



# Übung zur Empirischen Wirtschaftsforschung

## VIII. Einkommensfunktion

- 5.1 [Querschnittsdaten](#)
- 5.2 [Dummyvariablen](#)
- 5.3 [Humankapitaltheorie](#)
- 5.4 [Individualdaten des SOEP](#)  
*Sozio-ökonomisches Panel (SOEP)*
- 5.5 [Schätzung von Einkommensfunktionen](#)  
*Additives vs. multiplikatives Modell*
- 5.6 [Test auf Strukturkonstanz](#)

### [Literatur](#)

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

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

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

## 5.1 Querschnittsdaten

**Zeitreihendaten** (time series) sind Daten, die über die Zeit wiederholt erhoben werden, i.d.R. mit einer regelmäßigen Frequenz. Die Daten werden nach Möglichkeit bei denselben Merkmalsträgern und mit gleicher Methodik und Definition erhoben. Bei Zeitreihendaten liegt zu einer Variable ein Beobachtungswert pro Periode vor, z.B. der:

- aggregierte Konsum pro Quartal von 1960-2011,
- kurzfristige Dreimonatszinssatz pro Monat von 1975-2011,
- DAX-Wert pro Tag von 1993-2011,
- ...

In der Gleichung

$$y_t = \beta_0 + \beta_1 \cdot x_t + \varepsilon_t$$

ist  $t$  der Zeitindex.

**Querschnittsdaten** beziehen sich auf zeitlich ungeordnete Beobachtungen für unterschiedliche Merkmalsträger, die meist zu einem einheitlichen Zeitpunkt erhoben werden. Es liegen somit  $i$  Beobachtungen zu einem Zeitpunkt vor, z.B.:

- die Mietkosten verschiedener Wohnungen in einer Stadt am 1. Juli 2012,
- das Einkommen von Arbeitnehmern im Mai 2010,
- der Exportanteil von Unternehmen im Jahr 2008,
- ...

In der Gleichung

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$$

ist  $i$  kein Zeitindex, sondern die Nummer des  $i$ -ten Mieters, Arbeitnehmers bzw. der  $i$ -ten Firma.

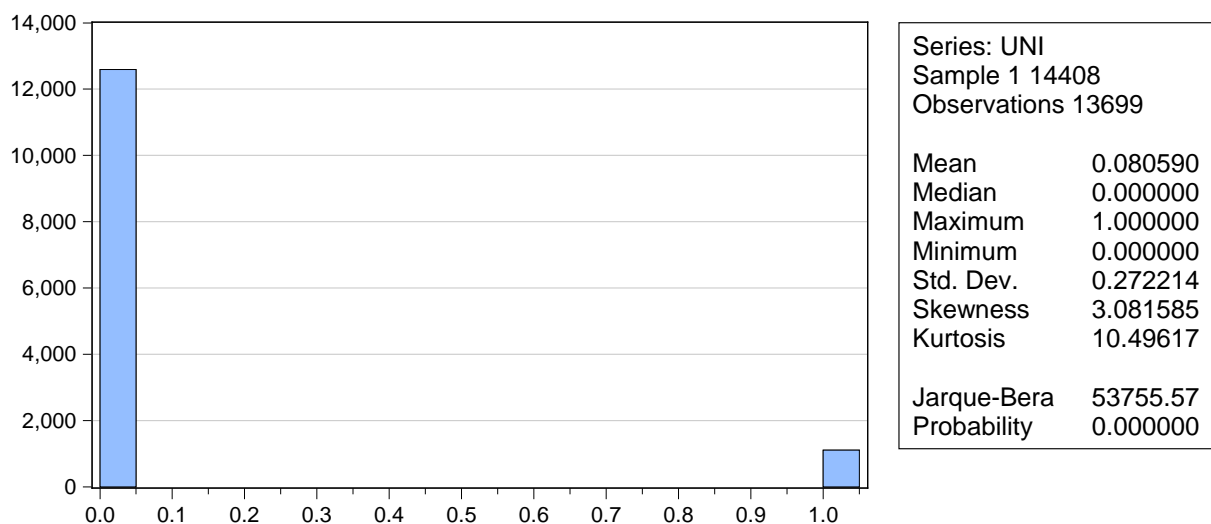
Querschnittsanalysen können ebenso wie Zeitreihenuntersuchungen in EViews durchgeführt werden. Hierzu wird ein Workfile mit *nicht datierten* Daten benötigt. Der Befehl "create u 1 32000" erzeugt hierbei ein Workfile mit 32.000 Beobachtungen.

