Aufgaben zur Ökonometrie I

7. Modelle mit qualitativen Variablen

7.1 Die Produktion y einer Industriebranche lässt sich in einem Quartal auf die um eine Periode verzögerten Auftragseingänge x sowie additiver saisonaler Einflüsse zurückführen. Geben Sie das zugehörige Design der Beobachtungsmatrix **X** für einen Zeitraum von drei Jahren wieder!

Lösung:

• Auftragseingänge in den Quartalen im Analysezeitraum: $x_1, x_2, ..., x_{n-1}, x_n$

Auftragseingänge vor Beginn des Analysezeitraums: x_0

• Die Saisonvariable wird bei 4 Quartalen durch 3 Saisondummies erfasst

Wählt man z.B. das 1. Quartal als Referenzquartal, werden Saisondummies für die Quartale 2, 3 und 4 gebildet:

$$s_j = \begin{cases} 1, \text{ falls Quartal } j \\ 0, \text{ sonst} \end{cases}, \quad j=2,3,4$$

Design der Beobachtungsmatrix **X**: $$\underset{12x5}{\mathbf{X}} = \begin{bmatrix} 1 & x_0 & 0 & 0 & 0 \\ 1 & x_1 & 1 & 0 & 0 \\ 1 & x_2 & 0 & 1 & 0 \\ 1 & x_3 & 0 & 0 & 1 \\ 1 & x_4 & 0 & 0 & 0 \\ 1 & x_5 & 1 & 0 & 0 \\ 1 & x_6 & 0 & 1 & 0 \\ 1 & x_7 & 0 & 0 & 1 \\ 1 & x_8 & 0 & 0 & 0 \\ 1 & x_9 & 1 & 0 & 0 \\ 1 & x_{10} & 0 & 1 & 0 \\ 1 & x_{11} & 0 & 0 & 1 \end{bmatrix}$$

7.2 Für die Keynes'sche Konsumfunktion ergaben sich mit EViews folgende Schätzergebnisse für die Gesamtperiode 1974 -1992,

Dependent Variable: KONSUMR
Method: Least Squares Sample: 1974 1992

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	37.31676	44.21441	0.843995	0.4104
VEINKR	0.864223	0.037617	22.97439	0.0000

R-squared	0.968797	S.D. dependent var	128.5075
Adjusted R-squared	0.966962	S.E. of regression	23.35807

sowie vor

Dependent Variable: KONSUMR
Method: Least Squares Sample: 1974 1979

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-118.7550	56.85496	-2.088736	0.1050
VEINKR	0.997260	0.054577	18.27246	0.0001

R-squared	0.988162	S.D. dependent var	63.73863
Adjusted R-squared	0.985202	S.E. of regression	7.753625

und nach dem Ölpreisschock 1980:

Dependent Variable: KONSUMR
Method: Least Squares Sample: 1980 1992

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	167.1905	47.73910	3.502170	0.0050
VEINKR	0.764885	0.038743	19.74249	0.0000

R-squared	0.972553	S.D. dependent var	106.1633
Adjusted R-squared	0.970057	S.E. of regression	18.37044

Überprüfen Sie mit Hilfe des Chow-Tests, ob der Ölpreisschock 1980 einen Strukturbruch in der Keynes'schen Konsumfunktion verursacht haben könnte ($\alpha = 0{,}01$)!

Lösung:

•Nullhypothese des Chow-Tests:

$$H_0 : \beta_j^1 = \beta_j^2 \quad \text{für alle } j = 1{,}2$$

Unter der Nullhypothese liegt kein Strukturbruch vor, so dass der autonome Konsum β_1 und die marginale Konsumquote β_2 vor und nach dem Ölpreisschock gleich sind.

