} etc.) gleich.

Omitted Variables: HSPRIV HSALL

F-statistic	0.269653	Prob. F(2,9)	0.769601
Log likelihood ratio	0.872944	Prob. Chi-Square(2)	0.646313

Test Equation:
 Dependent Variable: LEBEN
 Method: Least Squares
 Date: 01/06/10 Time: 12:47
 Sample: 1992 2006
 Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARZT	4.944828	1.550694	3.188784	0.0110
AUS	-0.911442	0.709270	-1.285042	0.2309
PKAUS	0.000547	0.001974	0.277313	0.7878
C	69.56570	5.672194	12.26434	0.0000
HSPRIV	-0.007013	0.029949	-0.234184	0.8201
HSALL	0.010446	0.017919	0.582938	0.5742

R-squared	0.981897	Mean dependent var	80.81333
Adjusted R-squared	0.971839	S.D. dependent var	0.966486
S.E. of regression	0.162188	Akaike info criterion	-0.510949
Sum squared resid	0.236744	Schwarz criterion	-0.227729
Log likelihood	9.832121	F-statistic	97.62896
Durbin-Watson stat	1.229789	Prob(F-statistic)	0.000000

3. Redundant variables Test

Geht man vom überladenen Modell (auf der ersten Seite) aus mit den fünf Regressoren und fragt, ob die letzten beiden **hspriv** und **hsall** redundant sind, so erhält man im Prinzip das gleiche Ergebnis:

Redundant Variables: HSPRIV HSALL

F-statistic	0.269653	Prob. F(2,9)	0.769601
Log likelihood ratio	0.872944	Prob. Chi-Square(2)	0.646313

Test Equation: Dependent Variable: LEBEN Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	72.02130	2.227328	32.33530	0.0000
ARZT	5.307830	1.158836	4.580311	0.0008
AUS	-0.958018	0.306725	-3.123373	0.0097
PKAUS	0.000624	0.000359	1.735713	0.1105

R-squared	0.980812	Mean dependent var	80.81333
Adjusted R-squared	0.975579	S.D. dependent var	0.966486
S.E. of regression	0.151036	Akaike info criterion	-0.719420
Sum squared resid	0.250930	Schwarz criterion	-0.530606
Log likelihood	9.395649	F-statistic	187.4231
Durbin-Watson stat	1.115907	Prob(F-statistic)	0.000000

Auch hier unterscheidet sich der equation output zu dem von Seite 2 (also vom underfitted model) nur durch den oberen Teil. Relevant ist vor allem die Angabe

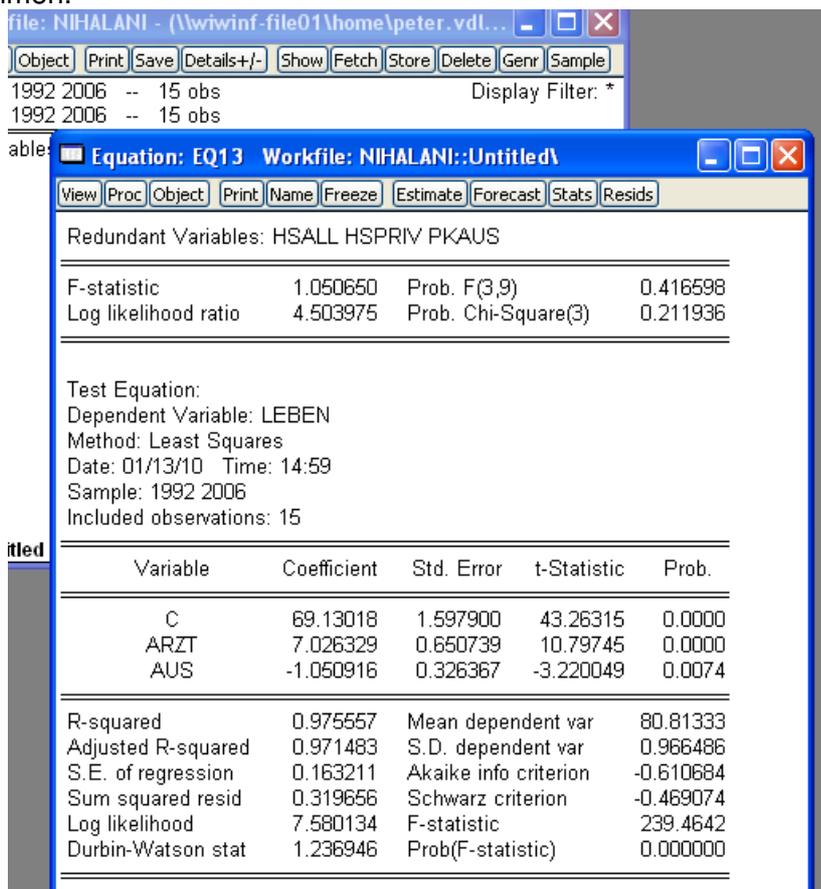
Log likelihood ratio	0.872944	Prob. Chi-Square(2)	0.646313
----------------------	----------	---------------------	----------

