

Technische Universität Dresden
Institut für Wirtschaft und Verkehr
Lehrstuhl für Ökonometrie und Statistik, insbes. im Verkehrswesen

Skript zur Vorlesung

Verkehrsökonomie

Methoden und Modelle

Dr. Martin Treiber

Wintersemester 2022/23

© 2008-2023 Martin Treiber.

Inhaltsverzeichnis

1	Allgemeines	1
1.1	Ablauf einer ökonomischen Untersuchung	2
1.2	Modelle und Variablen der Verkehrsökonomie	3
1.2.1	Endogene Variablen	3
1.2.2	Exogene Variablen	4
1.2.3	Modellparameter	5
1.2.4	Zufallsanteile	5
1.2.5	Modellfunktionen	6
1.2.6	Das Prinzip der Sparsamkeit	6
1.3	Modellgleichungen	6
1.3.1	Inhaltliche Struktur	7
1.3.2	Mathematische Struktur	8
1.4	Ein Anwendungsbeispiel: Verkehrsmittelwahl	12
1.4.1	Modellstruktur	13
1.4.2	Modellgleichungen und Parameter	14
1.4.3	Systemgleichungen	15
1.5	Verwendete Symbole	18
2	Stetige ökonomische Modelle	19
2.1	Struktur der Gleichungen	19
2.2	Vorgehen bei der ökonomischen Untersuchung	21
2.2.1	Ablauf bei vorgegebenem Modell	21
2.2.2	Modellentwicklung und Validierung	23
2.3	Modellspezifikation	24
2.3.1	Funktionale Spezifikation	25
2.3.2	Statistische Spezifikation	28
2.3.3	Datenspezifikation	31
2.3.4	Zusammenfassung	33
2.4	Parameterschätzung	33
2.4.1	Methode der kleinsten Fehlerquadrate	35
2.4.2	Parameterschätzung I: Formulierung mit Summen	37
2.4.3	Additionsregel, Residualvarianz und Bestimmtheitsmaß	39
2.4.4	Spezialfälle: Eine und zwei exogene Variable	40
2.4.5	Parameterschätzung II: Formulierung mit Vektoren und Matrizen	41
2.4.6	Anschauliche Interpretation des linearen Modells	44
2.5	Statistische Eigenschaften der OLS-Schätzer quasilinearer Modelle	47
2.5.1	Erwartungswert	48

Inhaltsverzeichnis

2.5.2	Varianzen und Kovarianzen der OLS-Schätzer	48
2.5.3	Verteilungs- und Dichtefunktionen der OLS-Schätzer	51
2.5.4	Niedrigdimensionale Spezialfälle	54
2.6	Gewinnung einer statistisch belastbaren Modellaussage	56
2.6.1	Schätzer für die Residualvarianz	57
2.6.2	Konfidenzintervalle	58
2.6.3	Extrapolation und Prognose	61
2.6.4	Statistische Tests eines Parameters	62
2.6.5	p-Werte	68
2.6.6	Gütekfunktion und Fehler zweiter Art	71
2.6.7	Konfidenzregionen von Parameterkombinationen	75
2.6.8	Simultane Tests verbundener Nullhypothesen: F -Test	77
2.6.9	Test auf Relevanz von Einflussfaktoren	81
2.7	Nichtmetrische exogene Variable	83
2.8	Logistische Regression	86
2.9	Bayes'sche Inferenz: Was sagt der p-Wert wirklich aus?	90
2.9.1	Problemstellung	90
2.9.2	Gaußverteilte a-priori-Werte	93
2.10	Einige Herleitungen	102
2.10.1	Statistischen Eigenschaften der Einfachregression "zu Fuß"	102
2.10.2	Additionsregel der Varianzen und Bestimmtheitsmaß	105
2.10.3	Fehlerquadratsumme und Schätzer der Residualvarianz	106
2.11	Rechnen mit Vektoren und Matrizen	108
2.11.1	Wichtige Definitionen	108
2.11.2	Additionen und Multiplikationen	110
2.11.3	Wichtige Matrix-Rechenregeln	110
2.11.4	Multivariate Normalverteilung	111
2.12	Verwendete Symbole	112
3	Datengewinnung: Verkehrserhebungen	114
3.1	Ablauf einer Erhebung	114
3.2	Erhebungsdesign	116
3.2.1	Aggregierungsebene	116
3.2.2	Abgrenzung bzw. Zeit- und Merkmalsträgerdimension	116
3.2.3	Ausmaß der Kontrolle über den Untersuchungsgegenstand	117
3.2.4	Ziehungsmethode	118
3.2.5	Modalität der Erhebung	119
3.3	Beispiel einer objektiven Messung: Verkehrsflussdaten	120
3.3.1	Mesung	120
3.3.2	Tagesganglinien und Bemessungsverkehrsstärke	121
3.4	Beispiel zu einer Revealed-Choice Erhebung: SrV	122
3.5	Direkte Nutzermessung: Experimentelles Design	127
3.6	Konkretes Anwendungsbeispiel: Wahlbasierte Conjoint-Analyse	128
3.6.1	Allgemeines	128

