

Masterprüfung

Fach: Mikroökonomie

Prüfer: Prof. Regina T. Riphahn, Ph.D.

Vorbemerkungen:

- Anzahl der Aufgaben:** Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen.
- Bewertung:** Es können maximal 60 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
- Erlaubte Hilfsmittel:** Tabellen der statistischen Verteilung (sind der Klausur beigelegt)
Taschenrechner
Fremdwörterbuch
- Wichtige Hinweise:** Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten exakten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1 (12,5 Punkte)

Für das Jahr 2015 werden mit einem Querschnittsdatensatz die Determinanten von Urlaubsausgaben mit einem Regressionsmodell geschätzt. 139 der 714 in der Stichprobe enthaltenen Individuen haben 2015 keine Urlaubsausgaben getätigt (d.h. $urlaub=0$). Die Variablen des Datensatzes sind wie folgt definiert:

- $urlaub$ = Höhe der Urlaubsausgaben für das Jahr 2015 in €
- $alter$ = Alter in Jahren
- $ek1$ = 1, falls monatliches Einkommen höchstens 1499 €; =0 sonst
- $ek2$ = 1, falls monatliches Einkommen zwischen 1500 und 2999 €; =0 sonst
- $ek3$ = 1, falls monatliches Einkommen mindestens 3000 €; =0 sonst

Ein Tobit-Modell liefert folgende Ergebnisse:

Tobit regression		Number of obs = 714			
Log likelihood = -745.947		LR chi2(?) = ???	Prob > chi2 = ???		
		Pseudo R2 = 0.7087			
urlaub	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
alter	77.85490	23.80924	3.27	0.001	17.40015 124.5218
ek1	-370.676	1.046733	-359.9	0.000	-378.815 -374.711
ek2	-129.863	23.42910	-5.43	0.000	-175.785 -85.2105
_cons	6.979211	7.60e-03	918.1	0.000	6.964301 6.994105
/sigma	.9952454	.0110585			.9735666 1.016924

- 1.1 Erläutern Sie knapp am Beispiel der Variable $urlaub$ den Unterschied zwischen Stützung und Zensierung. Welcher Fall liegt in der Aufgabenstellung vor? Begründen Sie Ihre Antwort. (2 Punkte)
- 1.2 Überprüfen Sie mit einem Wald-Test am 1%-Signifikanzniveau die Hypothese, dass sich die Höhe der Urlaubsausgaben in mindestens einer der niedrigeren Einkommensklassen (d.h. Einkommen höchstens 1499 € und/oder Einkommen zwischen 1500 und 2999 €) von der in der höchsten Einkommensklasse (d.h. Einkommen mindestens 3000 €) unterscheiden. Geben Sie Null- und Alternativhypothese, Freiheitsgrade, kritischen Wert, Teststatistik und Testentscheidung an. (5,5 Punkte)

Hinweise:

(1) Die Teststatistik des Wald-Tests lautet: $W = \hat{\beta}' \widehat{Var}(\hat{\beta})^{-1} \hat{\beta} \sim \chi_k^2$

(2) Verwenden Sie folgende geschätzte inverse Kovarianzmatrix des Schätzvektors $\hat{\beta} = (\hat{\beta}_{ek1} \hat{\beta}_{ek2})'$:

$$\widehat{Var}(\hat{\beta})^{-1} = \begin{pmatrix} 0,20 & 0 \\ 0 & -1,62 \end{pmatrix}$$

- 1.3 Führen Sie einen Likelihood-Ratio-Test auf Gesamtsignifikanz auf dem 10%-Signifikanzniveau durch. Geben Sie Null- und Alternativhypothese, Freiheitsgrade, kritischen Wert, Teststatistik und Testentscheidung an. *Hinweis:* Der Log-Likelihood-Wert eines Modells, welches nur eine Konstante enthält, beträgt -834,135. (4 Punkte)

- 1.4 Erklären Sie eine Schwäche des Tobit-Schätzers. (1 Punkt)

Aufgabe 2 (18 Punkte)

Als Umweltbeauftragte(r) der Stadt Nürnberg möchten Sie die Determinanten der Fahrradnutzung auf dem Arbeitsweg analysieren und befragen hierzu 320 vollzeitbeschäftigte Personen zwischen 25 und 65 Jahren. Die Variablen sind wie folgt definiert:

- rad* = Fahrradnutzung (1=gering, 2=mittel, 3=hoch)
- entfernung* = Entfernung zwischen Wohnort und Arbeitsstätte in km
- entfernung2* = Variable *entfernung* zum Quadrat
- bildung* = Bildung in Jahren
- kinder* = 1, falls mindestens ein Kind im Haushalt; =0 sonst

Die Schätzung eines ordered Probit-Modells liefert folgende Ergebnisse:

```
Iteration 0:  log likelihood = -5387.704

Ordered probit regression          Number of obs   =       320
                                   LR chi2(4)         =    1661.07
                                   Prob > chi2         =       0.0000
Log likelihood = -3155.410        Pseudo R2       =       ??????
```

rad	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
entfernung	.0812574	.0131231	6.19	0.000	.0555366	.1069783
entfernung2	-.0266431	.000761	-35.01	0.000	-.031134	-.028151
bildung	.0278783	.0035772	7.84	0.000	.025159	.0295975
kinder	-.055498	.011292	-4.91	0.000	-.07763	-.033366
/cut1	-1.45667	0.224305			-1.54319	-1.51835
/cut2	.0864140	0.124921			-.187535	0.85293

- 2.1 Erläutern Sie formal den Zusammenhang zwischen der beobachteten Fahrradnutzung y_i und der latenten abhängigen Variable y_i^* . (2 Punkte)
- 2.2 Leiten Sie die Wahrscheinlichkeiten für die Fahrradnutzungskategorien, $P(y_i = j)$ für $j=2$, in einem geordneten Probit-Modell her. Nutzen Sie dabei die folgende Form des latenten Modells:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

Wie spiegelt sich die Verteilungsannahme für ε_i in Ihrer Darstellung? Erläutern Sie anschließend den Unterschied zum geordneten Logit-Modell. (4 Punkte)

- 2.3 Stellen Sie die Likelihood-Funktion unter Berücksichtigung der Standardnormalverteilung für das geordnete Probit-Modell auf. (3 Punkte)
- 2.4 Bestimmen Sie diejenige Entfernung zwischen Wohnort und Arbeitsstätte in Metern, welche die latente Fahrradnutzung maximiert. (2 Punkte)
- 2.5 Berechnen Sie die kompensierende Bildungsvariation für eine Person mit Kindern im Haushalt. Interpretieren Sie das Ergebnis. (3 Punkte)
- 2.6 Definieren und erläutern Sie das McFadden R^2 und definieren Sie seine Bestandteile. Berechnen und interpretieren Sie das Maß für die vorliegende Schätzung. (4 Punkte)

Aufgabe 3 (15,5 Punkte)

Sie wollen die Determinanten der Ehedauer untersuchen. Ihnen liegen dazu folgende Daten zu 601 Personen vor:

- y_{mar}* = Anzahl der Jahre, die eine Person verheiratet ist
- married* = 1 wenn Person verheiratet, = 0 sonst
- rating* = Beurteilung der Ehe auf einer Skala von 1 = sehr schlecht bis 5 = sehr gut
- affairs* = Anzahl der außerehelichen Affären im letzten Jahr
- relig* = Religiosität von 1= nicht religiös bis 5 = sehr religiös
- children* = Anzahl der Kinder

Sie stellen folgendes Modell auf:

$$ymar_i = \beta_1 + \beta_2 children_i + \beta_3 affairs_i + \beta_4 rating_i + \varepsilon_i$$

4.1 Erläutern Sie in Bezug auf die Analyse der Ehedauer (*y_{mar}*) das Problem endogener Stichprobenselektion. Welche Auswirkung hätte dies auf eine KQ-Schätzung? (3 Punkte)

