Klausur zur Vorlesung Verkehrsökonometrie,

für Master-Studenten

WS 2008/09

Lösungsvorschlag

Aufgabe 1 (20 Punkte)

Gegeben sind folgende Modelle (Sie müssen nicht verstehen, was sie modellieren oder was die einzelnen Größen bedeuten):

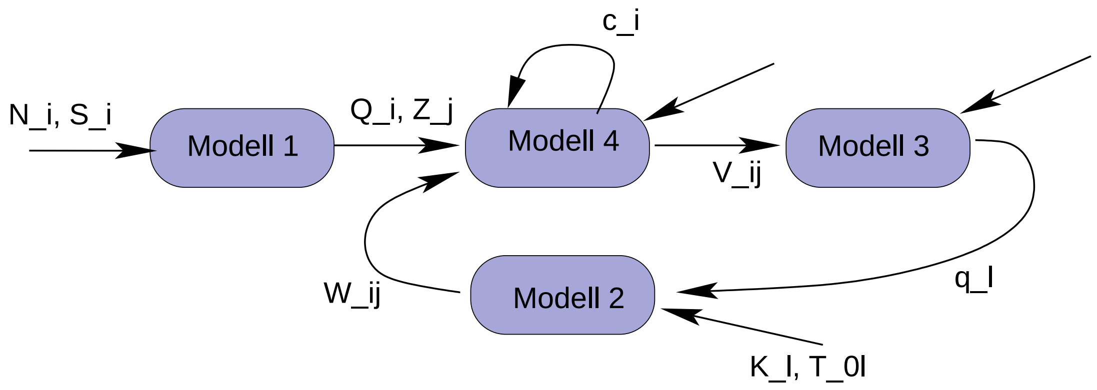
$$\begin{aligned}
&\underline{\text{Modell 1:}} & Q_i &= g(N_i, S_i), \quad Z_j = h(N_j, S_j) \\
&\underline{\text{Modell 2:}} & W_{ij} &= \sum_{l \in \{l_{ij}\}} \left[T_{0\,l} + \left(\frac{q_l}{K_l} \right)^4 \right] \\
&\underline{\text{Modell 3:}} & q_l &= \sum_{ij} f_{lij}(V_{ij}) \\
&\underline{\text{Modell 4:}} & V_{ij} &= c_i B_{ij} Q_i Z_j, \quad c_i = \frac{1}{\sum\limits_k B_{ik} Z_k}, \quad B_{ij} = B(W_{ij})
\end{aligned}$$

(a) *Geben Sie von jedem Modell die exogenen und endogenen Variablen an. Hinweis: Die Funktionen g, h, f_{lij} und B sind Bestandteile der jeweiligen Modelle.*

Die exogenen, also die von außen kommenden Input-Variablen stehen auf der rechten Seite als Funktionsargumente, und die endogenen (abhängigen, Output-) Varianblen auf de linken.

- Modell 1: Endogen: Q_i, Z_j, exogen N_i und S_i (nicht zusätzlich N_j und S_j; dies dieselben Variablen, nur mit anderen Laufindex). [Hintergrundinfo (nicht verlangt): Erzeugungsmodell; Q_i, Z_j sind Quell- und Zielsummen und die N_i und S_i Raumstrukturdaten Raumstrukturvariablen.]
- Modell 2 (Definition des Streckenwiderstands von i nach j, Teil des Umlegungsmodells): Endogen W_{ij} (Streckenwiderstände), exogen q_l (Belastungen der Kanten), Kapazitäten K_l, Mindestzeiten T_{0l}. Eventuell auch die Definition der Linkmengen l_{ij} am externen Pfeil zu Modell 3, da sie durch das Netzwerk als solches definiert werden, dies war aber hier nicht verlangt.
- Modell 3: Endogen q_l, exogen V_{ij} (Umlegungsmodell)
- Modell 4: Endogen V_{ij} (Verkehrsstrommatrix), exogen Q_i, Z_j (Verkehrsverteilungsmodell)

(b),(c) *Tragen Sie die Modelle 2 bis 4 in die noch freien Plätze des Flussdiagramms ein. Tragen Sie nun auch alle Variablen an die entsprechenden Pfeil-Linien ein (ggf. mehrere Variable pro Pfeil, es können auch Pfeile leerbleiben). Geben Sie alle Verkettungen und Rückkopplungen an.*



Verkettungen: Modell 1-4-3-2

Rückkopplungen: Über die Variablen W_{ij} von Modell 2 (endogene Variable) nach Modell 4 als exogene Variable.