asymptotisch äquivalent ist das Ergebnis des F-Tests. Die H_0 wonach **hspriv** und **hsall** beide redundant sind wird also angenommen.

Man kann natürlich auch prüfen ob nicht auch noch weitere Variablen (Regressoren) "überflüssig" sind. Wir betrachten **pkaus** als einen möglicherweise weiteren redundanten Regressor und erhalten nebenstehenden screen shot. Man sieht an dem großen Wert von Prob F(3,9) = 0,417, dass wohl auch noch **pkaus** ein redundanter Regressor ist. Wir haben jetzt $L = 3$ Restriktionen ($H_0: \beta_3 = \beta_4 = \beta_5 = 0$ und $T - K - 1 = 15 - 5 - 1 = 9$) so dass die Prüfgröße $F_{3,9}$ verteilt ist.

Die verbleibenden Regressoren **arzt** und **aus** ergeben übrigens die eq_01 des oben erwähnten downloads A2 ("Mehr Ärzte ..."), dort auf Seite 7 (= Abb.5). Man sieht, es handelt sich um die gleiche Regressionsfunktion.

Die Situation ist sehr viel komplexer wenn man zwei Modelle vergleicht, bei denen nicht eines in dem anderen "genestet" ist (wie z.B. das "wahre" Modell im redundanten).



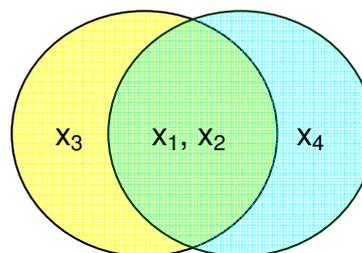
4. Der J- Test als non-nested Test zum Vergleich zweier linearer Modelle

Im Folgenden behandeln wir anhand eines anderen Datensatzes, nämlich der Variablen der **Übung F** (download F, bei dem es um die Erklärung der Zeitreihe der Geburten [geb = y] in Deutschland von 1965 bis 2003 geht)³ den Vergleich zweier Modelle, bei dem eines nicht im anderen "genestet" ist. Entscheidungen über non-nested Modelle kann man mit dem J-Test durchführen. Wir demonstrieren die rechnerische Durchführung des Tests mit EViews.

Gleichung	Allgemein	Im konkreten Beispiel
eq_1	$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_{1t}$	geb c ehen (x_1) ueber (x_2) quote (x_3)
eq_2	$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_4 x_{4t} + u_{2t}$	geb c ehen ueber kiged (kiged = x_4)

Das Verhältnis der Variablenmengen ist jetzt also etwas anders, wie das nebenstehende Bild veranschaulicht.

Das Verfahren des J-Test von Davidson und MacKinnon (1993) läuft wie folgt: Man schätzt eine Gleichung (eq_4) bei dem im **Modell 2** neben den dort vorgesehenen Regressoren auch die fitted values (Variable \hat{y}_1) aus der Gleichung eq_1 ist aufgenommen wird. (\hat{y}_1 ist **geffit_1** genannt), die auch den Einfluss von x_3 auf y enthält (Bild rechts unten auf dem screen shot auf der nächsten Seite [S. 5]).