Inhaltsverzeichnis

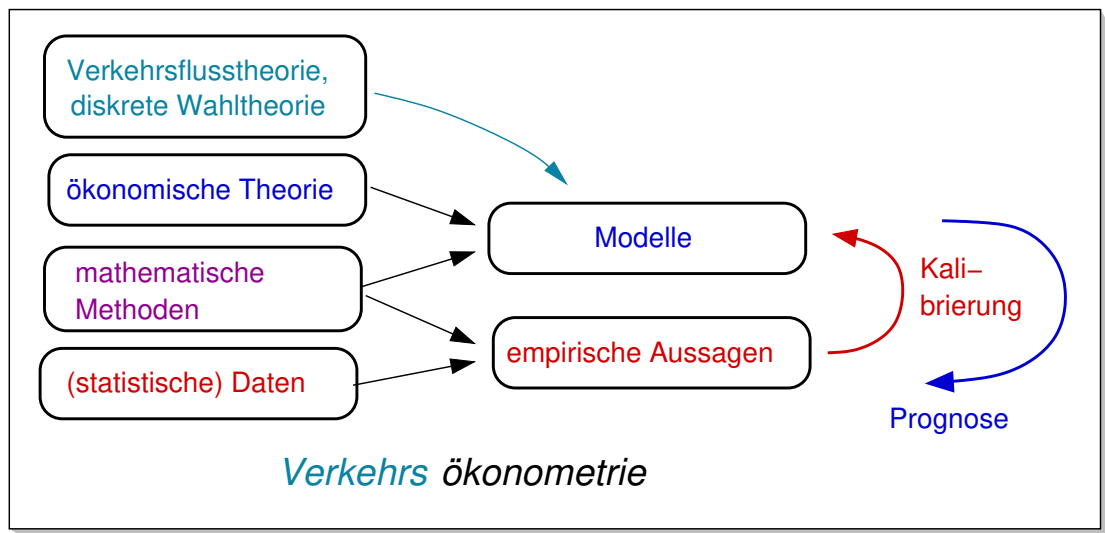
3.6.2	Beispiel: Studentenbefragung	128
3.6.3	Komplexeres Beispiel aus der Literatur: Modal-Split in der Schweiz	134
3.6.4	Experimentelles Design der Choice Sets	135
4	Modelle diskreter Entscheidungen	138
4.1	Einführung	138
4.2	Mathematische Beschreibung	140
4.2.1	Alternativen	142
4.2.2	Beeinflussungsfaktoren	144
4.2.3	Deterministische Nutzenfunktion	145
4.2.4	Zufallsnutzen	150
4.2.5	Modellierung der Wahlentscheidung	153
4.3	Probit-Modelle	155
4.3.1	Binomiales Probit-Modell	155
4.3.2	Multinomiales i.i.d Probit-Modell	155
4.4	Logit-Modelle	157
4.4.1	Binomiales Logit Modell	158
4.4.2	Multinomiales Logit Modell	158
4.4.3	Interpretation der Parameter bei Probit- und Logitmodellen	162
4.5	Elastizitäten	163
4.5.1	Definitionen im Rahmen der diskreten Wahltheorie	164
4.5.2	Mikroskopische Preiselastizitäten	166
4.5.3	Makroskopische Preiselastizitäten	168
4.5.4	Exkurs: Revenue Management	171
4.6	Einschub: Die Maximum-Likelihood-Methode	171
4.6.1	Regressionsmodelle: Vergleich mit der LSE-Methode	172
4.6.2	Anwendung auf das einfachste Wahlmodell	174
4.7	Parameterschätzung mit der Maximum-Likelihood-Methode	175
4.7.1	Maximum-Likelihood Funktionen	176
4.7.2	Log-Likelihood des allgemeinen Entscheidungsmodells	177
4.7.3	Grafische Lösung	177
4.7.4	Kalibrierungsbedingungen für das allgemeine Entscheidungsmodell	183
4.7.5	Schätzung des binomialen Probit-Modells mit quasilinearen Nutzen	184
4.7.6	Schätzen der Logit-Modelle mit quasilinearen Nutzen	185
4.7.7	Schätzung des Konstantenmodells	187
4.8	Numerische Lösung der Schätzgleichungen mit der Newton-Methode . . .	189
4.8.1	Beispiel: Revealed-Choice-Befragung mit zwei Alternativen	191
4.8.2	Beispiel 2: Revealed-Choice-Befragung mit vier Alternativen	193
4.8.3	Beispiel 3: Stated Choice Conjoint-Analyse	196
4.8.4	Diskussion	199
4.9	Statistische Eigenschaften der ML-Schätzer	200
4.9.1	Kovarianzmatrix der Parameterschätzer	200
4.9.2	Konfidenzintervalle und Tests	202
4.9.3	Spezialfall zweier Modellparameter	203