Querschnittsdaten haben in der Regel eine höhere Varianz als Zeitreihendaten. Daher ist ein geringerer Wert für das Bestimmtheitsmaß  $R^2$  zu erwarten.

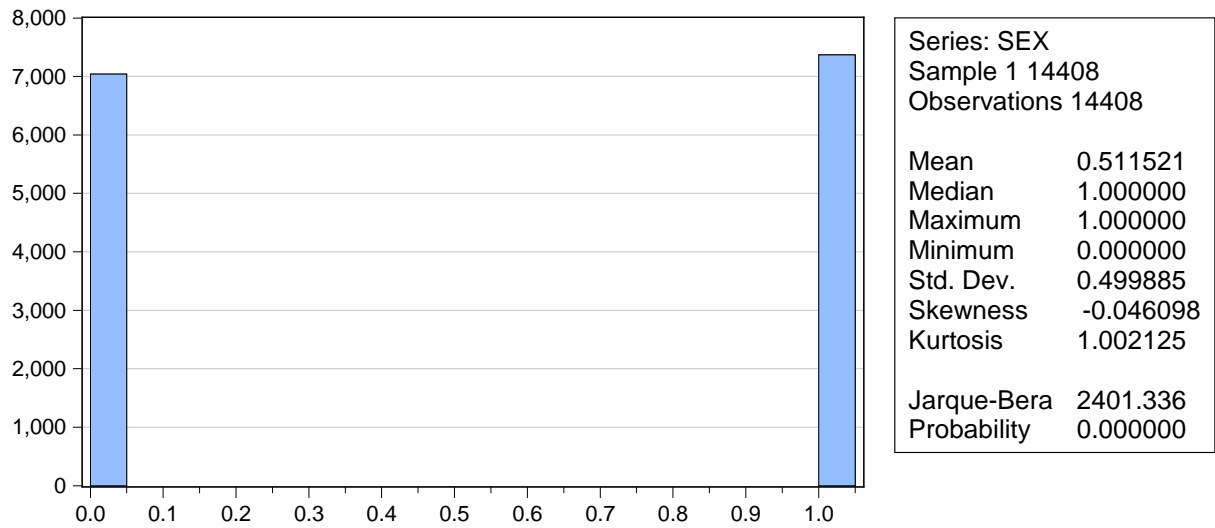
## 5.2 Dummyvariablen

Dummyvariablen sind keine quantitativen, sondern qualitative Variablen. Sie nehmen nur 2 Ausprägungen an, z.B.:

- @seas(1) ist 1 im ersten Quartal, 0 in den anderen drei Quartalen.
- Die Variable UNI hat die Ausprägungen 1 für einen Universitätsabschluss und 0 für keinen Universitätsabschluss.
- Die Variable SEX hat die Ausprägungen 1 für Frau und 0 für Mann.
- ...



- Der Maximalwert ist 1, d.h. ein Universitätsabschluss.
- Der Minimalwert ist 0, d.h. kein Universitätsabschluss.
- Der Median ist 0, d.h. es haben mehr Personen keinen Universitätsabschluss.
- Das arithmetische Mittel beträgt 0.080, d.h. 8% der 13.699 Personen haben einen Universitätsabschluss (1.095 Personen).



- Der Maximalwert ist 1, d.h. die Person ist weiblich.
- Der Minimalwert ist 0, d.h. die Person ist männlich.
- Der Median ist 1, d.h. unter den beobachteten Personen befinden sich mehr Frauen.
- Das arithmetische Mittel ist 0.511, d.h. 51,1% der 14.408 Personen sind Frauen.

## 5.3 Humankapitaltheorie

Nach der neoklassischen Theorie entspricht der Lohn im Gleichgewicht dem Grenzprodukt der Arbeit. Im Rahmen der Humankapitaltheorie (als Erweiterung des neoklassischen Modells) richtet sich der Lohn nach der Produktivität und dem Humankapitalbestand, d.h. Individuen investieren in ihr Humankapital, um durch eine höhere Produktivität einen höheren Reallohn zu erhalten.