Nichtverkettete (echt von außen kommende) exogene Variable: N_i, S_i, K_l (Streckenkapazitäten) und T_{0l} (Mindestzeiten auf den Streckenelementen). Eventuell auch die Definition der Linkmengen l_{ij} am externen Pfeil zu Modell 3, da sie durch das Netzwerk als solches definiert werden, siehe weiter oben (nicht für volle Punktzahl verlangt).

Aufgabe 2 (10 Punkte)

Die Bahn will in einer Kampagne notorische Autofahrer von ihren Dienstleistungen überzeugen. Die Frage, in wie weit sich Vorlieben für bestimmte Verkehrsmittel (z.B. aufgrund des Elternhauses oder einer politischen Überzeugung) langfristig überhaupt ändern lassen, soll anhand von empirischen Untersuchungen beantwortet werden. Ist dabei eine Quer- oder eine Längsschnittuntersuchung sinnvoll? Sind Paneldaten geeignet?

Eine Längsschnittuntersuchung, d.h. eine geringe Stichprobe, welche über lange Zeit kontinuierlich oder in mehreren zeitlich gestaffelten Erhebungen untersucht wird. Idealerweise eine **Panelerhebung**, d.h. die Personen der Stichprobe (das "Panel") sind in den gestaffelten Erhebungen immer die sleben. Nur so lassen sich langsfristige Änderungen von individuellen Neigungen (die ja an Personen gebunden sind!) messen.

Aufgabe 3 (30 Punkte)

Der Besitzer eines großen Autohauses versucht anhand der Verkaufszahlen des letzten Jahres zu ermitteln, wie stark der Preis und die Motorleistung in die Kaufentscheidung einfließt. Für die Fahrzeugtypen im mittleren Segment erhält er folgende Tabelle:

Preis x_1 (in 1000 €)	19	19	20	25	25	30	33	37	41	45
Leistung x_2 (kW)	90	75	90	70	140	160	135	85	170	140
Verkaufszahl y (Fahrzeuge)	89	110	50	31	47	70	50	5	45	8

Gegeben aus Aufgabenstellung:

$$\bar{x}_1 = 29.4, \quad \bar{x}_2 = 115.5, \quad \bar{y} = 50.5, \quad \underline{\underline{S}} = \begin{pmatrix} 79.2 & 190 \\ 190 & 1247 \end{pmatrix}, \quad \begin{pmatrix} s_{1y} \\ s_{2y} \end{pmatrix} = \begin{pmatrix} -198 \\ -167 \end{pmatrix}.$$

(a) *Bestimmen Sie die Koeffizienten der Regressionsfunktion*

$$\hat{y}(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Determinante der Varianz-Kovarianz-Matrix:

$$\text{Det } \mathbf{S} = s_{11}s_{22} - s_{12}^2 = \underline{\underline{62808}}.$$

Aus dem Skript, Koeffizienten für den bivariaten Fall

$$\begin{aligned}
\hat{\beta}_1 &= \frac{s_{1y}s_{22} - s_{2y}s_{12}}{\det\left(\underline{\underline{S}}\right)} = \underline{\underline{-3.42}}, \\
\hat{\beta}_2 &= \frac{s_{2y}s_{11} - s_{1y}s_{21}}{\det\left(\underline{\underline{S}}\right)} = \underline{\underline{0.386}}, \\
\hat{\beta}_0 &= \bar{y} - \beta_1\bar{x}_1 - \beta_2\bar{x}_2 = \underline{\underline{106.4}}.
\end{aligned}$$

(b) *Sagen Sie in Worten, was die Anstiegsparameter und ihre Zahlenwerte bedeuten. Falls Sie (a) nicht gelöst haben, nehmen Sie hier und im Folgenden $\beta_1 = -3$ und $\beta_2 = 0.4$ an (diese Werte sind nicht das Ergebnis von Teil (a)). Angesichts der rapide sinkenden Nachfrage sinnt der Autohausbesitzer nach Mitteln, den Umsatz zu steigern. Mit wieviel mehr verkauften Fahrzeugen könnte er rechnen, wenn er eine jeweils um 10 kW stärkere Version um 2 000 € billiger als die Originalversion anbieten würde?*

- Anstiegsparameter β_1: Wieviel Autos pro 1 000 € pauschalen Preisanstieg bei ceteris-paribus-Bedingungen mehr verkauft werden würden (offensicht;ich ist dieser Koeffizient negativ), also drei Fahrzeuge weniger pro 1 000 € Aufpreis.
- Anstiegsparameter β_2: Wieviel Autos pro 1 kW Motor-Mehrleistung bei allen Typen bei ansonsten gleichen Preis mehr verkauft werden würden: 0.4 Fahrzeuge pro kW bzw. 1 Auto pro 2.5 kW an Mehrleistung.
- Erwarteter Mehrverkauf bei $\Delta x_1 = -2$ und $\Delta x_2 = 10$:

$$\Delta y = -2\beta_1 + 10\beta_2 = \underline{\underline{10.7\,\text{Fahrzeuge}}}.$$

