



Übung zur Empirischen Wirtschaftsforschung

VI Einkommensfunktion – Teil 2

Schätzung von Einkommensfunktionen (Teil 2)

Test auf Strukturkonstanz

Literatur

Winker, P. (2007), Empirische Wirtschaftsforschung, Kap. 9.

Franz, W. (2003), Arbeitsmarktökonomik, 5. Auflage, Kap. 3.

Hübler, O. (2005), Einführung in die empirische Wirtschaftsforschung, Kap. 2 & 3.

Smolny, W., and Kirbach, M. (2004): Wage differentials between East and West Germany - Is it related to the location or to the people?

Schätzung von Einkommensfunktionen

Das multiplikative Modell

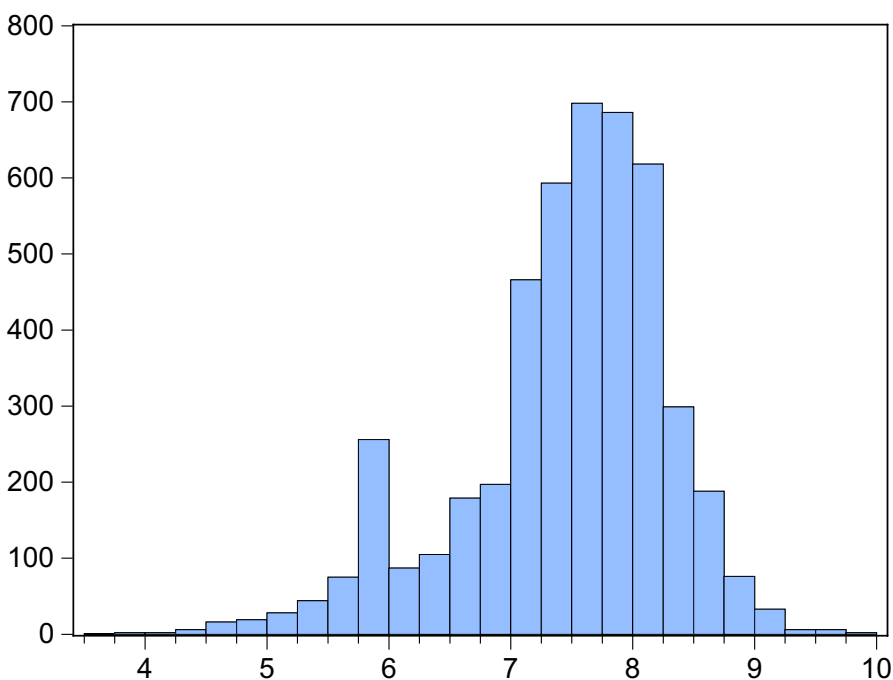
Ein multiplikatives Modell erklärt die Daten besser, siehe dazu auch die theoretische Herleitung im Lehrbuch in Franz, W. (2003), Arbeitsmarktökonomik, S. 90ff.

$$y = e^{\beta_0} \cdot e^{\beta_1 x_1} \cdot e^{\beta_2 x_2} \cdot e^{\beta_3 x_3} \cdot e^{\varepsilon}$$

bzw.

$$\log(\text{Eink}) = \beta_0 + \beta_1 \cdot \text{Schule} + \beta_2 \cdot \text{Beruf} + \beta_3 \cdot \text{Berufserfahrung} + \varepsilon$$

Die Koeffizienten β_1 bis β_3 sind als logarithmische Prozente zu interpretieren. Ein Koeffizient von 0,7 bedeutet eine Verdoppelung des Einkommens ($e^{0,7} \approx 2$); ein Koeffizient von 1,4 bedeutet eine Vervielfachung des Einkommens ($e^{1,4} \approx 4$).



Series: LOGEINK	
Sample 1 61545	
Observations 4688	
Mean	7.458258
Median	7.603898
Maximum	9.903488
Minimum	3.688879
Std. Dev.	0.839740
Skewness	-0.895145
Kurtosis	3.978039
Jarque-Bera	812.9187
Probability	0.000000

- Das Histogramm von $\log(\text{Eink})$ ähnelt nunmehr eher einer Normalverteilung als das Histogramm des Einkommens selbst.
- Das arithmetische Mittel hat den Wert 7,458.
- Das arithmetische Mittel des logarithmierten Einkommens beträgt $e^{7,458} = 1.733,68 \text{ €}$.