Die Akkumulation des Humankapitals eines Erwerbstätigen bestimmt sich analog zu Sachkapitalinvestitionen. Formal lässt sich das mittels folgender Gleichung darstellen:

$$HK_t = HK_{t-1} + I_{HK} - \delta \cdot HK .$$

Der Humankapitalbestand einer Periode  $t$  ergibt sich aus dem Humankapitalbestand der Vorperiode ergänzt um Investitionen in das Humankapital  $I_{HK}$  und korrigiert um Abschreibungen  $\delta$  auf Humankapital. Investitionen werden durchgeführt, falls der Ertrag höher ist als die anfallenden Kosten:

$$\sum_{t=1}^T = \frac{w_t^q - w_t}{(1+r)^t} - C_0 > 0 .$$

Der Ertrag der Investition entspricht der Summe der Differenzen des Reallohns  $w^q$  abzüglich des Reallohns  $w$ , den das Individuum ohne die Investition in Humankapital realisiert hätte bis zur Endperiode  $T$ , in der ein Erwerbstätiger aus dem Erwerbsleben ausscheidet. Der Ertrag jeder Periode wird um einen Abzinsungsfaktor  $r$  korrigiert. In der Periode, in der die Fort- und Weiterbildung durchgeführt wird, entstehen Kosten  $C_0$ .

Schul-, Aus-, Fort- und Weiterbildung sind Investitionen in das Humankapital, von denen Wirtschaftssubjekte zukünftige Erträge erwarten. Wenn die Summe der zukünftigen Erträge höher ist als die anfallenden Kosten, werden diese Investitionen durchgeführt, andernfalls wird, zumindest aus finanzieller Sicht, auf sie verzichtet. Die Verteilung der Arbeitseinkommen richtet sich demzufolge nach der Investition in (Aus-) Bildung, dem Humankapitalbestand. Personen mit einem höheren Schul- und Berufsabschluß sollten somit ein größeres Einkommen haben als Personen mit einer niedrigen Schul- bzw. Berufsausbildung. Einkommensunterschiede werden demzufolge durch die Dauer und die Kosten der Ausbildung bestimmt.

Im Folgenden soll anhand eines empirischen Modells untersucht werden, inwiefern Investitionen in Humankapital, z.B. in Form unterschiedlicher Schul- und Berufsabschlüsse zu einem höheren Einkommen führen. Als Datengrundlage dienen Zahlen aus dem Sozio-ökonomischen Panel (SOEP)

## 5.4 Individualdaten des SOEP

Das sozio-ökonomische Panel (SOEP) ist eine repräsentative Wiederholungsbefragung privater Haushalte in Deutschland, die seit 1983 im jährlichen Rhythmus bei den selben Personen und Familien vom DIW Berlin und Infratest durchgeführt wird. Sie dient der Untersuchung des Wandels objektiver Lebensbedingungen und der subjektiv wahrgenommenen Lebensqualität und liefert dabei u.a. Informationen über:

- Erwerbs- und Familienbiographien
- Erwerbsbeteiligung und berufliche Mobilität
- Einkommensverläufe
- Gesundheit
- Lebenszufriedenheit
- gesellschaftliche Partizipation und Zeitverwendung
- Persönlichkeitsmerkmale
- Haushaltszusammensetzung, Wohnsituation

Dabei enthält es unterschiedliche Stichproben:

- A Deutsche (West)
- B Ausländer (West)
- C Deutsche (Ost)
- D Zuwanderer 1984–1993
- E Ergänzung 1998
- F Ergänzung 2000
- G Hocheinkommensbezieher 2002

Auf der Website <http://panel.gsoep.de/soepinfo> kann man z. B. die entsprechenden Fragebögen einsehen. Im Workfile `Uebung8.wf1` sind ausgewählte Variablen aus der aktuellen Welle von 2010 enthalten. Das Sample ist mit Hilfe einer Zufallsvariablen auf die Hälfte der Beobachtungen beschränkt.

Daten im Workfile Uebung8.wf1

*EINK* Bruttomonatseinkommen in €

*ALTER* Alter in Jahren

*SEX* Geschlecht, Dummyvariable, 1 für Frauen, 0 für Männer

*OST* Ostdeutschland

*WEST* Westdeutschland

*OHNE* kein Schulabschluss

*HAUPT* Hauptschulabschluss

*REAL* Realschulabschluss

*FACH* Fachschule

*ABI* Abitur

*KEIN* ohne Berufsabschluss

*LEHRE* Lehre

*MEIST* Meister

*FH* Fachhochschulabschluss

*UNI* Universitätsabschluss

*TENUB* Betriebszugehörigkeit

*EX* Berufserfahrung

*DAUER* Dauer der Ausbildung in Jahren

*STUND* durchschnittliche wöchentliche Arbeitszeit

*PSAMPLE* Stichprobenart (1: Deutsche (West), 2: Ausländer (West),  
3: Deutsche (Ost), 4: Zuwanderer 1984-93, 5: Ergänzung 1998,  
6: Ergänzung 2000, 7: Hocheinkommensbezieher 2002)