(c) *Wieviel Euro ist unter ceteris paribus Bedingungen ein kW an Leistung "wert"?*

Berechnung der neutralen Fläche:

$$\Delta y = \beta_1\Delta x_1 + \beta_2\Delta x_2 \overset{!}{=} 0 \quad \Rightarrow \quad \beta_1\Delta x_1 = -\beta_2\Delta x_2 \Rightarrow \frac{\Delta x_1}{\Delta x_2} = -\frac{\beta_2}{\beta_1} = \underline{\underline{0.113}}$$

Also (da x_1 in 1 000 € gemessen wird) 113 € pro kW Mehrleistung.

(d) *Berechnen Sie die Konfidenzintervalle für die Anstiegsparameter zu einer Fehlerwahrscheinlichkeit von 5%. Hinweis: $\sum_i (y_i - \hat{y}(x_{1i}, x_{2i}))^2 = 3593$. Beachten Sie auch, dass die Breite der Intervalle mit der Gesamtzahl der Autokäufer (nicht der Autotypen) sinkt!*

Zunächst Schätzung der Residualvarianz: In der Formel aus dem Skript ist J die Zahl der erklärenden bzw. unabhängigen Variablen, also hier $J = 2$. Damit mit $n = 10$ Datenpunkte und dem gegebenen Wert 3593 für die Abweichungsquadratsumme:

$$\hat{\sigma}_R^2 = \frac{1}{n-1-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{3593}{7} = \underline{\underline{513}}.$$

Um aus dem Skript die Formel für die Varianz der Anstiegsparameter anwenden zu können, wird (zumindest in der zweiten Form der untenstehenden Gl. der Varianzen) noch die Korrelation der unabhängigen Variablen benötigt:

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}} = \underline{\underline{0.604}}$$

Damit aus dem Skript die Formel für die Varianz der Anstiegsparameter:

$$\begin{aligned} V\left(\hat{\beta}_1\right) &= \frac{\sigma_R^2 s_{22}}{n \text{ Det}(\mathbf{S})} = \frac{\sigma_R^2}{n s_{11}\left(1 - r_{12}^2\right)} = \underline{\underline{1.019}}, \\ V\left(\hat{\beta}_2\right) &= \frac{\sigma_R^2 s_{11}}{n \text{ Det}(\mathbf{S})} = \frac{\sigma_R^2}{n s_{22}\left(1 - r_{12}^2\right)} = \underline{\underline{0.06476}} \end{aligned}$$

Mit dem Quantil aus der beigefügten Tabelle,

$$t_{0.975}^{(7)} = 2.365,$$

erhält man

$$\begin{aligned} \text{KI}(\beta_1) &= [\hat{\beta}_1 - t_{0.975}^{(7)}\sqrt{V(\hat{\beta}_1)}, \hat{\beta}_1 + t_{0.975}^{(7)}\sqrt{V(\hat{\beta}_1)}] &= \underline{\underline{[-5.81, -1.03]}}, \\ \text{KI}(\beta_2) &= [\hat{\beta}_1 - t_{0.975}^{(7)}\sqrt{V(\hat{\beta}_2)}, \hat{\beta}_1 + t_{0.975}^{(7)}\sqrt{V(\hat{\beta}_2)}] &= \underline{\underline{[-0.216, 0.988]}} \end{aligned}$$

Nur das erste Konfidenzintervall ist also von Null signifikant unterschiedlich (nicht verlangt).

(e) *Um zu testen, ob der Treibstoffverbrauch x_3 auch ein wichtiges Kriterium des Autokaufs ist führt der Autohausbesitzer (i) eine univariate lineare Regression der Verkaufszahl bezüglich des Verbrauchs, (ii) eine bivariate Regression bezüglich Preis und Verbrauch, (iii) eine trivariate Regression bezüglich aller drei Variablen durch. Neben den Korrelationen $r_{13} = 0.65$ und $r_{23} = 0.71$ erhält er bei (i) und (iii), wie erwartet, einen negativen Anstiegsparameter β_3 bezüglich des Verbrauchs, im Fall (ii) jedoch einen positiven. Wie ist dies zu erklären?*