• Berechnung der Residualquadratsummen

- Residualquadratsumme Q für den ges. Beobachtungszeitraum 1974 – 1992 (n=19, k=2):

$$Q = SSR = (n-k)\cdot SER^2 = (19-2)\cdot 23{,}35807^2 = 9275{,}2$$

- Residualquadratsumme Q_1 für Teilzeitraum 1974 – 1979 (n=6, k=2):

$$Q_1 = SSR_1 = (n-k)\cdot SER_1^2 = (6-2)\cdot 7{,}753625^2 = 240{,}5$$

- Residualquadratsumme Q_2 für Teilzeitraum 1980 – 1992 (n=13, k=2):

$$Q_2 = SSR_2 = (n-k)\cdot SER_2^2 = (13-2)\cdot 18{,}37044^2 = 3712{,}2$$

• Prüfgröße des Chow-Tests (n=19, k=2):

$$F = \frac{[Q-(Q_1+Q_2)]/k}{(Q_1+Q_2)/(n-2k)} = \frac{[9275{,}2-(240{,}5+3712{,}2)]/2}{(240{,}5+3712{,}2)/(19-2\cdot 2)}$$

$$= \frac{(9275{,}2-3952{,}7)/2}{3952{,}7/15} = \frac{2661{,}25}{263{,}513} = 10{,}099$$

• Kritischer Wert ($\alpha = 0{,}01$):

$F_{2;15;0,99} = 6{,}36$

• Testentscheidung:

$(F = 10{,}099) > (F_{2;15;0,99} = 6{,}36)$ → H_0 ablehnen

Nach dem Ölpreisschock können der autonome Konsum β_1 und die marginale Konsumquote β_2 nicht mehr als gleich angesehen werden, so dass auf einen Strukturbruch in der Keynes'schen Konsumfunktion geschlossen werden muss.

7.3 Interpretieren Sie das EViews-Ergebnis einer OLS-Schätzung der Keynes'schen Konsumfunktion unter Verwendung der Dummy-Variablen D_1 (1 für 1974 bis 1979), 0 sonst und D_2 (1 für 1980 bis 1992, 0 sonst):

Dependent Variable: KONSUMR
Method: Least Squares Sample: 1974 1992

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	-118.7550	119.0320	-0.997673	0.3343
D2	167.1905	42.18468	3.963298	0.0012
VEINKR*D1	0.997260	0.114263	8.727736	0.0000
VEINKR*D2	0.764885	0.034235	22.34197	0.0000

Lösung:

• Autonomer Konsum

Die Regressionskoeffizienten der beiden Dummy-Variablen D_1 und D_2 geben den autonomen Konsum vor und nach dem Ölpreisschock 1980 an. Rein rechnerisch liegt der autonome Konsum nach dem Ölpreisschock um 285,9455 GE [= 167,1905 – (–118,7550) GE] höher als nach dem Ölpreisschock. Hierbei ist jedoch zu berücksichtigen, dass der Regressionskoeffizient der Dummy-Variablen D_1 nicht signifikant von 0 abweicht (p = 0,3343). Bei einem „wahren" absoluten Glied vor dem Ölpreisschock von 0 steigt der autonome Konsum nach dem Ölpreisschock um 167,1905 GE an.

• Marginale Konsumquote

Die marginalen Konsumquoten sind in den beiden Teilperioden durch die Regressionskoeffizienten der Interaktionsterme VEINR*D_1 und VEINKR*D_2 gegeben. Hier zeigt sich genau eine zum autonomen Konsum entgegengerichtete Tendenz. Die marginale Konsumquote ist in dem Teilzeitraum 1974 – 1979 bei einem Wert von 0,997 größer als in dem zweiten Teilzeitraum 1980 – 1992, in dem sie nur noch bei 0,765 liegt. Sie sinkt somit nach dem Ölpreisschock um 23,2 Prozentpunkte [=(0,997 – 0,765)·100%] ab.

7.4 Welche Einwände lassen sich gegen das lineare Wahrscheinlichkeitsmodell bei qualitativen abhängigen Variablen aus ökonometrischer Sicht erheben?