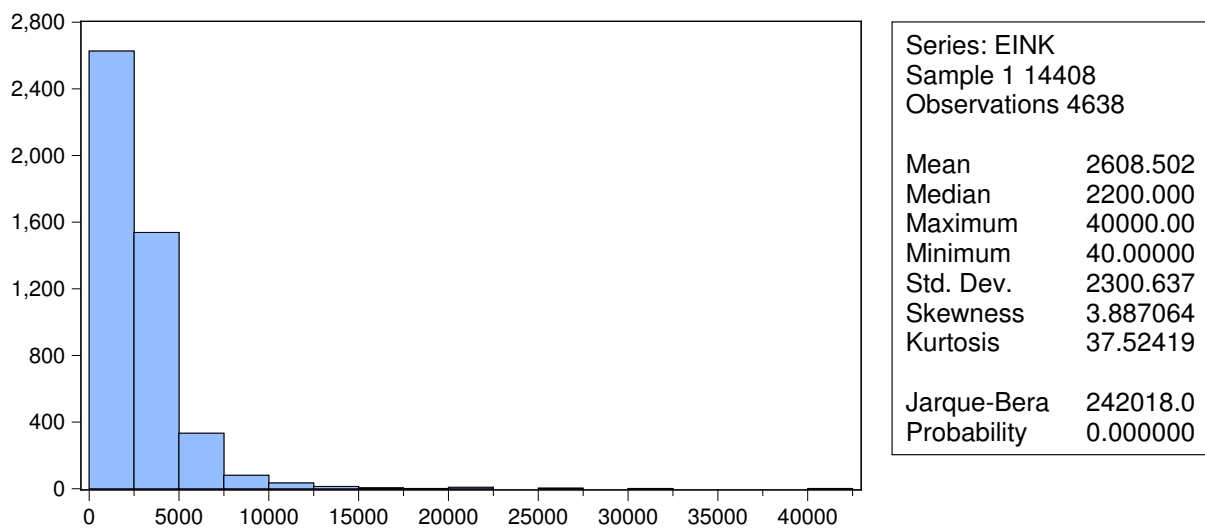
Quelle: Ausgewählte Daten des SOEP für das Jahr 2010.

## 5.5 Schätzung von Einkommensfunktionen

### Das additive Modell

Im einfachen additiven empirischen Modell wird die endogene Variable *Einkommen* durch verschiedene exogene Variablen erklärt:

$$Eink = \beta_0 + \beta_1 \cdot Schule + \beta_2 \cdot Beruf + \beta_3 \cdot Berufserfahrung + \beta_4 \cdot Geschlecht + \varepsilon .$$



- Im Beispiel liegen 4.638 Beobachtungen vor.
- Der Wert des arithmetischen Mittels ist 2.608,502 €.
- Die Verteilung ist links-steil bzw. rechts-schief, der Median ist kleiner als das arithmetische Mittel.
- Die Standardabweichung beträgt 2.300,637 €.



