# Lösungsvorschlag zur
# Klausur zur Vorlesung Methoden der Verkehrsökonometrie
# Master-Studiengang
# WS 2015/16

## Aufgabe 1 (20 Punkte)

(a) Sektor 1: Logistik, Sektor 2: Sonstige Wirtschaft. Die sonstigen Sektoren benötigen zum Transport ihrer Produkte 5 % des gesamten Warenwertes, also $A_{12} = 0.05$. Die Logistik-Unternehmer wiederum wenden 40% ihrer Erlöse für den Fuhrpark (also vom Kfz-Sektor) und weitere 20 % aus anderen Nicht-Logistik-Sektoren für den sonstigen Betrieb auf, also $A_{21} = 0.6$. Die relativewn Eigenverbräuche kann man direkt der Aufgabenstellung entnehmen, also $A_{11} = 5\,\%$ und $A_{22} = 30\,\%$.

Insgesamt also

$$\mathbf{A} = \begin{pmatrix} 0.05 & 0.05 \\ 0.6 & 0.3 \end{pmatrix}.$$

$A_{12} = x_{12}/x_2$ gibt den wertmäßigen Anteil an, denn die sonstigen Sektoren für ihre Gesamtproduktion vom Logistik-Sektor benötigt.

(b) Koeffizientenmatrix des vollen Aufwandes:

$$\mathbf{B} = (\mathbf{1} - \mathbf{A})^{-1} = \frac{1}{\det(\mathbf{1} - \mathbf{A})} \begin{pmatrix} 1 - A_{22} & A_{12} \\ A_{21} & 1 - A_{11} \end{pmatrix} = \begin{pmatrix} 1.10 & 0.0787 \\ 0.944 & 1.496 \end{pmatrix}.$$

Hierbei wurde

$$\det(\mathbf{1} - \mathbf{A}) = (1 - A_{22})((1 - A_{11}) - A_{12}A_{21} = 0.635$$

eingesetzt.

(c) Normiert man die bisherige externe Gesamtnachfrage auf 1, ergeben die Angaben aus der Aufgabenstellung

$$\boldsymbol{y}^{\text{alt}} = \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix}.$$

Eine Verdoppelung der Nachfrage nach Logistik-Dienstleistungen bei sonst gleicher Nachfrage entspricht

$$\boldsymbol{y}^{\text{neu}} = \begin{pmatrix} 0.04 \\ 0.98 \end{pmatrix},$$

welche nun natürlich nicht mehr normiert ist. Vielmehr nimmt die Gesamtnachfrage um 2 % zu.

Die bisherige Gesamt-Wirtschaftsleistung betrug

$$x^{\text{alt}} = x_1^{\text{alt}} + x_2^{\text{alt}} = (1,1)\ \mathbf{B}\ \boldsymbol{y}^{\text{alt}} = (1,1) \begin{pmatrix} 1.10 & 0.0787 \\ 0.944 & 1.50 \end{pmatrix} \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix} = 1.58$$

und die Änderung in beiden Sektoren durch die Erhöhung der Logistik Nachfrage beträgt

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = \mathbf{B} \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \end{pmatrix} = \begin{pmatrix} 1.10 & 0.0787 \\ 0.944 & 1.496 \end{pmatrix} \begin{pmatrix} 0.02 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.0220 \\ 0.0189 \end{pmatrix}.$$

Die Erhöhung der Gesamtproduktion (Gesamtwirtschaftsleistung) ist damit

$$\Delta x = (1,1) \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = (1,1) \begin{pmatrix} 0.0220 \\ 0.0189 \end{pmatrix} = 0.0409$$

und damit die relative Erhöhung der Gesamtwirtschaftsleistung damit zu

$$\frac{\Delta x}{x^{\text{ alt}}} = 2.58\,\%$$

Sie ist also höher als die 2 % Gesamt*nachfrage*-Erhöhung.

### Häufige Fehler

- Wenn es in der Aufgabesntellung "mit Herleitung" heißt, bitte dies beachten. Ansonsten gibt es in Zukunft 0 Punkte!
- Wie schon häufig bemerkt: Bitte *spezifisch* argumentieren! Also *nicht* $A_{12}$ beschreiben als Anteil, den Sektor 2 von Sektor 1 benötigt. Wichtig ist auch das Wort "Anteil", den der (absolute) Güterfluss (in Geldeinheiten!) ist $X_{12}$, nicht $A_{12}$.