³ Siehe auch dort (download F) zur Definition der Regressoren. Bei allen konkreten Beispielen – so auch bei diesem Übungsblatt H - geht es uns nicht um die Interpretation der Daten oder die Gewinnung sinnvoller Modelle. Die empirischen Beispiele werden hier nur zu dem Zweck der Demonstration einer Methode und ihrer konkreten Durchführung mit EViews präsentiert.

Geschätzt wird also die Gleichung

$$(eq_4) y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_4 x_{4t} + \lambda_1 \hat{y}_1 + v_{1t}$$

Ist der Koeffizient λ_1 von `gebfit_1` gegen Null gesichert (signifikant, Hypothese H_{02} wird verworfen) was hier der Fall ist (prob.value ist 0,0014), so ist, was im Modell 2 fehlt offenbar einflussreich, und das Modell 2 insofern abzulehnen.

Entsprechend verfährt man mit Modell 1 und einer Erweiterung um $\hat{y}_2 = \text{gebfit}_2$ (was den Einfluss von x_4 einschließt). Es geht jetzt um die Signifikanz von λ_2 ($H_{01}: \lambda_2 = 0$) in der Gleichung

$$(eq_3) y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \lambda_2 \hat{y}_2 + v_{1t}$$

Auf den Abbildungen unten ist zu sehen, wie man mit Befehlszeilen die entsprechenden Rechenvorgänge mit EViews durchführen kann.

Mit `equation eq_1.ls` wird die Gleichung 1 (`eq_1`) geschätzt und mit `eq_1.fit gebfit_1` werden die \hat{y} -Werte aus dieser Gleichung als Variable gesichert und `gebfit_1` genannt. Die übrigen Zeilen verstehen sich dann von selbst.

Realisierung des J-Tests mit EViews

The screenshot displays four EViews windows, each showing the results of a least squares regression. The windows are titled 'Equation: EQ_1', 'Equation: EQ_2', 'Equation: EQ_3', and 'Equation: EQ_4'. Each window contains a table of coefficients and various statistical measures.

Equation: EQ_1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	982950.0	153232.9	6.414745	0.0000
EHEN	0.395502	0.154948	2.552481	0.0152
UEBER	0.861302	0.036358	23.68961	0.0000
QUOTE	-5278.879	1590.111	-3.319818	0.0021

Equation: EQ_2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	623126.2	83945.01	7.423029	0.0000
EHEN	0.653748	0.134289	4.868212	0.0000
UEBER	0.780908	0.032348	24.14058	0.0000
KIGED	-6.721677	3.698377	-1.817467	0.0777

Equation: EQ_3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	492007.3	275417.1	1.786408	0.0830
EHEN	-0.464720	0.434846	-1.068700	0.2927
UEBER	0.065965	0.379673	0.173741	0.8631
QUOTE	-5293.743	1517.604	-3.488225	0.0014
GEBFIT_2	1.007781	0.479075	2.103600	0.0429

Equation: EQ_4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	134264.2	158060.7	0.849447	0.4016
EHEN	-0.202501	0.271893	-0.744781	0.4615
UEBER	-0.010778	0.228701	-0.047126	0.9627
KIGED	-6.773979	3.220184	-2.103600	0.0429
GEBFIT_1	1.002816	0.287486	3.488225	0.0014

Es mag anfänglich etwas irritierend sein, dass man keinen equation output oder eine Tabelle der Werte der geb_fit Variable⁴ erhält.⁵ Dass die Berechnungen durchgeführt worden sind sieht man am workfile window, wo die neuen Gleichungen und Variablen aufgelistet sind.