In einer ersten Schätzung wird das Einkommen durch den Schulabschluss erklärt:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/27/12   Time: 14:11
Sample: 1 14408 IF PSAMPLE<=3
Included observations: 1804
=====
Variable      Coefficient      Std. Error      t-Statistic      Prob.
=====
C              1827.593         206.6004         8.846026         0.0000
HAUPT          340.6690         225.1376         1.513159         0.1304
REAL           375.2235         217.5420         1.724833         0.0847
FACH           1295.541         271.2069         4.776948         0.0000
ABI            1569.361         225.0556         6.973215         0.0000
=====
R-squared          0.080119      Mean dependent var      2522.113
Adjusted R-squared 0.078074      S.D. dependent var      1936.536
S.E. of regression 1859.404      Akaike info criterion   17.89667
Sum squared resid  6.22E+09      Schwarz criterion       17.91190
Log likelihood     -16137.79     Hannan-Quinn criter.   17.90229
F-statistic        39.17220     Durbin-Watson stat     1.751058
Prob(F-statistic)  0.000000
=====

```

- HAUPT, REAL, FACH und ABI sind Dummyvariablen für einen Hauptschulabschluss, Realschulabschluss, Fachabitur bzw. Abitur.
- Personen ohne Schulabschluss werden nicht als Dummyvariable aufgenommen, da nur  $i-1$  Dummyvariablen bei  $i$  Ausprägungen modelliert werden dürfen. Personen ohne Schulabschluss haben ein Einkommen von 1.827,59 € (Koeffizient der Konstanten C).
- Personen mit Abitur haben ein Einkommen von  $1.827,59 + 1.569,36 = 3.396,95$  €.
- Der Erklärungsgehalt des Modells ist mit ca. 8% gering.

### Prüfung der Schätzergebnisse:

- erwartetes Vorzeichen?
- Größenordnung?
- ökonomische Plausibilität?

Die Erklärung des Einkommens durch den Berufsabschluss liefert folgendes Ergebnis:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/27/12   Time: 14:24
Sample: 1 14408 IF PSAMPLE<=3
Included observations: 1780
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1957.118	75.39210	25.95919	0.0000
LEHRE	246.0622	99.48171	2.473442	0.0135
MEIST	1104.799	170.6130	6.475466	0.0000
FH	1569.811	173.8513	9.029618	0.0000
UNI	2187.075	135.4859	16.14246	0.0000

```

=====
R-squared          0.161162   Mean dependent var    2551.435
Adjusted R-squared 0.159271   S.D. dependent var    1917.774
S.E. of regression 1758.431   Akaike info criterion 17.78504
Sum squared resid  5.49E+09   Schwarz criterion     17.80044
Log likelihood     -15823.68   Hannan-Quinn criter. 17.79073
F-statistic        85.25532   Durbin-Watson stat   1.779155
Prob(F-statistic)  0.000000
=====

```

- LEHRE, MEIST, FH und UNI sind Dummyvariablen für Lehre, Meister, Fachhochschul- bzw. Universitätsabschluss.
- Personen ohne Berufsabschluss haben ein Einkommen von 1.957,12 €.
- Personen mit Universitätsabschluss erhalten ein Einkommen in Höhe von  $1.957,12 + 2.187,08 = 4.144,20$  €.
- Alle Variablen sind signifikant.
- Der Erklärungsgehalt des Modells ist deutlich gestiegen und beträgt nun ca. 16%.

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

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/27/12   Time: 14:25
Sample: 1 14408 IF PSAMPLE<=3
Included observations: 1733
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1692.834	203.9876	8.298712	0.0000
HAUPT	289.3455	220.8610	1.310080	0.1903
REAL	232.0988	213.9859	1.084645	0.2782
FACH	766.6344	269.2892	2.846881	0.0045
ABI	522.8806	242.7379	2.154095	0.0314
LEHRE	240.4423	102.5196	2.345330	0.0191
MEIST	1057.697	174.0880	6.075642	0.0000
FH	1354.660	190.9907	7.092807	0.0000
UNI	1978.991	180.7791	10.94701	0.0000

```

=====
R-squared          0.168191   Mean dependent var    2579.868
Adjusted R-squared 0.164331   S.D. dependent var    1930.982
S.E. of regression 1765.206   Akaike info criterion 17.79510
Sum squared resid  5.37E+09   Schwarz criterion     17.82345
Log likelihood     -15410.46   Hannan-Quinn criter. 17.80558
F-statistic        43.57381   Durbin-Watson stat   1.805380
Prob(F-statistic)  0.000000
=====

```

- Personen ohne Schul- oder Berufsabschluss haben ein Einkommen von 1.692,83 € (der Koeffizient der Konstanten C).
- Personen mit Abitur und Universitätsabschluss haben ein Einkommen von  $1.692,83 + 522,88 + 1.978,99 = 4.194,70$  €.
- Die Variablen FACH, ABI, LEHRE, MEIST, FH und UNI sind signifikant, während die Koeffizienten HAUPT und REAL insignifikant sind.
- Der Erklärungsgehalt des Modells steigt im Vergleich zum vorherigen Modell nur leicht auf ca. 16,8%.

Zur Erklärung des Einkommens kommen neben Schul- und Berufsabschluss weitere Variablen in Betracht:

```

=====
Dependent Variable: EINK
Method: Least Squares
Date: 06/27/12   Time: 14:26
Sample: 1 14408 IF PSAMPLE<=3
Included observations: 1731
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1939.733	227.9484	8.509527	0.0000
HAUPT	162.4705	207.5821	0.782681	0.4339
REAL	424.2284	201.7652	2.102584	0.0356
FACH	940.4349	254.9638	3.688504	0.0002
ABI	820.9039	232.1858	3.535547	0.0004
LEHRE	138.7159	96.45940	1.438075	0.1506
MEIST	694.6947	164.2624	4.229176	0.0000
FH	1112.115	179.2444	6.204461	0.0000
UNI	1650.793	170.6837	9.671648	0.0000
SEX	-1253.523	81.21309	-15.43499	0.0000
EX	14.20143	3.877991	3.662058	0.0003

```

=====
R-squared          0.274204   Mean dependent var    2582.502
Adjusted R-squared 0.269984   S.D. dependent var    1930.528
S.E. of regression 1649.462   Akaike info criterion 17.66062
Sum squared resid  4.68E+09   Schwarz criterion     17.69529
Log likelihood     -15274.27   Hannan-Quinn criter. 17.67344
F-statistic        64.98121   Durbin-Watson stat    1.767784
Prob(F-statistic)  0.000000
=====

```

- SEX ist eine Dummyvariable, die den Wert 1 für Frauen und den Wert 0 bei Männern annimmt. Frauen verdienen bei sonst gleichen Voraussetzungen 1.253,52 € weniger als Männer. Eine Aussage darüber, ob geringeres Humankapital (Babypause) oder Diskriminierung der Grund für den geringeren Verdienst ist, kann aufgrund dieser Schätzung nicht gemacht werden.
- EX ist eine Variable für die Berufserfahrung: Ein Jahr mehr Berufserfahrung erhöht den Verdienst um 14,20 €.
- Der Erklärungsgehalt des Modells steigt im Vergleich zum vorherigen Modell auf ca. 27%.

## Das multiplikative Modell

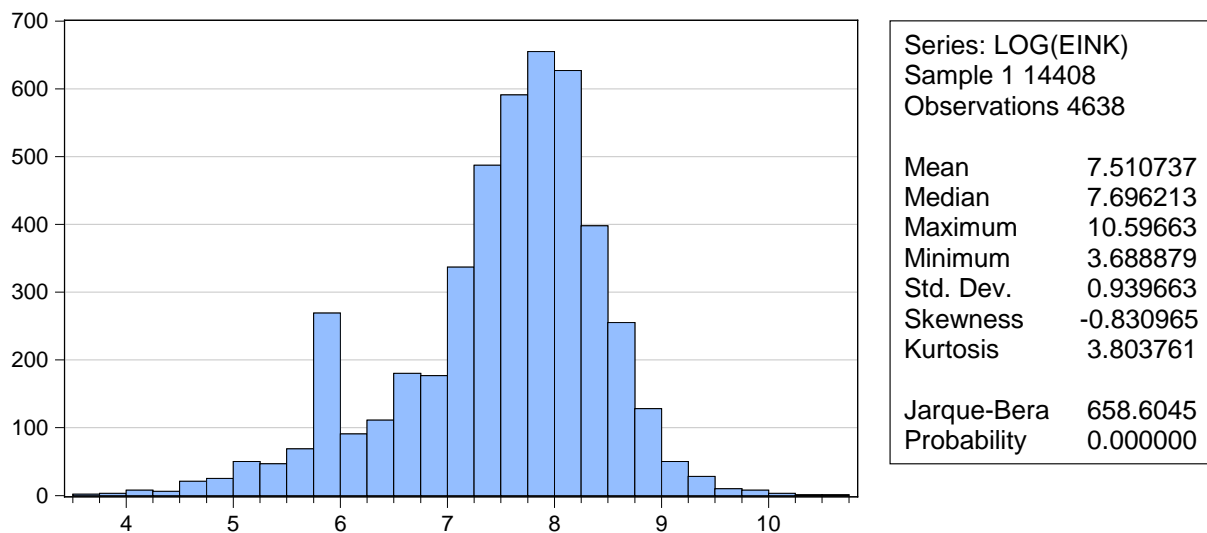
Ein multiplikatives Modell ist aus theoretischer Sicht besser geeignet. Siehe dazu auch die Herleitung im Lehrbuch von Franz, W. (2003), Arbeitsmarktökonomik, S. 90 ff.