## Aufgabe 2 (45 Punkte)

(a)
- Exogen: (i) Jahreszahl $t$ bzw. $x_1 = t - 2000$: kardinalskaliert bzw. metrisch, genauer: absolutskaliert (alle Formulierungen ergeben volle Punktzahl)
- (ii) Temperatur: kardinalskaliert (metrisch)
- (iii) Wetter: nominal bzw. qualitativ (ordinal wäre auch OK, da man eine natürliche Rangfolge sonnig-bedeckt-Regen oder Regen-bedeckt-sonnig annehmen kann)
- endogen: Zahl $y$ der Räder, welche kardinalskaliert (absolutskaliert) ist.

*Hinweis:* Bei einer linearen Regressionsrechnung muss die endogene Variable immer kardinalskaliert sein. Da das geschätzte Modell Erwartungswerte der Zahlen angibt, muss die endogene Variable hingegen nicht notwendigerweise stetig sein, sondern kann auch natürlichzahlig, z.B. wie hier absolutskaliert, sein.

(b) Da die Zahl der Vorlesungen unter Umständen stark vom Wochentag und der Uhrzeit abhängt und man den entsprechenden störenden Einfluss (engl.: *confounding factors*) möglichst beseitigen will. Außerdem sollte der Erhebungszeitpunkt während einer Vorlesung liegen, was hier (während der 2. DS) erfüllt ist.

(c)
- $\beta_0$: erwartete Anzahl an Rädern im Jahr 2 000 ($x_1 = 0$) bei 0 °C ($x_2 = 0$) und nicht-regnerischem Wetter ($x_3 = 0$),

- $\beta_2$: erwarteter Anstieg mit der Zeit (Räder pro Jahr),
- $\beta_2$: Temperatursensitivität: erwarteter Anstieg der Zahl der Räder pro °C an Temperaturanstieg,
- $\beta_3$: Unterschied der mittleren Zahl an Fahrrädern bei regnerischem im Vergleich zu sonstigen Wetter.

(d) Bei allen drei Nullhypothesen ist die wahre Varianz unbekannt und es gibt $n = 8$ Datensätze bei $J + 1 = 4$ Parametern. Also ist bei (grenzwertigem) Zutreffen der jeweiligen Nullhypothese die relative Abweichung $T$ vom Rand der Nullhypothese in Einheiten der geschätzten Standardabweichung Student-t-verteilt mit $\nu = n - J - 1 = 4$ Freiheitsgraden.

(i) Keine Jahreszahlabhängigkeit, also $H_{01} : \beta_1 = 0$ mit folgender Realisierung $t_1$ von $T$:

$$t_1 = \frac{\hat{\beta}_1}{\sqrt{\hat{V}_{11}}} = 1.13.$$

Es ist ein symmetrischer Test auf 0, welcher abgelehnt ist, falls

$$|t_1| > t^{(\nu)}_{1-\alpha/2} = t^{(4)}_{0.975} = 2.78.$$

Dies ist nicht der Fall. [*nicht verlangt:* Der $p$-Wert $p_1 = 2(1 - F^{(4)}_T(|t_1|)) = 32\,\%$ ist dementsprechend größer als die Fehlerwahrscheinlichkeit $\alpha = 5\,\%$.]

(ii) Bei Regen nimmt die Räderzahl um mindestens 25 ab, also $H_{02} : \beta_3 \leq \beta_{30} = -25$ mit folgender Realisierung $t_2$ von $T$:

$$t_2 = \frac{\hat{\beta}_3 - \beta_{30}}{\sqrt{\hat{V}_{33}}} = \frac{3.3}{\sqrt{14.0}} = 0.883.$$

Es ist ein unsmmetrischer Test auf "$\leq$", welcher ablehnbar ist, falls

$$t_2 > t^{(\nu)}_{1-\alpha} = t^{(4)}_{0.95} = 2.13.$$

Dies ist nicht der Fall. [*nicht verlangt:* Der $p$-Wert $p_2 = 1 - F^{(4)}_T(t_2) = 21.3\,\%$ ist dementsprechend größer als $\alpha$.]

