

Diplomprüfung

Fach: Mikroökonomie

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

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	18.07.2007
Raum	0.424
Unterschrift	

Vorbemerkungen:

Beachten Sie die **Anweisungen auf der Klausurangabe.**

Aufgabe	Punkte
Aufgabe 1	
Aufgabe 2	
Aufgabe 3	
Aufgabe 4	
Aufgabe 5	
Aufgabe 6	
Aufgabe 7	
Summe	
Note	

Klausur in Mikroökonomie am 18.07.2007

Hinweise:

- (1) Teil 1 (90 Punkte) muss von allen gelöst werden, Teil 2 (30 Punkte) nur von denjenigen, die 4 oder 6 CP erwerben wollen.
- (2) Die Punktverteilung der Aufgaben entspricht dem empfohlenen zeitlichen Gewicht bei der Beantwortung, es stehen 90 bzw. 120 Minuten zur Verfügung.
- (3) Erlaubte Hilfsmittel sind Taschenrechner und die anhängenden Verteilungstabellen.

Bitte tragen Sie Ihren Namen und Ihre Matrikelnummer ein.

Name

Matrikelnummer

Zahl der Kreditpunkte, die Sie erwerben möchten: 3 CP 4 CP 6 CP

ERSTER TEIL: 5 AUFGABEN FÜR ALLE

Aufgabe 1

[22 Punkte]

Sie interessieren sich für die Auswirkungen der Benzinpreise auf das Fahrverhalten von Privatpersonen. Sie sammeln daher Informationen über 500 Haushalte (i) in 24 aufeinanderfolgenden Monaten (t) zu folgenden Variablen:

- Kilometer_{it} gefahrene Autokilometer
- Benzin_{it} durchschnittlicher Benzinpreis pro Liter in der Wohnsitzgemeinde (in €)
- Regen_{it} Niederschlagsmenge in der Wohnsitzgemeinde (in l/m²)
- Einkommen_{it} Einkommen des Haushalts (in 100€)
- Ferien_t im Beobachtungsmonat t sind Schulferien (1=ja, 0=nein)
- Kinder_i im Haushalt i leben Kinder (1=ja, 0=nein)
- Distanz_i Entfernung zwischen Wohnsitzgemeinde des Haushalts und Arbeitsort des Haushaltsvorstandes (in km)

(Während des Beobachtungszeitraums sind die Variablen Kinder_i und Distanz_i für jeden Haushalt zeitkonstant.)

Sie schätzen zwei verschiedene Modelle, die die gefahrenen Autokilometer erklären sollen und erhalten folgende Schätzergebnisse:

Variable	gepooltes OLS		Fixed Effects	
	Koeffizient	Standardfehler	Koeffizient	Standardfehler
Benzin	-98,34	13,4	-112,76	1,8
Regen	15,4	6,1	16,34	5,9
Einkommen	23,87	8,3	13,48	5,9
Ferien	8,28	4,9	7,32	4,7
Kinder	27,31	8,4	-	-
Ferien*Kinder	45,49	2,3	49,61	3,9
Distanz	40,82	11,7	-	-
Konstante	-13,48	2,76	-28,43	2,93

F-Test auf H₀: "Fixed Effects=0", p-Wert: 0,04

- a) Während Sie für einen Teil der Variablen auf amtliche Statistiken zurückgreifen können, müssen Sie unter anderem die gefahrenen Autokilometer und das Haushaltseinkommen durch Haushaltsbefragungen ermitteln. Erläutern sie kurz zwei Probleme, die sich aus der Datenqualität ergeben könnten, und ihre Auswirkungen auf die Schätzung. (4 Punkte)

- b) Halten Sie es für sinnvoll, die Panelstruktur der Daten im vorliegenden Fall zu berücksichtigen? Beziehen Sie in Ihre Beurteilung ein Testergebnis, aber auch mögliche andere Konsequenzen der Panelkorrektur ein. (4 Punkte)
- c) Interpretieren Sie die Steigungskoeffizienten der Fixed Effects Schätzung ausführlich statistisch und inhaltlich. (8 Punkte)
- d) Nehmen Sie an, die tatsächlich gefahrenen Autokilometer seien für Ihre Untersuchung nicht hinreichend genau ermittelbar, Sie hätten aber verlässliche Information darüber, ob ein Haushalt in einem Monat mehr als 1000km gefahren ist oder nicht. Welches Fixed Effects Modell würden Sie zur Schätzung nun heranziehen? Begründen Sie Ihre Wahl und erläutern Sie knapp, wie in diesem Modell die unbeobachtete Heterogenität berücksichtigt wird. Geben Sie einen Vor- und einen Nachteil des gewählten Verfahrens an. (6 Punkte)

Aufgabe 2**[22 Punkte]**

Es wird für 300 Familien die Wahrscheinlichkeit untersucht, drei verschiedene Restauranttypen zu besuchen (Familienrestaurant, Schnellrestaurant, Delikatessenrestaurant). Die erklärenden Variablen sind:

- Durchschnittliche Kosten pro Person in Euro in den unterschiedlichen Restauranttypen
- Die Entfernung des Wohnorts der Familien zu den jeweiligen Restaurants in Kilometern
- Das Familieneinkommen in 100 Euro
- Die Anzahl der Kinder.