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

bzw.

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

Die Koeffizienten  $\beta_1$  bis  $\beta_3$  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 Vervierfachung des Einkommens ( $e^{1,4} \approx 4$ ).



- Das Histogramm von  $\log(\text{Eink})$  ähnelt nunmehr eher einer Normalverteilung als das Histogramm des Einkommens selbst.
- Das arithmetische Mittel hat den Wert 7,510.
- Das arithmetische Mittel des Einkommens beträgt  $e^{7,510} = 1.826,21$  €.

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

```

=====
Dependent Variable: LOG(EINK)
Method: Least Squares
Date: 06/27/12   Time: 14:29
Sample (adjusted): 1 14403
Included observations: 4220 after adjustments
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.013625	0.052608	133.3193	0.0000
HAUPT	0.245961	0.058880	4.177325	0.0000
REAL	0.313083	0.055704	5.620438	0.0000
FACH	0.337522	0.072052	4.684440	0.0000
ABI	0.401103	0.061222	6.551600	0.0000
LEHRE	0.129079	0.033202	3.887739	0.0001
MEIST	0.519459	0.056292	9.227989	0.0000
FH	0.588662	0.054022	10.89666	0.0000
UNI	0.749238	0.051528	14.54052	0.0000

```

=====
R-squared          0.130772   Mean dependent var    7.579903
Adjusted R-squared 0.129120   S.D. dependent var    0.905081
S.E. of regression 0.844630   Akaike info criterion 2.502294
Sum squared resid  3004.127   Schwarz criterion     2.515832
Log likelihood      -5270.841   Hannan-Quinn criter. 2.507080
F-statistic         79.19091   Durbin-Watson stat    1.918824
Prob(F-statistic)  0.000000
=====

```

- Personen, die weder einen Schul- noch einen Berufsabschluss haben verdienen  $e^{7,014} = 1.112,09 \text{ €}$ .
- Personen, die einen Hauptschulabschluss, aber keinen Berufsschulabschluss haben verdienen  $e^{7,014+0,246} = e^{7,26} = 1.422,26 \text{ €}$ .
- Anders aufgeschrieben verdienen diese Personen  $e^{7,014} \cdot e^{0,246} = 1.112,09 \cdot 1,279 = 1.422,26 \text{ €}$ .  
Sie haben also ein 27,89% höheres Einkommen.
- Personen mit Abitur und Universitätsabschluß verdienen  $e^{7,014+0,401+0,749} = e^{8,107} = 3.512,21 \text{ €}$ .
- Anders aufgeschrieben verdienen diese Personen  $e^{7,014} \cdot e^{0,401} \cdot e^{0,749} = 1.112,09 \cdot 1,493 \cdot 2,115 = 3.512,21 \text{ €}$ .  
Sie haben ein  $1,493 \cdot 2,115 = 3,157$  mal höheres Einkommen.

## 5.6 Test auf Strukturkonstanz

Unter den bisherigen Schätzungen wurde immer von der Annahme ausgegangen, daß die Parameter des "wahren" ökonomischen Modells  $\beta$  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.

### Wald-Test

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: LOG(EINK)

Method: Least Squares

Date: 06/27/12 Time: 15:49

Sample: 1 14408 IF PSAMPLE<=3 AND (OST=0 OR OST=1)

Included observations: 1802

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.522491	0.087589	74.46699	0.0000
DAUER	0.104286	0.006707	15.54773	0.0000
SEX	-0.611940	0.035712	-17.13559	0.0000

R-squared	0.227088	Mean dependent var	7.534937
Adjusted R-squared	0.226229	S.D. dependent var	0.861231
S.E. of regression	0.757575	Akaike info criterion	2.284276
Sum squared resid	1032.483	Schwarz criterion	2.293427
Log likelihood	-2055.133	Hannan-Quinn criter.	2.287654
F-statistic	264.2813	Durbin-Watson stat	1.826400
Prob(F-statistic)	0.000000		

### Einkommensgleichung für Ost (ohne Berlin)

===== (2)

Dependent Variable: LOG(EINK)

Method: Least Squares

Date: 06/27/12 Time: 15:51

Sample: 1 14408 IF PSAMPLE<=3 AND OST=1

Included observations: 568

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.044728	0.163175	37.04435	0.0000
DAUER	0.116158	0.011992	9.686175	0.0000
SEX	-0.331501	0.062396	-5.312821	0.0000

R-squared	0.173096	Mean dependent var	7.407815
Adjusted R-squared	0.170169	S.D. dependent var	0.815440
S.E. of regression	0.742825	Akaike info criterion	2.248556
Sum squared resid	311.7609	Schwarz criterion	2.271489
Log likelihood	-635.5898	Hannan-Quinn criter.	2.257505
F-statistic	59.13579	Durbin-Watson stat	1.906784
Prob(F-statistic)	0.000000		



Einkommensgleichung für West (ohne Berlin)

===== (3)

Dependent Variable: LOG(EINK)  
 Method: Least Squares  
 Date: 06/27/12 Time: 15:53  
 Sample: 1 14408 IF PSAMPLE<=3 AND OST=0  
 Included observations: 1234

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.604182	0.102599	64.36868	0.0000
DAUER	0.108862	0.008013	13.58544	0.0000
SEX	-0.730009	0.042317	-17.25097	0.0000

R-squared	0.282769	Mean dependent var	7.593450
Adjusted R-squared	0.281604	S.D. dependent var	0.875646
S.E. of regression	0.742182	Akaike info criterion	2.243984
Sum squared resid	678.0772	Schwarz criterion	2.256427
Log likelihood	-1381.538	Hannan-Quinn criter.	2.248665
F-statistic	242.6618	Durbin-Watson stat	1.891567
Prob(F-statistic)	0.000000		

$$F = \frac{(1032,483 - (311,760 + 678.077))/3}{(311,760 + 678.077)/(1234 + 568 - 6)} = 25,79$$

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

Neben dem Wald-Test besteht eine weitere Möglichkeit Indizien über einen potenziellen Strukturbruch zu erhalten. Die Aufnahme der entsprechenden Variable in die Schätzgleichung (über das gesamte Beobachtungssample!) erlaubt Rückschlüsse über mögliche strukturelle Unterschiede. In unserem Fall wird die Ost-Variable als weiterer exogener Term betrachtet.