Nach diesem output ist das Modell 1 mit "quote" (=x₃) dem Modell 2 mit "kiged" (= x₄) vorzuziehen, weil gebfit_1 auf dem 5% und auch auf dem 1% Niveau signifikant ist, während gebfit_2 dies (beim 1% Niveau) nicht ist (prob.value 0,0429). Das Ergebnis ist nicht überraschend, weil "kiged" ja auch schon in der Ausgangsregression (eq_2) nichtsignifikant ist (prob.value nur 0,0777). Angesichts der Werte für \bar{R}^2 (Adjusted R-squared) ist der Unterschied zwischen den Modellen 1 und 2 nicht sehr groß (0,985826 vs. 0,981510), so dass nicht so offensichtlich ist, das Modell 1 zu bevorzugen ist. Um zu sehen, was der J-Test zeigt, wenn das Ergebnis ziemlich eindeutig ist vergleichen wir

$$eq_1: y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_{1t}$$

mit den (für geb) sehr relevanten Variablen x₂ = ueber und x₃ = quote) mit der ganz offensichtlich sehr viel schlechteren Gleichung

$$eq_5: y_t = \alpha + \beta_1 x_{1t} + \beta_4 x_{4t} + \beta_5 x_{5t} + u_{5t}$$

mit den wenig relevanten Variablen kiged (x₄) und kigaus (x₅) und den fitted values gebfit_5.

Tabellarische Übersicht über die Ergebnisse

Modell eq_1 bzw. erweitert um gebfit_5 (enthält Variablen des Modells 5 also x₁, x₄, x₅)

Variable	Coeff.	Prob.	Coeff.	Prob.
x ₁ ehen	0.395502	0.0152	0.225579	0.2330
x ₂ ueber	0.861302	0.0000	0.814331	0.0000
x ₃ quote	-5278.879	0.0021	-5429.998	0.0014
gebfit_5			0.110293	0.1222
\bar{R}^2 0.984611			\bar{R}^2 0.985248	

Modell eq_5 bzw. erweitert um gebfit_1 (enthält x₁, x₂, x₃)

Variable	Coeff.	Prob.	Coeff.	Prob.
x ₁ ehen	1.71075	0.0001	-0.108970	0.4625
x ₄ kiged	108.3115	0.0000	2.265332	0.7984
x ₅ kigaus	-980.420	0.0000	-73.19495	0.2788
gebfit_1			0.949733	0.0000
\bar{R}^2 0,847742			\bar{R}^2 0.986464	

Man sieht: wird das Modell 5 erweitert um gebfit_5 also die Variablen x₁, x₂, x₃ dann ist keiner der Regressoren x₁, x₄, x₅ noch signifikant (große prob. values, türkis markiert), sondern nur noch gebfit_1 ist signifikant.

Demgegenüber ist Modell 1 erheblich besser. Nimmt man gebfit_1 hinzu (was auch ehen enthält), so ist der Koeffizient für x₁ (= ehen) nicht mehr signifikant (der Einfluss ist in gebfit_5 enthalten) und gebfit_5 (also die Einflüsse von x₁, x₄, x₅) ist ebenfalls nichtsignifikant. Die aufgrund von \bar{R}^2 und den Ergebnissen des downloads F bereits vermutete Überlegenheit des Modells 1 gegenüber 5 wird also bestätigt.

Nicht immer sind die Verhältnisse so klar und eindeutig. Das Problem beim J Test ist, dass er neben der gewünschten Entscheidung für ein Modell und damit gegen das andere Modell (wie etwa hier für 1 und gegen 5) auch zu der unbefriedigenden Entscheidung führen kann, dass **beide** Modelle oder **keines** der beiden Modelle angenommen werden sollte.

⁴ gebfit-1 oder gebfit-2 wäre nicht zulässig, weil – statt _ als minus gelesen wird.

⁵ das Programm führt die Rechnung sofort nachdem man mit "enter" eine Befehlszeile abgeschlossen hat aus.

5. Interpretation des "(equation) estimation output" und inhaltliche Erwägung bei der Modellwahl: Wird Zufriedenheit mehr bestimmt vom Bruttoinlandsprodukt oder von einem Wohlfahrtsmaß?