Lösung:

Im linearen Wahrscheinlichkeitsmodell lassen sich die Erwartungswerte der abhängigen qualitativen Variablen y als Wahrscheinlichkeiten bei gegebenen Werten der Regressionskoeffizienten β_j und exogenen Variablen x_{jt} interpretieren. Die Regressionswerte $\hat{y}_t$ müssten als Schätzer dieser Wahrscheinlichkeiten zwischen 0 und 1 liegen, damit das Modell sachgerecht interpretiert werden kann. Bei einer Kleisnt-Quadrate-Schätzung kann diese Bedingung jedoch verletzt werden. Die geschätzten Wahrscheinlichkeiten können f+r einzelne Querschnittseinheiten (z.B. Konsumenten, Unternehmen, Beschäftigte) negativ oder größer als eins werden, so dass die Schätzergebnisse des linearen Wahrscheinlichkeitsmodells nicht mehr sinnvoll interpretierbar sind.

Beim multiplen linearen Regressionsmodell wird standardmäßig angenommen, dass die Störgröße homoskedastisch ist, d.h. die gleiche Varianz für alle Einheiten besitzt. Diese Annahme ist beim linearen Wahrscheinlichkeitsmodell stets verletzt, da die Störvarianz von den Werten der exogenen Variablen abhängt. Sie ist damit in der Regel bei allen Querschnittseinheiten unterschiedlich. Der bedingte Erwartungswert setzt im linearen Wahrscheinlichkeitsmodell von vornherein inadäquate Restriktionen bezüglich der Störgröße.

7.5 X_i, i=1,2,…,n ist eine Bernoulli-verteilte Zufallsvariable,die den Wert 1 mit der Wahrscheinlichkeit p und den Wert 0 mit einer Wahrscheinlichkeit 1-p annimmt. Bestimmen Sie für die gegebene Zufallsstichprobe (x_1, x_2, …, x_n) aus einer Bernoulli-verteilten Grundgesamtheit den Maximum-Likelihood-Schätzer (ML-Schätzer) für die Wahrscheinlichkeit p!

Lösung:

Aus der induktiven Statistik ist bekannt, dass die Summe Bernoulli-verteilter Zufallsvariablen binomialverteilt ist.

Likelihood-Funktion:

$$L(p \mid x, n) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad \text{mit} \quad x = \sum_{i=1}^{n} \underset{\substack{\uparrow \\ 0 \text{ od. } 1}}{x_i}$$

Notwendige Bedingung für ein Maximum:

$$\frac{dL}{dp} = \binom{n}{x} \cdot x \cdot p^{x-1} \cdot (1-p)^{n-x} + \binom{n}{x} \cdot p^x \cdot (n-x) \cdot (1-p)^{n-x-1} \cdot (-1) = 0$$

$$x \cdot (1-p) - (n-x) \cdot p = 0$$

$$x - n \cdot p = 0 \Leftrightarrow n \cdot p = x$$

$$\hat{p}_{ML} = \frac{x}{n} \quad \text{(Stichprobenanteilswert)}$$

7.6 Die Wahl eines Verkehrsmittels V (ÖPNV oder Pkw) für den Weg zur Arbeitsstätte ist mit EViews auf der Basis des Logit-Modells in Abhängigkeit von der Entfernung E (in km) und der beruflichen Stellung S mit der Maximum-Likelihood-Methode geschätzt worden (Daten s. Anhang):

Dependent Variable: VERKEHRSART
Method: ML - Binary Logit (Newton-Raphson)
Sample: 1 24 Included observations: 24
Convergence achieved after 6 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-5.005772	2.122852	-2.358041	0.0184
ENTFERNUNG	0.265619	0.134538	1.974306	0.0483
STELLUNG	4.590040	1.957320	2.345063	0.0190