(iii) Der Übersichtlichkeit halber vergleichen wir zwei positive Größen: 1. Unterschied $-\beta_3$ der Nutzerzahl zwischen Nicht-Regen und Regen, 2. Erhöhung $10\beta_2$ der Nutzerzahl bei 10 Grad Temperaturanstieg. Also gilt, wenn der Unterschied Regen-Nicht Regen die Nutzerzahl stärker beeinflussen soll als 10 Grad Temperaturunterschied folgende Nullhypothese $H_{03} : |\beta_3| > 10|\beta_2|$ bzw. bei Beachtung der sinnvollen Vorzeichen: $H_{03} : -\beta_3 > 10\beta_2$ oder äquivalent

$$H_{03} : U = 10\beta_2 + \beta_3 < 0.$$

Damit

$$\begin{aligned} \hat{U} &= 10\hat{\beta}_2 + \hat{\beta}_3 = -9.33, \\ \sqrt{V(\hat{U})} &= \sqrt{100V_{22} + V_{33} + 2 * 10 * 1 * V_{23}} = 3.38, \\ t_3 &= \frac{\hat{U}}{\sqrt{V(\hat{U})}} = -2.75. \end{aligned}$$

Es ist ein unsmmetrischer Test auf "<", welcher völlig gleichwertig zu einem auf '$\leq$' ist und welcher ablehnbar ist, falls

$$t_3 > t_{1-\alpha}^{(\nu)} = t_{0.95}^{(4)} = 2.13.$$

Dies ist nicht der Fall. [*nicht verlangt:* Der $p$-Wert $p_3 = 1 - F_T^{(4)}(t_3) = 97.4\,\%$. Damit hätte die *gegenteilige* Nullhypothese $\bar{H}_{03} : U \geq 0$ einen $p$-Wert von $2.6\,\%$ und wäre damit bei $\alpha = 5\,\%$ abzulehnen: Damit beeinflusst in der Tat Regen die Radnutzerzahl signifikant stärker als 10 Grad Temperaturunterschied!]

(e) Erweitertes Modell:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon, \quad \epsilon \sim \text{ i.i.d. } N(0, \sigma^2),$$

wobei $x_1$ und $x_2$ wie bisher definiert sind, und die neuen Wetter-Dummyvariablen gegeben sind durch

$$x_3 = \begin{cases} 1 & \phantom{x} \\ 0 & \text{sonst,} \end{cases} \quad , \quad x_4 = \begin{cases} 1 & \phantom{x} \\ 0 & \text{sonst,} \end{cases}$$

Dementsprechend bezeichnet nun $\beta_3$ den erwarteten Unterschied der Räderanzahl bei Regen im Vergleich zu sonnigen Wetter, und $\beta_4$ den erwarteten Unterschied bei bedeckten im Vergleich zu sonnigen Wetter.

(f) Es könnten sich auch die Zahl der Studenten in den Vorlesungen ändern, sodass eine Änderung der absoluten Radanzahl nicht notwendigerweise mit einer Verschiebung des Modal Split gleichzusetzten ist. Merke: signifikante absolute Veränderungen implizieren nicht notwendigerweise auch signifikante relative Änderungen, wenn sich auch der Bezug, d.h. der Nenner, ändern kann!

### Häufige Fehler

- Teil (f): Für Temperatur und Wetter gibt es die exogenen Variablen $x_2$ bzw. $x_3$. Damit sind Effekte wie Klimaerwärmung oder mehr Schönwettertage in den letzten jahren automatisch berücksichtigt
- Teil (b), (c) und allgemein: **So spezifisch wie möglich!**, also nicht einfach sagen "zeitabhängige Faktoren werden vernachlässigt" sondern explizit auf die zeitabhängigen Studentenzahlen hinweisen. Temperatur gilt sowieso nicht, da diese extra berücksichtigt wurde, siehe auch Teil (f). Bei (c): Nicht einfach $\beta_3$ als "Einfluss vom Wetter" erklären, sondern Einfluss welchen Wetters (hier Regen gegenüber Nicht-Regen)
- Teil (b) und allgemein: Nichts reininterpretieren. Es steht nirgendwo was von Panel-Daten, was viele erwähnt hatten. Bei der Zählung von Fahrrädern und über die jahre wechselnden Studenten wäre dies sowieso sinnlos.
- Teil (d). Wieder der Klassiker: Signifikanztests kann man nur ablehnen, nicht aber annehmen!
- Teil (e): Wieder genau spezifizieren, nicht einfach nur "$\beta_3 x_3 + \beta_4 x_4$" hinschreiben, sondern $x_3$ und $x_4$ spezifizieren!