Die folgenden Überlegungen sind sowohl statistisch als auch inhaltlich interessant. Es wird immer wieder darüber geklagt, dass das Bruttoinlandsprodukt (BIP oder GDP) nicht tauglich ist um das "Glück" (bzw. das Gefühl glücklich und zufrieden zu sein) zu messen und behauptet, das BIP solle besser durch ein Maß der Wohlfahrt ersetzt werden.⁶

In einer empirischen Untersuchung⁷ aus den USA wurde die Eignung des

x_1 : pro-Kopf-BIP (GDP per capita = **PCGDP92** und **PCGDP95**), als das übliche Maß des Wachstums und der

x_2 : "human development index" **HDI**⁸

auf ihre Eignung überprüft die Variable **HAP**⁹ als abhängige Variable y zu bestimmen. Als weitere Regressoren wurden geprüft:

x_3 : **GINI** die Einkommensdisparität gemessen am Gini-Koeffizienten sowie

x_4 : **HEALTH** ein Indikator für den Gesundheitszustand (Anteil der Gesundheitsausgaben am BIP)

sowie einige andere und es wurden auch alternative Maße für y betrachtet.¹⁰

Folgende Schätzergebnisse wurden mitgeteilt:

	(1)	(2)	(3)	(1a)	(2a)	(3a)
C	10.972*	19.819	53.966*	-4.902	-4.618	40.900*
x_1 PCGDP92	0.001* (4.689)		0.002* (7.853)	0.001* (3.328)		0.001* (6.894)
x_2 HDI		4.062 (0.237)	- 63.206* (- 5.326)		5.802 (0.294)	- 59.648* (-4.796)
x_3 GINI				0.432* (3.692)	0.357** (2.104)	0.264 (1.850)
R²	0.275	0,001	0,495	0.332	0.057	0.536
n	40	38	38	36	34	34

* bzw. ** heißt signifikant auf dem 1% bzw. 5% Niveau (zweiseitig). Es sind auch (abgesehen von C) – wie üblich – die **t-Werte in Klammern mitgeteilt**. Es ist z.B. klar, dass $t = 4,689$ bei $40-2 = 38$ Freiheitsgraden signifikant ist.

Es gibt selten so viele Hinweise für die Überlegenheit eines Regressors gegenüber einem andren wie das hier der Fall ist.

1) **t-Test**: Wie man sieht ist das pro-Kopf BIP (CPGDP oder x_1) für sich genommen signifikant (Gl. 1), HDI dagegen nicht (Gl. 2). Das gilt auch wenn man GINI (x_3) als Maß der Einkommensdisparität hinzu nimmt (Gl. 1a und 2a). Auffallend ist auch der geringe Erklärungswert

⁶ Zuletzt gab der sog. Stiglitz Report (auf Veranlassung des französischen Präsidenten Sarkozy) wieder Anlass, dieses Thema zu diskutieren. Vgl das "Zeitgespräch" im Heft Januar 2010 der Zeitschrift "Wirtschaftsdienst".

⁷ Dipterto, W. R. and E. Anoruo, GDP per capital and its challengers as measures of happiness, International Journal of Social Economics, Vol 33, no. 10, 2006, pp. 698 – 709. Verwendet wurden Daten für n Länder.

⁸ In den HDI der UNO gehen Variablen ein wie Lebenserwartung, Bildungsstand und Kaufkraft (Realeinkommen). Er ist gedacht als Wohlfahrtsmaß und als eine Alternative zum BIP zur Wohlfahrtsmessung (wobei mit dem BIP ja generell gar nicht die Absicht verbunden ist, die Wohlfahrt zu messen, sondern die Produktionsleistung eines Landes).

⁹ Anteil der Menschen, die sich in einem Land bei einer weltweiten Befragung (1990-93 World Value Survey of the University of Michigan) als "very happy" bezeichnet hatten.

¹⁰ Es ist hier nicht der Ort, die Studie insgesamt zu referieren.

von x_2 wenn man die Bestimmtheitsmaße R^2 vergleicht: nur 0.001 (!) bei Gl. 2 gegenüber 0.275 bei Gl. 1.

2) F-Test: Im Falle der einfachen Regression laufen natürlich die Betrachtung von R^2 und der F-Test einerseits und die Betrachtung der Signifikanz (t-Test) des einen Regressors andererseits auf das Gleiche hinaus.