R-squared	0.635631	Mean dependent var	0.416667
S.D. dependent var	0.503610	S.E. of regression	0.280037
Akaike info criterion	0.744953	Sum squared resid	1.646835
Schwarz criterion	0.892210	Log likelihood	-5.939442
Restr. log likelihood	-16.30064	LR statistic	20.72239
Avg. log likelihood	-0.247477	Prob(LR statistic)	3.16E-05

Obs with Dep=0	14	Total obs	24
Obs with Dep=1	10		

a) Geben Sie das numerisch spezifizierte Logit-Modell der Verkehrsmittelwahl

- mit den geschätzten Wahrscheinlichlkeiten $\hat{p}_t$,

- mit den geschätzten log Odds $\ln[\hat{p}_t/(1-\hat{p}_t)]$

als Regressionswerte an!

Lösung:

Abhängige Variable: Verkehrsart (y)

Unabhängige Variablen: Entfernung (x_2), Berufliche Stellung (x_3)

- Logit-Modell mit geschätzten Wahrscheinlichlkeiten $\hat{p}_t$ als Regressionswerte:

$$\hat{p}_t = \frac{e^{-5{,}005772+0{,}265619\cdot x_2+4{,}590040\cdot x_3}}{1+e^{-5{,}005772+0{,}265619\cdot x_2+4{,}590040\cdot x_3}}$$

- Logit-Modell mit geschätzten log Odds $\ln[\hat{p}_t/(1-\hat{p}_t)]$ als Regressionswerte:

$$\ln\left(\frac{\hat{p}_t}{1-\hat{p}_t}\right) = -5{,}005772 + 0{,}265619\cdot x_2 + 4{,}590040\cdot x_3$$

b) Wie lassen sich die Effektkoeffizienten der Variablen Entfernung (x_2) und Berufliche Stellung (x_3) interpretieren?

Lösung:

Effektkoeffizient der Entfernung: $e^{\hat{\beta}2} = e^{0,265619} = 1,304$

Wenn sich die Entfernung zur Arbeitsstätte (x_2) um 1 km erhöht, steigen die Odds um den Faktor 1,304, d.h. es steigt das Chancenverhältnis für die Wahl des Pkw anstelle des ÖPNV als Verkehrsmittel für den Weg zur Arbeitsstätte ceteris paribus um 30,4%.

Effektkoeffizient der Entfernung: $e^{\hat{\beta}3} = e^{4,590040} = 98,498$

Die Odds steigen für einen Beschäftigten mit einer hohen beruflichen Stellung ($x_3 = 1$) um den Faktor 98,498 gegenüber einen Beschäftigten mit einer niedrigen beruflichen Stellung ($x_3 = 0$). Das Chancenverhältnis für die Wahl des Pkw anstelle des ÖPNV steigt damit bei einem Übergang von einem Beschäftigten in einer niedrigen beruflichen Stellung zu einem Beschäftigten mit einer hohen beruflichen Stellung ceteris paribus um 97,498%.

c) Geben Sie den marginalen Effekt der Entfernung zum Arbeitsplatz (x_2) auf die Wahrscheinlichkeit für die Wahl eines Pkws für Arbeitnehmer an, die das Profil

- des Beschäftigten 1,
- des Beschäftigten 2

besitzen?

Lösung:

Marginaler Effekt der Entfernung zum Arbeitsplatz (x_2) auf die Wahrscheinlichkeit p für die Wahl eines Pkws:

$$\frac{\partial \hat{p}_t}{\partial x_{2t}} = \frac{e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{2t} + \hat{\beta}_3 \cdot x_{3t}}}{\left(1 + e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{2t} + \hat{\beta}_3 \cdot x_{3t}}\right)^2} \cdot \hat{\beta}_2$$