```

=====
Dependent Variable: LOG(EINK)
Method: Least Squares
Date: 07/03/14   Time: 12:05
Sample: 1 14408 IF PSAMPLE<=3 AND (OST=1 OR OST=0)
Included observations: 1802
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.507826	0.086485	75.24774	0.0000
DAUER	0.111851	0.006710	16.66966	0.0000
SEX	-0.603845	0.035270	-17.12046	0.0000
OST	-0.267237	0.038449	-6.950394	0.0000

```

=====
R-squared          0.247311   Mean dependent var    7.534937
Adjusted R-squared 0.246056   S.D. dependent var    0.861231
S.E. of regression 0.747807   Akaike info criterion 2.258873
Sum squared resid  1005.468   Schwarz criterion     2.271074
Log likelihood      -2031.245   Hannan-Quinn criter. 2.263377
F-statistic         196.9234   Durbin-Watson stat    1.855680
Prob(F-statistic)   0.000000
=====

```

Die Variable *Ost* übt einen signifikanten negativen Einfluss auf die endogene Variable  $\log(Eink)$  aus. Dies deutet auf einen Unterschied in der Datenstruktur der beiden Regionen hin.

Ausgehend von Einkommensgleichung (3) können in der Schätzung weitere Variablen betrachtet werden.

```

=====
Dependent Variable: LOG(EINK)
Method: Least Squares
Date: 07/03/14   Time: 12:56
Sample: 1 14408 IF PSAMPLE<=3 AND OST=0
Included observations: 1208
=====

```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.719825	0.105915	44.56224	0.0000
DAUER	0.091670	0.006165	14.86936	0.0000
SEX	-0.187574	0.036414	-5.151100	0.0000
STUND	0.042881	0.001406	30.50652	0.0000
EX	0.010673	0.001431	7.460398	0.0000

```

=====
R-squared          0.607578   Mean dependent var    7.603944
Adjusted R-squared 0.606273   S.D. dependent var   0.873444
S.E. of regression 0.548066   Akaike info criterion 1.639288
Sum squared resid  361.3527   Schwarz criterion     1.660384
Log likelihood      -985.1301   Hannan-Quinn criter. 1.647232
F-statistic         465.6436   Durbin-Watson stat   1.925122
Prob(F-statistic)  0.000000
=====

```

Sowohl die wöchentliche *Arbeitszeit*, als auch die *Berufserfahrung* beeinflussen das Einkommen positiv und signifikant.