- **Fall 1, univariate Regression** $\hat{y}(x_3) = \beta_0 + \beta_3 x_3$: Hier ist der Anstiegsparameter negativ, was aber weniger auf den Verbrauch, sondern auf die Korrelation des Verbrauchs mit dem Preis zurückzuführen ist ($r_{13} = 0.65$). Die gemeinsame Ursache ist natürlich zum großen Teil die Motorleistung.
- **Fall 2, bivariate Regression** $\hat{y}(x_1, x_3) = \beta_0 + \beta_1 x_1 + \beta_3 x_3$ Hier wird der Preis berücksichtigt, d.h. β_3 gibt die Änderung der Verkaufszahl mit dem Verbrauch bei konstantem Preis an. Dieser kann durchaus positiv sein, nicht, weil die Leute Spritfresser lieben, sondern weil sie zum gleichen Preis lieber ein leistungsstärkeres Auto wollen und mehr Leistung sehr stark mit mehr Verbrauch gekoppelt ist ($r_{23} = 0.71$, was viele in Kauf nehmen.
- **Fall 3, trivariate Regression.** Hier ist der Einfluss der Leistung durch einen extra Parameter β_2 beschrieben. Bei konstantem Preis *und* konstanter Leistung bevoprzugen die Leute natürlich niedrigen Verbrauch, weshalb hier β_3 wieder negativ ist.

Aufgabe 4 (30 Punkte)

Ein Ökonometriker fragt im Bekanntenkreis fünf Personen (zugegeben eine kleine Stichprobe) nach ihrer Verkehrsmittelwahl für den Arbeitsweg der letzten 2 Wochen, also jeweils 10 Entscheidungen. Das Ergebnis ist wie folgt:

Merkmal / Person	Reisezeit (h) Auto/ÖV	Kosten (€) Auto/ÖV	Geschlecht	Gewählt Auto (i=1)	Gewählt ÖV (i=2)
$n = 1$	0.6/0.8	3.00/3.50	m	10	0
$n = 2$	0.5/0.5	3.00/1.50	w	1	9
$n = 3$	1000/0.5	1000/2.00	m	0	10
$n = 4$	1.2/1.5	4.00/3.00	w	6	4
$n = 5$	0.5/1.0	3.00/6.00	w	8	2

(a) *Welche generischen (alternativenspezifischen) bzw. sozioökonomischen Variablen wurden gelistet? Welche weitere sozioökonomische Variable wurde durch die hohen Einträge von 1 000 h bzw. 1 000 € implizit berücksichtigt?*

Generische Variablen: Reisezeit und Kosten

Sozioökonomischen Variablen: Geschlecht

Weitere sozioökonomischen Variable, welche implizit im Modell durch die hohe Disutility von 1000 h und 1000 € enthalten ist: Kfz-Verfügbarkeit (mit der Ausprägung “nein” bei Person 3).

(b) *Die Daten werden mit dem Binomial-Logit-Modell ausgewertet. Folgende Nutzenfunktion kommt dabei zum Einsatz:*

$$V_{ni} = \beta_1 T_{ni} + \beta_2 T_{ni}\delta_{n1} + \beta_3 K_{ni} + \beta_4 K_{ni} G_n + \beta_5 \delta_{n1}.$$

Dabei bezeichnet T_{ni} die Zeit in h, K_{ni} die Kosten in € und G_n hat die Werte 1 bzw. 0 für weibliche bzw. männliche Personen. Geben Sie die fünf Gleichungen zum Kalibrieren der 5 Parameter an.

Die Merkmalssummen aller fünf Merkmale bzw. Kriterien in der Nutzenfunktion muss im Modell im Mittel gleich den realisierten Merkmalssummen sein. Zunächst sind die Zahlen der Entscheidungen einer Person n gegeben durch

$$y_n = 10, \quad n = 1, ..., 5$$

Damit die Kalibrierungsbedingungen $x_m^{\text{Logit}}(\boldsymbol{\beta}) = x_m^{\text{data}}$, $m = 1, ..., 5$:

$$\begin{aligned}
\sum_n \sum_i y_n P_{ni}(\boldsymbol{\beta}) T_{ni} &= \sum_n \sum_i y_{ni} T_{ni} &&= \underline{\underline{35.2}} \quad \text{(Zeit in h gesamt)},\\
\sum_n y_n P_{n1}(\boldsymbol{\beta}) T_{n1} &= \sum_n y_{n1} T_{n1} &&= \underline{\underline{17.7}} \quad \text{(Zeit in Kfz)}\\
\sum_n \sum_i y_n P_{ni}(\boldsymbol{\beta}) K_{ni} &= \sum_n \sum_i y_{ni} K_{ni} &&= \underline{\underline{138.5}} \quad \text{(Kosten in € gesamt)},\\
\sum_{n=2,4,5} \sum_i y_n P_{ni}(\boldsymbol{\beta}) K_{ni} &= \sum_{n=2,4,5} \sum_i y_{ni} K_{ni} &&= \underline{\underline{88.5}} \quad \text{(Kosten in € durch weibl. Teilnehmer)},\\
\sum_n y_n P_{n1}(\boldsymbol{\beta}) &= \sum_n y_{n1} &&= \underline{\underline{25}} \quad \text{(Gesamtzahl der Kfz-Entscheidungen)}
\end{aligned}$$