- Merkmalsprofil des Beschäftigten 1: $\mathbf{x}'_1 = (1 \;\; 26 \;\; 1)$

$$\frac{\partial \hat{p}_1}{\partial x_{21}} = \frac{e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{21} + \hat{\beta}_3 \cdot x_{31}}}{\left(1 + e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{21} + \hat{\beta}_3 \cdot x_{31}}\right)^2} \cdot \hat{\beta}_2 = \frac{e^{-5{,}00577 + 0{,}26562 \cdot 26 + 4{,}59004 \cdot 1}}{\left(1 + e^{-5{,}00577 + 0{,}26562 \cdot 26 + 4{,}59004 \cdot 1}\right)^2} \cdot 0{,}26562$$

$$= \frac{658{,}78024}{(1 + 658{,}78024)^2} \cdot 0{,}26562 = \frac{658{,}78024}{435309{,}9617} \cdot 0{,}26562 = 0{,}000402$$

- Merkmalsprofil des Beschäftigten 1: $\mathbf{x}'_2 = (0 \;\; 5 \;\; 0)$

$$\frac{\partial \hat{p}_2}{\partial x_{22}} = \frac{e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{22} + \hat{\beta}_3 \cdot x_{32}}}{\left(1 + e^{\hat{\beta}_1 + \hat{\beta}_2 \cdot x_{22} + \hat{\beta}_3 \cdot x_{32}}\right)^2} \cdot \hat{\beta}_2 = \frac{e^{-5{,}00577 + 0{,}26562 \cdot 5 + 4{,}59004 \cdot 0}}{\left(1 + e^{-5{,}00577 + 0{,}26562 \cdot 5 + 4{,}59004 \cdot 0}\right)^2} \cdot 0{,}26562$$

$$= \frac{0{,}02528}{(1 + 0{,}02528)^2} \cdot 0{,}26562 = \frac{0{,}02528}{1{,}05120} \cdot 0{,}26562 = 0{,}00639$$

d) Wie groß ist das in EViews mit „R-squared“ bezeichnete Pseudo-R^2? Wie lässt es sich interpretieren?

Lösung:

$PR^2 = 0.635631$

$L^* = -5.939442$ (s. EViews-Output)

$$L_0^* = \ln\left(\frac{n_1}{n}\right)\cdot n_1 + \ln\left(\frac{n_2}{n}\right)\cdot n_2 = \ln\left(\frac{14}{24}\right)\cdot 14 + \ln\left(\frac{10}{24}\right)\cdot 10 = -0{,}538997\cdot 14 - 0{,}875469\cdot 10$$
$$= -7{,}545958 - 8{,}75469 = -16{,}300648$$

Pseudo-R^2: $PR^2 = 1 - \frac{L^*}{L_0^*} = 1 - \frac{-5{,}939442}{-16{,}300648} = 1 - 0{,}364368 = 0{,}636$

Das PseudoR^2 von 0,636 gibt an, dass der Wert der Likelihood-Funktion um 63,6% reduziert werden kann, wenn man die Wahrscheinlichkeit für die Wahl des Pkw als Verkehrsmittel zur Fahrt zum Abeitsplatz unter Verwendung der erklärenden Variablen mit dem Logit-Modell anstelle der bloßen Häufigkeit von 10/24 = 0,417 schätzt.

Anhang: Daten zur Wahl des Verkehrsmittels (Aufg. 7.6)

Beschäf-tigter (t)	Art des Ver-kehrsmittel (y)	Entfernung (x_1)	Berufliche Stellung (x_2)
1	1	26	1
2	0	5	0
3	1	17	1
4	0	1	1
5	0	14	0
6	1	28	1
7	0	19	0
8	0	2	0
9	1	22	0
10	0	8	0
11	0	6	0
12	1	18	1
13	0	4	0
14	0	12	0
15	0	3	0
16	1	15	1
17	1	8	0
18	0	10	0
19	0	2	0
20	1	10	1
21	1	16	1
22	0	11	0
23	0	4	0
24	1	5	1

Art des Verkehrsmittel (y): 0 ÖPNV, 1 Pkw

Berufliche Stellung (x_2): 0 niedrige berufliche Stellung, 1 hohe berufliche Stellung