Im Datensatz wählen 10% der Familien ein Schnellrestaurant, 70% ein Familienrestaurant und 20% ein Delikatessenrestaurant.

Geschätzt wird ein Conditional Logit und ein Mixed Logit Modell. Die Ergebnisse finden Sie in untenstehender Tabelle:

Variable	Conditional Logit	Mixed Logit
Kosten	-0,03 (0,010)	-0,08 (0,025)
Entfernung	-0,20 (0,040)	-0,19 (0,047)
Einkommen Schnellrest.		-0,04 (0,009)
Einkommen Delikatessenrest.		0,04 (0,008)
Kinder Schnellrest.		-0,30 (0,103)
Kinder Delikatessenrest.		-0,41 (0,113)
Konstante Schnellrestaurant		-1,09 (0,583)
Konstante Delikatessenrest.		-2,55 (0,518)
Log-likelihood	-578,91	-491,23

Standardfehler in Klammern

Referenzkategorie: Familienrestaurant

- a) Geben Sie an, in welcher Richtung sich im Mixed Logit Modell die Wahrscheinlichkeiten der einzelnen Kategorien ändern, wenn (6 Punkte)
- i) sich das Familieneinkommen erhöht.
 - ii) sich die Entfernung zum Delikatessenrestaurant erhöht.

- b1) Geben Sie drei Testmethoden an, mit denen Sie Hypothesen bezüglich der beiden Schätzungen testen können. Skizzieren Sie kurz die Vorgehensweise und die Idee der drei Methoden. (9 Punkte)
- b2) Führen Sie einen Test durch, um zwischen den beiden Schätzmodellen zu wählen. Geben Sie dabei die Null- und Alternativhypothese, die Teststatistik, den kritischen Wert und die Freiheitsgrade an. (3 Punkte)
- c) Sie nehmen noch eine Kategorie Pizzeria auf, die dann von 50% der Familien gewählt wird. Erläutern Sie die betroffene Modellannahme. Wie müssen sich die Wahrscheinlichkeiten der anderen Kategorien ändern, wenn diese gelten soll? (4 Punkte)

Aufgabe 3**[20 Punkte]**

Sie schätzen ein Poissonmodell, mit dem Sie die Anzahl der Ausflüge von Familien an den Brombachsee im Jahr 2006 erklären wollen. Ihr Datensatz enthält Informationen zu 800 Familien. Erklärende Variablen sind:

- Das Einkommen in 9 Kategorien (1=niedrig bis 9=hoch)
- Die Zahl der Kinder (0=keine Kinder bis 7 Kinder)
- Der Wohnort (1=Mittelfranken, 0=andere Region)

Variable	Poisson	Negbin
Einkommen	-0,11 (-1,80)	-0,09 (-0,58)
Kinder	0,47 (27,60)	0,50 (15,9)
Wohnort	0,90 (11,37)	0,87 (1,85)
Konstante	0,26 (2,83)	-1,12 (-5,08)

t-Werte in Klammern

- a) Interpretieren Sie die Steigungskoeffizienten des Poissonmodells statistisch und inhaltlich. (5 Punkte)
- b) Skizzieren Sie algebraisch, wie eine Likelihoodfunktion aufgestellt wird und erläutern Sie verbal, was die Likelihoodfunktion repräsentiert und wie sie analytisch genutzt wird. Stellen Sie die Likelihoodfunktion unter Berücksichtigung von x und β für das Poissonmodell auf. (6 Punkte)
- c) Sie vermuten, dass in Ihren Daten Überstreuerung vorliegt. Was bedeutet das für die Schätzergebnisse? Beschreiben Sie ein Verfahren, mit dem Sie auf Überstreuerung testen können. (5 Punkte)
- d) Erläutern Sie einen Unterschied zwischen dem Poisson Modell und dem Negbin Modell. Interpretieren Sie den Unterschied zwischen den Negbin und Poisson Resultaten in obiger Tabelle. (4 Punkte)