(c) *Die Parameterwerte seien nun*

$$\beta_1 = -10, \quad \beta_2 = -5, \quad \beta_3 = -0.5 \quad \beta_4 = -0.5, \quad \beta_5 = 2.$$

Was sagen die fünf Modellparameter aus? Wie hoch ist der Zeitwert (€/h), den die männlichen und weiblichen Personen jeweils für Fahrten mit dem Auto- bzw. dem ÖV veranschlagen? Wie hoch (in Zeiteinheiten) ist die generelle Bevorzugung für das Verkehrsmittel "Auto"?

- β_1: Nutzenanstieg pro h Reisezeit (offensichtlich ist β_1 negativ),
- β_2: Zusätzlicher Nutzenanstieg pro h Reisezeit beim Auto
- β_3: Nutzenanstieg pro € Mehrkosten (offensichtlich ist β_3 negativ),
- β_4: Zusätzlicher Nutzenanstieg pro € Mehrkosten bei Frauen im Vergleich zu Männer
- β_5: Pauschaler "Bonus" für die Alternative "Auto".

Die Zahlenwerte ergeben nur im Vergleich, nicht absolut einen quantitativen Sinn und erschließen sich durch die "neutrale Fläche":

$$\Delta V_{ni} = (\beta_1 + \beta_2 \delta_{n1}) \Delta T_{ni} + (\beta_3 + \beta_4 G_n) \Delta K_{ni} + \beta_5 \delta_{n1}$$

Bei gegebenem festen "Kfz-Bonus" β_5 erhält man also

$$\frac{\Delta K_{ni}}{\Delta T_{ni}} = -\frac{\beta_1 + \beta_2 \delta_{n1}}{\beta_3 + \beta_4 G_n}$$

Der Zeitwert ergibt sich aus dem Quotienten $-\Delta K_{ni}/\Delta T_{ni}$, da man ja eine *Verringerung* der Reisezeit mit einem höheren Preis bezahlt (oder umgekehrt). Also

- Zeitwert ♂, Kfz: $\frac{\beta_1+\beta_2}{\beta_3} = \underline{\underline{30\,€/\mathrm{h}}}$,
- Zeitwert ♂, ÖV: $\frac{\beta_1}{\beta_3} = \underline{\underline{20\,€/\mathrm{h}}}$,
- Zeitwert ♀, Kfz: $\frac{\beta_1+\beta_2}{\beta_3+\beta_4} = \underline{\underline{15\,€/\mathrm{h}}}$,
- Zeitwert ♀, ÖV: $\frac{\beta_1}{\beta_3+\beta_4} = \underline{\underline{10\,€/\mathrm{h}}}$,

Generelle Bevorzugung des Kfz: Bezieht man die Zeiteinheiten auf den ÖV, erhält man

$$\Delta T_{\text{kfz}} = -\frac{\beta_5}{\beta_1} = \underline{\underline{0.2\,\mathrm{h} = 12\,\mathrm{min}}}$$

Bezieht man die Zeit auf die "Kfz-Zeit" (beides gibt volle Punktzahl), erhält man

$$\Delta T_{\text{kfz}} = -\frac{\beta_5}{\beta_1 + \beta_2} = \underline{\underline{0.13.3\,\mathrm{h} = 8\,\mathrm{min}}}$$

(d) *Was würde sich verändern, wenn man in der ersten Datenzeile die Zeiten von 0.6 h für das Auto und 0.8 h für den ÖV auf 0 bzw. 0.2 h setzt? Begründen Sie Ihre Antwort mit dem vom Logit-Modell prognostizierten Quotienten der Auswahlwahrscheinlichkeiten.*

Gar nichts, denn im Logit-Modell kommt es nur auf Nutzendiffeenzen an. Speziell ist der Quotienten der Wahrscheinlichkeiten zweier Alternativen i und m gegeben durch

$$\frac{P_i}{P_m} = e^{V_i - V_m}$$

was die Abhängigkeit nur von der Differenz explizit macht.