Die Erklärung des Einkommens durch den Schul- und Berufsabschluss im multiplikativen Modell ergibt:

```

=====
Dependent Variable: LOGEINK
Method: Least Squares
Date: 06/16/09   Time: 12:56
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4471
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.087915	0.049453	143.3266	0.0000
HAUPT	0.122211	0.054227	2.253710	0.0243
REAL	0.097313	0.052486	1.854066	0.0638
FACH	0.396862	0.069861	5.680730	0.0000
ABI	0.206554	0.057708	3.579279	0.0003
LEHRE	0.211999	0.025817	8.211660	0.0000
MEIST	0.431721	0.045774	9.431494	0.0000
FH	0.378372	0.043122	8.774405	0.0000
UNI	0.715303	0.045729	15.64211	0.0000

```

=====
R-squared          0.119233      Mean dependent var 7.485283
Adjusted R-squared 0.117654      S.D. dependent var 0.821856
S.E. of regression 0.771996      Akaike info criteri2.322337
Sum squared resid  2659.256      Schwarz criterion  2.335231
Log likelihood     -5182.585      Hannan-Quinn criter2.326882
F-statistic        75.50504      Durbin-Watson stat 1.828949
Prob(F-statistic)  0.000000
=====

```

- Personen, die weder einen Schul- noch einen Berufsabschluss haben verdienen $e^{7,088} = 1.197,51 \text{ €}$.
- Personen, die einen Hauptschulabschluss, aber keinen Berufsschulabschluss haben, verdienen $e^{7,088+0,122} = e^{7,188} = 1.352,89 \text{ €}$.
- Anders aufgeschrieben verdienen diese Personen $e^{7,088} \cdot e^{0,122} = 1.197,51 \cdot 1,1298 = 1.352,89 \text{ €}$.
Sie haben also ein 12,98% höheres Einkommen.
- Personen mit Abitur und Universitätsabschluß verdienen $e^{7,088+0,207+0,715} = e^{8,010} = 3.010,92 \text{ €}$.
- Anders aufgeschrieben verdienen diese Personen $e^{7,088} \cdot e^{0,207} \cdot e^{0,715} = 1.197,51 \cdot 1,230 \cdot 2,044 = 3.010,92 \text{ €}$.
Sie haben ein $1,230 \cdot 2,044 = 2,514$ mal höheres Einkommen.

Die Erklärung des Einkommens durch den Schul- und Berufsabschluss sowie Berufserfahrung im multiplikativen Modell ergibt:

```

=====
Dependent Variable: LOGEINK
Method: Least Squares
Date: 06/16/09   Time: 13:10
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4463
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.329474	0.063755	99.27868	0.0000
HAUPT	0.083321	0.052449	1.588604	0.1122
REAL	0.068052	0.051188	1.329447	0.1838
FACH	0.423212	0.068079	6.216503	0.0000
ABI	0.306358	0.057509	5.327143	0.0000
LEHRE	0.157390	0.024805	6.344996	0.0000
MEIST	0.330522	0.044015	7.509333	0.0000
FH	0.276673	0.041631	6.645844	0.0000
UNI	0.590413	0.044220	13.35168	0.0000
EX	0.073560	0.003704	19.85995	0.0000
EX^2	-0.001335	7.39E-05	-18.07114	0.0000

```

=====
R-squared          0.191064      Mean dependent var 7.488501
Adjusted R-squared 0.189247      S.D. dependent var 0.818876
S.E. of regression 0.737330      Akaike info criteri2.230900
Sum squared resid  2420.357      Schwarz criterion  2.246683
Log likelihood     -4967.254      Hannan-Quinn criter2.236464
F-statistic        105.1527      Durbin-Watson stat 1.891915
Prob(F-statistic)  0.000000
=====

```

- Die Berufserfahrung ist keine Dummyvariable.
- Der Koeffizient der Variablen EX ist positiv, d.h. ein Jahr mehr Berufserfahrung bedeutet ein höheres Einkommen.
- Der Koeffizient des quadratischen Terms EX^2 ist negativ, d.h. ein Jahr mehr Berufserfahrung für einen Neuling hat einen größeren Effekt auf das Einkommen als ein Jahr mehr Berufserfahrung bei jemandem, der z. B. bereits 20 Jahre Berufserfahrung hat.
- Der Effekt auf das Einkommen beträgt: $\frac{\delta \log(EINK)}{\delta (EX)} = 0,0736 - 2 \cdot 0,0013 \cdot EX$.
- Das erste Jahr Berufserfahrung erhöht das Einkommen um $0,0736 - 0,0026 \cdot 1 = 0,0710$ bzw. 7,1%.
- Das zwanzigste Jahr Berufserfahrung erhöht das Einkommen um $0,0736 - 0,0026 \cdot 20 = 0,0216$ bzw. 2,16%.