Aufgabe 4**[17 Punkte]**

Wahr oder falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr) und für nicht zutreffende Aussagen den Buchstaben f (für falsch) ein. Für jede richtige Antwort gibt es einen Punkt, für jede falsche Antwort wird ein Punkt abgezogen. Nicht beantwortete Fragen werden nicht berücksichtigt. Die Gesamtpunktzahl kann nicht negativ werden.

w/f	
	Beim Probit Schätzer kann der wahre Koeffizient β nicht bestimmt werden.
	Solange ein lineares Regressionsmodell eine Konstante enthält, ist es unproblematisch, wenn der auf X bedingte Erwartungswert von ε konstant und größer als Null ist.
	Bei Heteroskedastie im Standard Probit Modell ist die Verteilungsannahme für die abhängige Variable falsch.
	Im linearen Random Effects Modell hat es keine Auswirkung auf die Eigenschaft des Schätzers, wenn die unbeobachtbare Heterogenität mit den erklärenden Variablen korreliert ist.
	Die Likelihoodfunktion hat ein eindeutiges globales Maximum, falls die Hessematrix der zweiten Ableitungen nach den Parametern negativ definit ist.
	Je flacher die Likelihoodfunktion in der Region um das Maximum, umso präziser ist der ML-Schätzer.
	Die marginalen Effekte von Logit- und Probitmodellen können sehr ähnlich sein, auch wenn sich die geschätzten Parameter voneinander unterscheiden.
	Der Wald-Test überprüft, ob der „Schattenpreis“ der Restriktion nahe Null ist.
	Im Nested Logit Modell werden keine Korrelationen zwischen den Alternativen der abhängigen Variable zugelassen.
	Bei einer Difference-in-Difference Schätzung wird der kausale Effekt einer Maßnahme ermittelt, indem man Differenzen in erklärenden Variablen bildet.
	Da ML-Schätzer asymptotisch normalverteilt sind, ist die t-Teststatistik in diesen Modellen asymptotisch t verteilt.
	Im geordneten Probitmodell wird pro Kategorie der abhängigen Variable ein Parametervektor geschätzt.
	Ein Messfehler in der erklärenden Variable kann dazu führen, dass ein bei korrekter Messung signifikant von Null verschiedener Koeffizient insignifikant wird.
	Die Schätzung eines multinomialen Probitmodells für geordnete Daten ergibt inkonsistente Ergebnisse, da nicht alle Informationen berücksichtigt wurden.
	Ein Schätzer $\hat{\theta}$ ist effizient, wenn er eine kleinere Varianz hat als jeder andere Schätzer.
	Das McFadden R^2 nimmt den Wert Null an, wenn der unrestringierte Log-Likelihoodwert Null ist.
	Man nennt erklärende Variablen endogen, wenn sie untereinander korreliert sind.

Aufgabe 5**[9 Punkte]**

Sind folgende Aussagen richtig? Geben Sie in der ersten Spalte zunächst w (für wahr) oder f (für falsch) an. Erläutern Sie in der zweiten Spalte stichwortartig Ihre Auffassung (Bsp.: "Stimmt, weil..." oder "Stimmt nicht, weil..."). Nur bei korrekter Begründung wird die Antwort mit 1,5 Punkten je Frage honoriert.

w/f	Begründung
	Je kleiner die Schwellenparameterdifferenz im Verhältnis zum geschätzten Koeffizienten im geordneten Probitmodell ist, desto einflussreicher ist die betrachtete erklärende Variable.
	Das Nested Logit Modell löst das Independence of Irrelevant Alternatives Problem (IIA).
	Beim Random Effects Modell braucht man das Incidental Parameter Problem nicht zu beachten.
	Stichprobenselektion kann nur durch natürliche Experimente, nicht durch soziale Experimente vermieden werden.
	Die Freiheitsgrade einer Wald-Teststatistik bestimmen sich aus der Zahl der geschätzten Parameter.
	Das Grid Search Verfahren ist robust gegenüber lokalen Maxima der (Log-)Likelihood-Funktion.

ZWEITER TEIL: 2 AUFGABEN NUR FÜR 4 ODER 6 CP KLAUSUR

Aufgabe 6

[20 Punkte]

Sie planen die Gründung einer Partyagentur und führen im Vorfeld verschiedene Untersuchungen über das Verhalten ihrer potenziellen Kunden durch. Unter anderem interessieren Sie sich für Faktoren, die die Verweildauer auf einer Feier bestimmen. Sie besuchen daher selbst um 22.00h eine Uniparty und befragen 50 zufällig ausgewählte Gäste nach ihrem Alter, Geschlecht und ihren bisherigen Getränkeausgaben. Sie wiederholen die Befragung zu den Getränkeausgaben bei denselben Personen um 23.00h, um 0.00h und um 1.00h unter den jeweils noch Anwesenden, sodass Sie Informationen zu drei abgeschlossenen Zeitintervallen zur Verfügung haben. Zudem notieren Sie sich, wer seit der letzten Befragung die Party verlassen hat. Danach verlassen Sie die noch andauernde Feier.