## Aufgabe 3 (55 Punkte)

(a) Nach Aufgabenstellung gab es 10 männliche und 10 weibliche Test-Teilnehmer. Das erste, dritte und fünfte Choice Set ($n = 1, 3, 5$) beantworteten also nur die jeweils 10 Männer ($y_n = y_{n1} + y_{n2}$ jeweils =10) und die geradzahligen Choice Sets nur Frauen. Damit bekam jeder Teilnehmer 3 Choice Sets vorgelegt. Das ergibt sich auch aus $N = \sum_n \sum_i y_{ni} = 60$ Einzelentscheidungen, geteilt durch 10+10 Befragte.

(b) Die Referenzalternative für die AC $\delta_{i1}$ ist $i = 2$, also die zweite (große) Stadt. Ohne die Erwähnung unterschiedlich großer Städte wäre die Information über die Qualität der Alternativen *vollständig* durch die Attribute Entfernung und Miete gegeben. Eine alternativenspezifische Konstante, die sonst unbeobachtete systematische Einflüsse umfasst, und die bei qualitativ unterschiedlichen Alternativen (wie MIV und ÖV) *immer* gegeben ist, ergäbe dann logisch *keinen* Sinn: Die Reihenfolge $i = 1, 2$ der Städte ist austauschbar, wenn man die Miet- und Entfernungsattribute mittauscht, also $\beta_4 = 0$.[1]

(c) $\beta_1$: generisch formulierte Entfernungssensitivität: Zunahme der Attraktivität mit der Entfernung in NE/km. Beiderlei Vorzeichen sind denkbar: Negativ, wenn die Studis tendenziell "Muttersöhnchen" sind, positiv, wenn sie in ihrer Mehrheit "die große, weite Welt" kennenlernen wollen. (Jede begründete Aussage gibt volle Punktzahl.)

$\beta_2$: generisch formulierte Mietpreissensitivität: Zunahme der Attraktivität mit der Miete in NE/Euro pro Monat: eindeutig negativ.

$\beta_3$: Das Maß (in NE), mit dem Frauen die mittlere Stadt stärker gegenüber der großen Stadt bevorzugen (oder weniger ablehnen) als Männer. Kein offensichtliches Vorzeichen

$\beta_4$: Das Maß (in NE), mit welcher Männer die mittlere gegenüber der großen Stadt bevorzugen. Negativ, wenn große Städte tendenziell bevorzugt werden (wieder gibt jede sinnvoll erläuterte Behauptung volle Punktzahl).

(d) Vier realisierte Merkmalssummen $M_j^{\text{data}}$, welche beim kalirierten Modell mit den vier modellierten Merkmalssummen $M_j^{\text{mod}}$ übereinstimmen müssen.

(i) Realisierte Gesamtentfernung der gewählten Städte bei den 60 Entscheidungen: $M_1^{\text{data}} = \sum_{n=1}^{6} \sum_{i=1}^{2} R_{ni} y_{ni} = 13\,000\,\text{km}$,

(ii) realisierte Gesamtmiete in den gewählten Städten: $M_2^{\text{data}} = \sum_{n=1}^{6} \sum_{i=1}^{2} M_{ni} y_{ni} = 23\,900\,\text{Euro}$,

(iii) Gesamtzahl an Entscheidungen für Stadt 1 von Frauen: $M_3^{\text{data}} = y_{21} + y_{41} + y_{61} = 17$,

(iii) Gesamtzahl an Entscheidungen für Stadt 1 von allen Teilnehmern: $M_4^{\text{data}} = \sum_{n=1}^{6} y_{n1} = 37$.

[1] Wohlgemerkt nur, wenn es bei den beiden Städten keine weitere qualifizierenden Informationen gibt, wie Lage, Größe etc der beiden Städte.

Das $\boldsymbol{\beta} = 0$ nicht den geschätzten Parametern entsprechen kann, sieht man am besten bei den "einfachen" Merkmalssummen, z.B. $M_4$: Da bei $\boldsymbol{\beta} = 0$ die modellierten Entscheidungswahrscheinlichkeiten $P_{ni}$ alle=1/2 sind, gilt

$$M_4^{\text{mod}} = \sum_{n=1}^{6} y_n P_{n1} = \frac{6 * 10}{2} = 30 \neq M_4^{\text{data}} \quad \Rightarrow \quad \text{Modell für } \boldsymbol{\beta} = 0 \text{ nicht kalibriert.}$$

(e) (i) Männliche Teilnehmer: Der Mietpreisunterschied $\Delta M_n = M_{n1} - M_{n2}$ kompensiert gerade die alternativenspezifische Konstante:

$$\beta_2 \Delta M_n + \beta_4 = 0 \ \Rightarrow \ \Delta M_n = -\frac{\beta_4}{\beta_2} = 50\,\text{Euro}$$