3) Hinzukommen eines Regressors: Aussagefähig ist auch, dass beim Übergang von Gl. 2 zu Gl. 3 R^2 von 0,001 auf 0,495 zunimmt was nach der Rekursionsformel

$$R_{y.12}^2 = R_{y.2}^2 + (1 - R_{y.2}^2)r_{y1.2}^2$$

für $r_{y1.2}^2 = 0.494$ spricht. Im Unterschied dazu steigert sich die Bestimmtheit beim Übergang von Gl. 1 zu 3 wenig, von 0.275 auf 0.495 was für eine geringere partielle Bestimmtheit des HDI Maßes (x_2) spricht: $r_{y2.1}^2 = 0,303$

4) Vorzeichen der Regressionskoeffizienten: Sehr bemerkenswert ist, dass HDI ein negatives Vorzeichen erhält wenn es zusammen mit dem BIP in der Regressionsgleichung erscheint, d.h. mehr Wohlfahrt im Sinne von HDI trägt eher zu weniger statt zu mehr Glück bei. Mithin ist der HDI nicht nur weniger mit dem Glück korreliert als das viel geschmähte BIP, er misst sogar in die falsche Richtung. Das negative Vorzeichen spricht *erheblich* gegen den HDI als Indikator für das Glücks- bzw. Zufriedenheitsempfinden der Bevölkerung.

Das Argument gegen den HDI wird noch dadurch verstärkt, dass sich die Aussagen über den Erklärungsbeitrag und das Vorzeichen auch wiederholen, wenn man statt Gl. 3 (ohne GINI) Gl. 3' (mit GINI) betrachtet. Der HDI leistet also einen *hochsignifikanten negativen* Beitrag zur Erklärung der Zufriedenheit.

5) Beträge der Regressionskoeffizienten: der absolute Betrag eines Regressionskoeffizienten ist nicht aussagekräftig und vor allem ist Vorsicht geboten beim Vergleich dieser Koeffizienten untereinander, etwa von 0,002 für x_1 und -63,206 für x_2 in Gl. 3 auf Seite 7, denn:

1. sind die betrachteten Koeffizienten stets nur Stichproben-Schätzwerte $\hat{\beta}_1, \hat{\beta}_2, \dots$ und nicht gleichzusetzen mit den wahren Regressionskoeffizienten β_1, β_2, \dots der Grundgesamtheit
2. sind sie abhängig von der Größenordnung und damit Streuung der zu vergleichenden Variablen x_1, x_2 etc. Man kann nicht die "rohen" Regressionskoeffizienten β_i miteinander vergleichen,¹¹ sondern bestenfalls die standardisierten Regressionskoeffizienten β_i^* für die gilt: $\beta_i^* = \beta_i \cdot \frac{s_i}{s_y}$. Es kann also sehr wohl $\beta_i < \beta_j$ und trotzdem $\beta_i^* > \beta_j^*$ gilt, und zwar einfach deshalb, weil die Standardabweichung von x_i größer ist als die von x_j , und
3. ist stets das Problem der Kollinearität zu berücksichtigen, d.h. Vergleiche hinsichtlich der relativen Bedeutung (Wichtigkeit) eines Einflussfaktors wären eigentlich nur zulässig, wenn diese statistisch unabhängig *wären* (was sie so gut wie nie sind).

6) ein unerwartetes Ergebnis: Bezieht man die Variable HEALTH in die Betrachtung ein, so erhält man die Regressionsfunktion $y = 5.413 + 0.002x_1 + -0.245x_4$ ($R^2 = 0.665$), also auch hier (überraschend) ein negatives Vorzeichen.

Fazit: Diese Übung H gibt einige Hinweise zur Spezifikation und Fehlspezifikation. Sie zeigt, dass neben formaler Kriterien (R^2 , Signifikanz von Regressoren [Regressionskoeffizienten] einzeln [t-Test] und zu mehreren zusammengefasst [F-Test] auch inhaltliche Aspekte und Fragen der Interpretation (Plausibilität) betrachtet werden sollten (z.B. hat ein Regressionskoeffizienten das "richtige" Vorzeichen?).

¹¹ Es wäre z.B. Unsinn, aufgrund von Gl. 3 zu folgern, dass der HDI mit 63 bedeutsamer sei als das BIP mit nur 0,002.