- a) Erläutern Sie anhand des Beispiels, was man unter rechts- bzw. linksenzierten Episoden versteht. (2 Punkte)
- b) Erläutern Sie anhand des Beispiels, was man unter negativer bzw. positiver Duration Dependence versteht. (2 Punkte)
- c) Ist Ihre Stichprobe ein Stock Sample oder ein Flow Sample? Erläutern Sie kurz. (2 Punkte)
- d) Sie möchten zunächst nur die Verweildauer abbilden. Berechnen Sie hierzu den Kaplan-Meier-Schätzer für die Survivalfunktion für die folgenden, beispielhaft ausgewählten sechs Personen und bilden Sie ihn grafisch ab. (4 Punkte)

Person	Hat die Party verlassen zwischen ...	
1	22.00h und 22.59h	(während des ersten Befragungszeitraums)
2	22.00h und 22.59h	(während des ersten Befragungszeitraums)
3	23.00h und 23.59h	(während des zweiten Befragungszeitraums)
4	0.00h und 0.59h	(während des dritten Befragungszeitraums)
5	0.00h und 0.59h	(während des dritten Befragungszeitraums)
6	nach 1.00h	(nach Ende des letzten Befragungszeitraums)

- e) Betrachten Sie nun wieder den gesamten Datensatz über alle 50 befragten Personen. Sie möchten nun die Hazardrate für das Verlassen der Party durch die bisherigen Getränkeausgaben der Personen bis zur Befragung t (G_t) erklären. Sie schätzen ein diskretes, zeitlich vollständig interagiertes Modell.
 - e1) Beschreiben Sie kurz notwendige Schritte zur Datenaufbereitung, die abhängige Variable und stellen Sie die Modellgleichung auf. (6 Punkte)
 - e2) Interpretieren Sie die folgenden Ergebnisse nur inhaltlich: (4 Punkte)

Variable	Koeffizient	Standardfehler
bisherige Getränkeausgaben (G_t)	-0,042	0,012
$G_t * T_{23-24h}$	0,007	0,015
$G_t * T_{0-1h}$	0,083	0,019

T_{23-24h} ist Eins im zweiten Befragungszeitraum, Null sonst.

T_{0-1h} ist Eins im dritten Befragungszeitraum, Null sonst.

Aufgabe 7**[10 Punkte]**

Wahr oder falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr) und für nicht zutreffende Aussagen den Buchstaben f (für falsch) ein. Für jede richtige Antwort gibt es einen Punkt, für jede falsche Antwort wird ein Punkt abgezogen. Nicht beantwortete Fragen werden nicht berücksichtigt. Die Gesamtpunktzahl kann nicht negativ werden.

w/f	
	Der Erwartungswert einer Variable x steigt, wenn sie von oben gestutzt wird und fällt, wenn sie von unten gestutzt wird.
	Werden bei Vorliegen zensierter Daten die zensierten Beobachtungen nicht berücksichtigt, ist das Schätzergebnis inkonsistent.
	Im Tobitmodell kann durch den Fixed Effects Schätzer das „Incidental Parameter Problem“ umgangen werden.
	Wird eine nicht-zufällige Stichprobe aufgrund eines exogenen, beobachteten Kriteriums erhoben, lässt sich eine Verzerrung des KQ-Schätzers durch Gewichtung korrigieren.
	Beim einfachen Heckman-Selektionsmodell wird Selektionsverzerrung korrigiert, indem man den Koeffizienten um die „inverse Mill's Ratio“ erhöht.
	Das Weibull-Modell unterstellt eine konstante Hazard-Rate.
	Die Likelihoodfunktion von Verweildauermodellen in stetiger Zeit berücksichtigt links- und rechtszensierte Beobachtungen jeweils mit der Wahrscheinlichkeit, dass ein Spell zensiert wird.
	Die Überlebenswahrscheinlichkeit $S(t)$ lässt sich aus der Hazardrate und ihrem zeitlichen Verlauf ermitteln.
	Ein Vorteil der Heckman-Selektionskorrektur besteht darin, dass die erklärenden Variablen für den Selektionsmechanismus nur für diejenigen Beobachtungen vorliegen müssen, für die auch die interessierende abhängige Variable beobachtet wird.
	Bei zensierten Daten ist der Erwartungswert für die Bevölkerung das Produkt aus dem Erwartungswert der gestutzten Verteilung und der Wahrscheinlichkeit, unzensiert zu sein.