(ii) Bei weiblichen Teilnehmern kommt der frauenspezifische Term $g_n \delta_{i1}$ hinzu:

$$\beta_2 \Delta M_n + \beta_3 + \beta_4 = 0 \ \Rightarrow \ \Delta M_n = -\frac{\beta_3 + \beta_4}{\beta_2} = 110\,\text{Euro.}$$

*Anmerkung*: Da $\Delta M_n$ als Mietpreisunterschied der mittleren zur große Stadt definiert wurde, würden in diesem Datensatz also sogar *mittlere* Städte um das Geldäquivalent 50 bzw. 110 Euro bevorzugt. Dies ist aber vermutlich eher unrealistisch. Da dies etwas missverständlich formuliert wurde, ergibt jedes Vorzeichen (also ± 50 Euro bzw. ± 110 Euro) die volle Punktzahl.

(f) Die deterministischen Nutzenfunktionen des ersten Choice Sets lauten

$$V_{11} = 100\beta_1 + 400\beta_2 + \beta_4 = -2.96,$$
$$V_{12} = 100\beta_1 + 700\beta_2 = -5.76$$

Damit der Nenner $\mathcal{N}$ der Logit-Formel:

$$\mathcal{N} = e^{V_{11}} + e^{V_{12}} = 0.05497$$

und

$$P_{12} = \frac{e^{V_{12}}}{\mathcal{N}} = 5.7\,\%.$$

(g) Mikroskopische Preis-Eigenelastizität der Stadt 1 bezüglich Verschiebungen (zur Stadt 2) beim letzten Choice Set $n = 6$ gemäß Skript oder Aufzeichnungen:

$$\epsilon_{611}^{(M)} = \beta_2 M_{61}(1 - P_{61}).$$

Mit $\beta_2 = -0.008$, $M_{61} = 300\,\text{Euro}$ und $P_{61} = 0.44$ aus der Aufgabenstellung ergibt dies

$$\epsilon_{611}^{(M)} = -1.344$$

Die Verschiebungs-Preiselastizität ist also überproportional: ein 1 % höherer Mietspiegel in Stadt 1 führte für Personen gemäß der Gruppe 6 zu einer erwarteten Abwanderung von 1.34 % in die andere Stadt.

**Häufige Fehler**

- Teil (b): Um zu argumentieren, ob eine AC sinnvoll ist oder nicht, darf man nicht mit unterschiedlichen Entfernungen oder Mietpreisen kommen, da diese ja explizit modelliert wird!

- Teil (c): Wieder wird häufig zu ungenau argumentiert: $\beta_3$ ist das Maß in NE, mit welcher Frauen die mittlere Stadt stärker gegenüber der großen Stadt bevorzugen (oder weniger ablehnen) als Männer. Nicht einfach "geschlechtsabhängige Bevorzugung" usw. Zur genauen Spezifikation ist hier jeder einzelne der Begriffe "Nutzeneinheiten (NE)" (nicht etwa Euro), Frauen gegenüber Männer (nicht etwa Männer gegenüber Frauen oder Bevorzugung bei Frauen allgemein, was beides falsch wäre) sowie mittlere gegenüber der großen Stadt (nicht umgekehrt) notwendig. Quizfrage: Ist die Spezifikation $\beta_3$: "Das Maß (in NE), mit dem Männer die mittlere Stadt stärker gegenüber der großen Stadt ablehnen (oder weniger bevorzugen) als Frauen" korrekt? gilt dies auch für $\beta_3$: "Das Maß (in NE), mit dem Männer die große Stadt stärker gegenüber der mittleren Stadt bevorzugen (oder weniger ablehnen) als Frauen" oder für $\beta_3$: "Das Maß (in NE), mit dem Frauen die mittlere Stadt stärker gegenüber der großen Stadt ablehnen (oder weniger bevorzugen) als Männer"?

- Teil (f): Obwohl die Alternative $i = 2$ die Referenzalternative ist, gilt nicht $V_{n2} = 0$, denn es gibt ja hier auch generische Variablen (Entfernung und Mietpreis), die bei beiden Alternativen ungleich null sind. Die Referenz bezieht sich *nur* auf die AC und ggf auf weitere alternativenspezifisch formulierte Variablen, z.B. sozioökonomische Variablen.