Eine weitere Möglichkeit der Einkommenserklärung bietet die Variable DAUER. Operationalisiert wird die unabhängige Variable durch die bei einem speziellen Schul- und Berufsabschluss üblicherweise notwendige Zahl an Schuljahren. Für das logarithmierte Einkommen im multiplikativen Modell ergibt sich damit folgende Schätzung:

```

=====
Dependent Variable: LOGEINK
Method: Least Squares
Date: 06/16/09   Time: 13:17
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4365
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.575248	0.063751	56.08139	0.0000
DAUER	0.085375	0.003037	28.11334	0.0000
SEX	-0.210873	0.017180	-12.27422	0.0000
EX	0.063064	0.002586	24.38350	0.0000
EX^2	-0.001058	5.28E-05	-20.03435	0.0000
STUND	0.088768	0.002120	41.87789	0.0000
STUND^2	-0.000748	2.75E-05	-27.24868	0.0000

```

=====
R-squared          0.612058      Mean dependent var 7.492762
Adjusted R-squared 0.611524      S.D. dependent var 0.818811
S.E. of regression 0.510347      Akaike info criteri1.494152
Sum squared resid  1135.060      Schwarz criterion  1.504386
Log likelihood     -3253.987      Hannan-Quinn criter1.497763
F-statistic        1145.940      Durbin-Watson stat 1.527837
Prob(F-statistic)  0.000000
=====

```

- Der Koeffizient der Variablen DAUER ist positiv, d.h. ein Jahr mehr an Schul- bzw. Berufserfahrung bedeutet ein höheres Einkommen. Ein weiteres (Aus-)bildungsjahr erhöht das Einkommen demzufolge um jeweils 8,5%.
- Frauen müssen mit einem Einkommensabschlag von ca. 21% rechnen.
- Das erste Jahr Berufserfahrung erhöht das Einkommen um $0.0630 - 0.0011 \cdot 1 = 0,0619$ bzw. 6,19%, das zwanzigste Jahr erhöht das Einkommen hingegen nur um $0.0630 - 0.0011 \cdot 20 = 0,041$ bzw. 4,10%.
- Der Effekt einer Arbeitsstunde auf das Einkommen beträgt: $\frac{\delta \log(EINK)}{\delta(Stund)} = 0.089 - 2 \cdot 0.00075 \cdot Stund.$

Test auf Strukturkonstanz

Unter den bisherigen Schätzungen wurde immer von der Annahme ausgegangen, daß die Parameter des "wahren" ökonomischen Modells β für alle Beobachtungseinheiten hinweg konstant ist. Diese Annahme ist jedoch nicht zwingend.

Für Individualdaten kann es Gründe geben, daß es Unterschiede im Verhalten einzelner Teilgruppen gibt. Wird die Möglichkeit von Strukturbrüchen vernachlässigt, ergibt die Kleinste-Quadrate-Schätzung verzerrte Werte für die Parameter.

Um zu überprüfen, ob die Unterschiede zwischen Ost und West signifikant sind, wird der F-Test verwendet, wobei die Summen der Residuenquadrate der gemeinsamen und der getrennten Schätzung miteinander verglichen werden. Ein Verfahren zur Überprüfung stellt der Wald-Test dar. Wird die Nullhypothese der Parameterkonstanz, d.h die Koeffizienten für Ost- und Westdeutschland sind gleich, verworfen, liegt ein Strukturbruch der Koeffizienten vor.

$$F_{k, N+M-2k} = \frac{(SSR_R - SSR_{UR})/k}{(SSR_{UR})/(N + M - 2k)}$$

$$SSR : \sum \hat{\epsilon}_i^2 = \sum (y_i - \hat{y}_i)^2$$

wobei SSR_R : Sum of squared residuals, restricted model, SSR_{UR} : Sum of squared residuals, unrestricted model, N, M: Anzahl der Beobachtungen im ersten (zweiten) restriktiven Modell, k: Anzahl der Koeffizienten.

Beispiel

Einkommensgleichung für Gesamtdeutschland (ohne Berlin)

(1)