Sie schätzen die Determinanten der Ehedauer mit einem Heckman-Selektionsmodell und erhalten folgenden Output:

```

Heckman selection model -- two-step estimates      Number of obs      =      601
(regression model with sample selection)         Censored obs       =      167
                                                Uncensored obs     =      434

                                                Wald chi2(3)       =      222.22
                                                Prob > chi2        =      0.0000
    
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

ymar						
children	6.681246	.5084635	13.14	0.000	5.684676	7.677816
affairs	.1970781	.0752819	2.62	0.009	.0495283	.344628
rating	-.4778142	.2194	-2.18	0.029	-.9078303	-.047798
_cons	6.322939	1.108401	5.70	0.000	4.150513	8.495364

married						
relig	.6662088	.0621876	10.71	0.000	.5443233	.7880943
children	-.0112215	.1359819	-0.08	0.934	-.2777411	.2552981
affairs	.0034766	.0183915	0.19	0.850	-.0325701	.0395232
rating	.0880271	.0571779	1.54	0.124	-.0240396	.2000938
_cons	-1.63591	.3174654	-5.15	0.000	-2.258131	-1.013689

mills						
lambda	-2.534384	.6749309	-3.76	0.000	-3.857224	-1.211544

rho	-0.54315					
sigma	4.6660573					

4.2 Nennen Sie eine zentrale Annahme des Heckman-Schätzverfahrens. Welche Folge hat eine Annahmeverletzung? (1,5 Punkte)

4.3 Anhand welcher Statistik im oben stehenden Output kann die Selektivität der Stichprobe beurteilt werden? Was misst die Statistik? Interpretieren Sie den gegebenen Wert. (3 Punkte)

Sie erwägen, obige Fragestellung mit einer Verweildaueranalyse zu untersuchen.

- 4.4 Erläutern Sie anhand des Beispiels der Ehedauer, was unter rechtszensierten Episoden zu verstehen ist. (1 Punkt)
- 4.5 Erläutern Sie allgemein und in Bezugnahme auf das obige Beispiel die Begriffe der negativen und positiven Duration Dependence. (2 Punkte)
- 4.6 Sind die Ihnen vorliegenden Daten geeignet, um einen Kaplan-Meier-Schätzer zu ermitteln? Erläutern Sie kurz. (3 Punkte)
- 4.7 Erläutern Sie am Beispiel der Ehedauer, wie sich die abhängigen Variablen von diskreten und kontinuierlichen Verweildauermodellen unterscheiden. (2 Punkte)

Aufgabe 4 (14 Punkte)

Sie schätzen ein Poissonmodell, in dem Sie die Anzahl der Arztbesuche erklären wollen. Sie haben folgende Informationen zu 14.675 Personen gegeben:

- doc* = Anzahl der Arztbesuche
gkv = 1, falls Person gesetzlich versichert, = 0 sonst
female = 1, falls Person weiblich, = 0 sonst
geburt = Geburtsjahr
educ = Bildung in Jahren

Sie erhalten folgenden Output:

```
Iteration 0: log pseudolikelihood = -38894.858
Iteration 1: log pseudolikelihood = -38894.858

Poisson regression                               Number of obs =      14675
                                                Wald chi2(4)      =      100.67
                                                Prob > chi2       =      0.0000
Log pseudolikelihood = -38894.858              Pseudo R2        =      0.0058

-----+-----
          |           Robust
          |           Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      gkv |  -.0943417   .0270359   -3.49  0.000   -0.147331   -0.0413523
  female |  .1142728   .0199458   5.73  0.000   .0751797   .1533659
  geburt |  -.0068225   .0009559   -7.14  0.000   -.008696   -.004949
    educ |  -.0105121   .0027562   -3.81  0.000   -.0159142   -.00511
    _cons |  14.56069   1.87507    7.77  0.000   10.88562   18.23576
-----+-----
```

- 3.1 Interpretieren Sie den Koeffizienten des Versicherungsstatus *gkv* inhaltlich und statistisch. Runden Sie auf die dritte Nachkommastelle. (2 Punkte)
- 3.2 Geben Sie die Dichtefunktion der Poissonverteilung an und stellen Sie die Likelihoodfunktion unter Berücksichtigung von x und β für das Poissonmodell auf. (5 Punkte)
- 3.3 Erläutern Sie kurz, was unter Überstreuerung zu verstehen ist. Was bedeutet das für die Schätzergebnisse? Beschreiben Sie ein Verfahren ausführlich, mit dem Sie auf Überstreuerung testen können. (5 Punkte)
- 3.4 Erläutern Sie den Unterschied zwischen einem Poisson und einem Negbin Modell. (2 Punkte)