

Masterprüfung Wintersemester 2020/2021

Fach: Mikroökonomie und Maschinelles Lernen

Semester: Wintersemester 2020/2021

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

Vorbemerkungen:

- Anzahl der Aufgaben:** Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen.
Es wird nur der Lösungsbogen eingesammelt.
- Bewertung:** Es können maximal 60 Punkte erworben werden. Die maximale Punktzahl für jede Aufgabe ist in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
- Erlaubte Hilfsmittel:**
- Tabellen der statistischen Verteilungen (sind der Klausur beigelegt)
 - Taschenrechner
 - Fremdwörterbuch
- Wichtige Hinweise:**
- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beigelegt sind, den gesuchten 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 (18 Punkte)

Der Zusammenhang zwischen der Art der Beschäftigung nach einer Arbeitslosigkeit und der Aufnahme eines Minijobs während der Arbeitslosigkeit (*minijob* = 1, falls Person einen Minijob ausgeübt hat; 0 sonst) soll mit einem Regressionsmodell analysiert werden. Drei Beschäftigungsarten werden unterschieden: 1 = Beschäftigung in Vollzeit (*fulltime*), 2 = Beschäftigung in Teilzeit (*parttime*), 3 = Sonstiges (*other*). Die Schätzung eines Multinomialen-Logit-Modells liefert folgende Ergebnisse:

```
Multinomial logistic regression          Number of obs =      13326
                                         LR chi2(2)      =      159.24
                                         Prob > chi2     =      0.0001
Log likelihood = -850.18                 Pseudo R2      =      0.0027
```

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

Sonstiges	(base outcome)					

Fulltime						
minijob	.1358066	.0209329	6.49	0.000	.1768344	.0947788
_cons	-1.616398	.0394833	-40.94	0.000	-1.539012	-1.693784

Parttime						
minijob	.0286888	.003749	7.65	0.000	.0213405	.0360371
_cons	-2.040842	.086105	-23.70	0.000	-2.209605	-1.872079

- 1.1 Berechnen Sie die Wahrscheinlichkeit des Zustands *Sonstiges* für eine Person, die keinen Minijob ausgeübt hat (d. h. *minijob*=0). (3 Punkte)
- 1.2 Wie lautet die unmittelbare inhaltliche Interpretation von Vorzeichen und Größe des geschätzten Koeffizienten der Variable *minijob* in der Antwortkategorie *parttime*? (2,5 Punkte)
- 1.3 Interpretieren Sie den nachfolgend ausgewiesenen marginalen Effekt für eine *Vollzeit Beschäftigung* inhaltlich. Berechnen Sie den marginalen Effekt für die Referenzkategorie *Other*, wenn der marginale Effekt für die Kategorie *parttime* den Wert 0,0153 annimmt. (3,5 Punkte)

```
Marginal effects after mlogit
y = Pr(insurance==fulltime) (predict, outcome(2))
= .44412634
-----
variable |          dy/dx              X
-----+-----
minijob*|          .0325797          .167126
-----
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- 1.4 Was besagt die IIA-Annahme? Erklären Sie am Beispiel, wann diese Annahme verletzt sein könnte. (4 Punkte)
- 1.5 Sie möchten einen Hausman Test der IIA-Annahme auf dem 5% Signifikanzniveau durchführen. Hierzu schätzen Sie zusätzlich zu dem oben ausgewiesenen Output ein Modell, welches die Alternative *parttime* ignoriert. Die Teststatistik lautet: $W = (b - B)' \cdot (V_b - V_B)^{-1} \cdot (b - B) = 6,26$. Stellen Sie Null- und Alternativhypothese auf und erläutern Sie die Eigenschaften der Schätzer unter den beiden Hypothesen. Geben Sie Freiheitsgrade, Entscheidungsregel mit kritischem Wert und Testentscheidung an. (5 Punkte)

Aufgabe 2 (12 Punkte)

Die Kreditvergabe (*kredit*, gemessen in Euro) von multinationalen Unternehmen wird als Funktion des logarithmierten Umsatzes (*ln_umsatz*) und der Firmengröße (*groesse*, gemessen als Anzahl der Mitarbeiter) modelliert. In einem Datensatz vergeben 1885 von 5800 Unternehmen keine Kredite. Die Schätzung eines Tobit-Modells liefert folgende Ergebnisse für die abhängige Variable *kredit*:

Variable	Tobit	
	Koeffizient	Std.fehler
ln_umsatz	6.501	(0.027)
groesse	-0.070	(0.000)
Konstante	-37.676	(0.105)
σ	0.995	(0.011)
$\Phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)$	0.350	
Anzahl Beobachtungen (N)	5,800	

- 2.1 Erläutern Sie am Beispiel der Variable *kredit* knapp den Unterschied zwischen Stutzung und Zensierung. Berechnen Sie für das vorliegende Beispiel die Anzahl der Beobachtungen, die bei der gestutzten Regression verwendet wird. (3 Punkte)
- 2.2 Wie lassen sich die Koeffizienten des Tobit-Modells direkt interpretieren? Interpretieren Sie den Koeffizienten der Variable für den logarithmierten Umsatz (*ln_umsatz*) inhaltlich. (2 Punkte)
- 2.3 Berechnen und interpretieren Sie den marginalen Effekt der Firmengröße auf die zensierte Höhe der Kredite (d. h. für alle Unternehmen). (3 Punkte)
- 2.4 Das Tobit-Modell wurde unter Verwendung der folgenden Log-Likelihood-Funktion geschätzt. Erklären Sie kurz, was die beiden Summen $\sum_{y_i=0}$ bzw. $\sum_{y_i>0}$ jeweils berechnen. Welche Annahme hinsichtlich $\boldsymbol{\beta}_1$ und $\boldsymbol{\beta}_2$ wird im Tobit-Modell gemacht? (3 Punkte)

$$\ln L = \sum_{y_i=0} \ln \Phi \left(\frac{0 - \mathbf{x}'_i \boldsymbol{\beta}_1}{\sigma} \right) + \sum_{y_i>0} \ln \frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}'_i \boldsymbol{\beta}_2}{\sigma} \right)$$

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

Aufgabe 3 (12 Punkte)

Sie schätzen ein Poissonmodell, in dem Sie die Anzahl der Kinder einer Frau erklären wollen. Sie haben folgende Informationen zu 14.675 Frauen:

kids = Anzahl der Kinder
age = Alter in Jahren
educ = Bildung in Jahren

Sie erhalten folgenden Output:

```
Iteration 0:  log pseudolikelihood = -38999.8
Iteration 1:  log pseudolikelihood = -38895.8

Poisson regression                               Number of obs =      14675
                                                Wald chi2(2)    =      100.67
                                                Prob > chi2     =      0.0000
Log pseudolikelihood = -38895.8                Pseudo R2      =      0.0058

-----
            |               Robust
            |               Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
    age |   .0943417   .0270359   3.49  0.000   .0413523   .147331
    educ |  -.0106121   .0027562  -3.81  0.000  -.0159142  -.00511
    _cons |  14.56069   1.87507   7.77  0.000   10.88562   18.23576
-----
```

- 3.1 Nennen Sie einen Vorteil des Poisson-Modells im Vergleich zum linearen Modell. (1 Punkt)
- 3.2 Interpretieren Sie den Koeffizienten der Bildung inhaltlich und statistisch. Runden Sie auf die dritte Nachkommastelle. (2 Punkte)
- 3.3 Erläutern Sie kurz, was unter Unterstreuerung zu verstehen ist. Was bedeutet das für die Schätzergebnisse? (2 Punkte)
- 3.4 Nennen Sie die Schritte eines Verfahrens, mit dessen Hilfe Sie prüfen können, ob Unterstreuerung vorliegt. (3 Punkte)
- 3.5 Benennen Sie zwei Unterschiede zwischen einem Poisson und einem Negbin Modell. (2 Punkte)
- 3.6 Sie möchten die Anpassungsgüte Ihres Modells mit dem Akaike Informationskriterium (AIC) prüfen. Berechnen Sie den Wert des AIC für das geschätzte Modell. Nennen Sie ein alternatives geeignetes Maß für die Schätzgüte Ihrer Schätzung und einen Vorteil gegenüber dem AIC. (*Hinweis*: $AIC = 2k - 2\ln L$.) (2 Punkte)

Aufgabe 4 (12 Punkte)

Sie analysieren die Dauer der Arbeitslosigkeit (gemessen in Wochen) mit einer Weibull-Regression. Ihr Datensatz enthält Informationen zu 3674 Arbeitslosen. Es liegen folgende erklärende Variablen vor:

$betreuung_i$ = 1, falls Betreuung durch Arbeitsvermittler; 0 sonst
 $alter_i$ Alter in Jahren von Person i

```

Weibull regression -- log relative-hazard form

No. of subjects =      3674                Number of obs =   3674
No. of failures =      2571
Time at risk   =        557
LR chi2(1)     =      243.86
Log likelihood = -1043.4721                Prob > chi2  =   0.0000

-----+-----
_t      | Coef.      Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
betreuung |  2.317153   .4855492    4.77  0.000    1.365494   3.268812
alter     | -0.2874122 .0391988   -3.41  0.001   -0.2106459 -0.0569896
_cons    | -0.2610042  1.479498   -0.18  0.860   -3.160767   2.638759
-----+-----
/ln_p    |  .4570259   .1665073    2.74  0.006    .1306776   .7833742
-----+-----
p        |  1.57937    .2629766    1.1396  2.188845
1/p      |  .6331639   .1054264    .4568619 .8775007
-----+-----

```

- 4.1 Erklären Sie die Begriffe *Flow Sample* und *Stock Sample*. Nennen sie eine Problematik, die sich beim *Stock Sample* ergibt. (3 Punkte)
- 4.2 Interpretieren Sie den Koeffizienten der Variable *alter* inhaltlich. (2 Punkte)
- 4.3 In einer weiteren Weibull-Schätzung wird das Geschlecht des Patienten als zusätzliche erklärende Variable in das Modell aufgenommen. Die Schätzung ergibt einen Log-Likelihood-Wert von -1040.6522 . Überprüfen Sie, ob sich der Erklärungsgehalt des Modells signifikant verbessert hat. Geben Sie die Teststatistik, Freiheitsgrade und kritischen Wert zum Signifikanzniveau $\alpha = 0.10$ an. Berechnen Sie den empirischen Wert der Teststatistik und treffen Sie eine Testentscheidung. (5 Punkte)

Aufgabe 5 (6 Punkte)

- 5.1 Die Ansätze des maschinellen Lernens und der Ökonometrie haben unterschiedliche Schwerpunkte. Benennen Sie knapp die unterschiedlichen Zielsetzungen der beiden Verfahren. (2 Punkte)
- 5.2 Lasso- und Ridge-Regression sind zwei Erweiterungen des linearen Regressionsmodells. Erläutern Sie, wie sich die beiden Verfahren vom herkömmlichen KQ Schätzer abheben und worin sich die beiden Verfahren voneinander unterscheiden? (3 Punkte)
- 5.3 Benennen Sie knapp zwei Verfahren, mit denen man im Bereich des maschinelle Lernens heterogene kausale Effekte identifiziert. (1 Punkt)