```
=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 13:24
Sample: 1 61545 IF PSAMPLE<=3 AND (OST=0 OR OST=1)
Included observations: 4332
=====
      Variable      Coefficient Std. Error t-Statistic Prob.
=====
           C          -50.55042   109.4696  -0.461776   0.6443
          DAUER         239.2957    8.483654   28.20669   0.0000
           SEX        -1244.961   44.47103  -27.99488   0.0000
=====
R-squared          0.265766   Mean dependent var 2335.646
Adjusted R-squared 0.265427   S.D. dependent var 1704.276
S.E. of regression 1460.688   Akaike info criteri17.41190
Sum squared resid  9.24E+09   Schwarz criterion  17.41631
Log likelihood     -37711.17   Hannan-Quinn criter17.41345
F-statistic        783.4701   Durbin-Watson stat 1.551796
Prob(F-statistic)  0.000000
=====
```

Einkommensgleichung für Ost (ohne Berlin)

(2)

```
=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 13:25
Sample: 1 61545 IF PSAMPLE<=3 AND OST=1
Included observations: 1233
=====
      Variable      Coefficient Std. Error t-Statistic Prob.
=====
           C          -765.8280   175.3094  -4.368436   0.0000
          DAUER         228.0826   13.19875   17.28062   0.0000
           SEX        -535.6160   63.86990  -8.386047   0.0000
=====
R-squared          0.227601   Mean dependent var 1910.740
Adjusted R-squared 0.226345   S.D. dependent var 1274.460
S.E. of regression 1120.985   Akaike info criteri16.88423
Sum squared resid  1.55E+09   Schwarz criterion  16.89668
Log likelihood     -10406.13   Hannan-Quinn criter16.88892
F-statistic        181.2210   Durbin-Watson stat 1.821849
Prob(F-statistic)  0.000000
=====
```


Einkommensgleichung für West (ohne Berlin)

(3)

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 13:25
Sample: 1 61545 IF PSAMPLE<=3 AND OST=0
Included observations: 3099
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.866180	129.4631	0.037587	0.9700
DAUER	260.9421	10.15061	25.70705	0.0000
SEX	-1500.132	54.30561	-27.62388	0.0000

```

=====
R-squared          0.315567   Mean dependent var 2504.703
Adjusted R-squared 0.315125   S.D. dependent var 1820.494
S.E. of regression 1506.590   Akaike info criteri17.47405
Sum squared resid  7.03E+09   Schwarz criterion  17.47990
Log likelihood     -27073.04   Hannan-Quinn criter17.47615
F-statistic        713.7253   Durbin-Watson stat 1.564403
Prob(F-statistic)  0.000000
=====

```

$$F = \frac{(9,24 - (1,55 + 7,03))/3}{(1,55 + 7,03)/(3099 + 1233 - 6)} = 110,92$$

Da die F-Statistik einen Wert ausweist, der größer ist als der kritische Wert ($F_{krit,95\%} = 2,10$)¹, muß die Nullhypothese verworfen werden. Es liegt hier keine Strukturkonstanz vor; eine Schätzung der Einkommen sollte daher getrennt für West- und Ostdeutschland durchgeführt werden.

¹Der kritische Wert ist den Tabellen zu entnehmen, in EViews z. B. mit `genr fkrit=@qfdist(0.95,6,4326)`.

Vergleich der Koeffizienten und Überprüfung auf Strukturkonstanz

(4)

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/16/09   Time: 13:34
Sample: 1 61545 IF PSAMPLE<=3
Included observations: 4332
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.866180	120.9687	0.040227	0.9679
DAUER	260.9421	9.484605	27.51218	0.0000
SEX	-1500.132	50.74251	-29.56360	0.0000
OST	-770.6941	251.2001	-3.068049	0.0022
DAUER*OST	-32.85954	19.09688	-1.720676	0.0854
SEX*OST	964.5155	94.91133	10.16228	0.0000

```

=====
R-squared                0.318505      Mean dependent var 2335.646
Adjusted R-squared      0.317717      S.D. dependent var 1704.276
S.E. of regression      1407.739      Akaike info criteri17.33874
Sum squared resid       8.57E+09      Schwarz criterion 17.34757
Log likelihood          -37549.72     Hannan-Quinn criter17.34186
F-statistic             404.3612     Durbin-Watson stat 1.604855
Prob(F-statistic)      0.000000
=====

```

In EViews kann die Nullhypothese $H_0 : \beta_{OST} = \beta_{DAUER*OST} = \beta_{SEX*OST} = 0$ im Equation-Fenster mit einem Klick auf View → Coefficient Test → Wald - Coefficient Restrictions... oder dem Befehl `equation.wald` mit Hilfe eines Wald-Tests gegen die entsprechende Alternative H_A getestet werden.

(5)

```

=====
Wald Test:
Equation: Untitled
=====

```

Test Statistic	Value	df	Probability
F-statistic	111.5917 (3, 4326)		0.0000
Chi-square	334.7752	3	0.0000

```

=====

```