Inhaltsverzeichnis

4.10	Parameter-nichtlineare Nutzenfunktionen	203
4.10.1	Prinzip-Beispiel	203
4.10.2	Reales Beispiel: Wahl der Mensa bzw. Imbissbude	209
4.11	Modellvergleich: Likelihood-Ratio-Test	215
4.11.1	Vorgehen	216
4.11.2	Test auf Relevanz der Einflussfaktoren mit dem Top-Down-Ansatz	220
4.12	Modellqualität: Goodness-of-Fit Maße	222
4.13	Generalized-Extreme-Value (GEV) Modelle	226
4.13.1	Motivation	226
4.13.2	Allgemeine Formulierung der Modellklasse	227
4.13.3	Nested-Logit-Modell	229
4.13.4	Weitere GEV-Modelle	239
4.14	Mixed-Logit Modell	240
4.14.1	Paneldaten	241
4.14.2	Parameterschätzung des Mixed-Logit-Modells	242
4.14.3	Spezialfall: Schätzung bei einer Panel-Struktur	243
4.15	Charakterisierung der Zuverlässigkeit	244
4.15.1	Annahmen und Modell	244
4.15.2	Zwei Spezialfälle	246
4.15.3	Beispiel: exponentialverteilte Verspätungen	248
4.16	Einige Herleitungen	249
4.17	Verwendete Symbole	258
5	Ökonometrische Verflechtungsmodelle	260
5.1	Allgemeines	260
5.2	Formulierung des Verflechtungsmodells	260
5.2.1	Kompakte Formulierung in Vektor-Matrix-Notation	264
5.3	Lösung des Verflechtungsmodells	265
5.3.1	Veranschaulichung der Koeffizientenmatrix des vollen Aufwandes	265
5.3.2	Explizite Lösungen für zwei und drei Sektoren	267
5.4	Lebenszyklusanalyse und Einbindung des Input-Output-Modells	269
5.4.1	Lebenszyklusanalyse (LCA)	270
5.4.2	Econometric Input-Output LCA (EIO-LCA)	272
5.4.3	Emissionen bei direkter Verwendung des IOM (IOM-LCA)	274
5.4.4	Konsistenz der Methode bei Preissteigerungen	274
5.5	Dynamische Verflechtungsmodelle	276
5.5.1	Spezialfall: Dynamisches Ein-Sektor-Modell	278
5.5.2	Varianten dynamischer Verflechtungsmodelle	278
5.5.3	Produktionsketten (Supply Chains)	281

1 Allgemeines

Das Wort **Ökonometrie** hat seinen Ursprung von der Ökonomie (Wirtschaft) und dem lateinischen Wort *metiri*, welches "messen" bedeutet. Damit beinhaltet Ökonometrie nicht nur *quantitative* wirtschaftliche Theorien und Konzepte, sondern vor allem deren **empirische** (also auf Messung, Beobachtung und Erfahrung beruhende) *Überprüfung*. Bei den durch Messung und Beobachtung gewonnenen Daten handelt es sich fast immer um **statistische Daten**, also um eine Beschreibung von *Massenphänomenen*. Damit spielt neben der quantitativen Formulierung wirtschaftlicher Theorien durch **mathematische Modelle** auch die Auswertung der Daten durch **statistische Modelle** eine Rolle. Das Zusammenspiel der verschiedenen Elemente der Ökonometrie wird in folgendem Flussdiagramm deutlich.



In der Verkehrsökonomie werden diese allgemeinen Konzepte auf den Verkehrssektor spezialisiert:

Die **Verkehrsökonomie** umfasst die Gesamtheit mathematischer Modelle und statistischer Verfahren, um auf einer empirischen Grundlage den Verkehr und seine volkswirtschaftlichen Auswirkungen quantitativ zu analysieren und zu prognostizieren.

1 Allgemeines

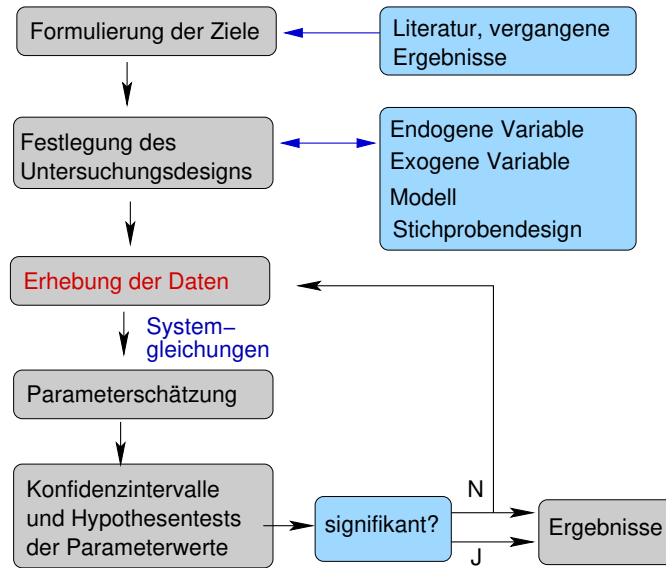


Abbildung 1.1: Allgemeiner Ablauf einer ökonometrischen Untersuchung.

1.1 Ablauf einer ökonometrischen Untersuchung

Im Prinzip läuft eine ökonometrische Untersuchung nach dem in Abb. 1.1 dargestellten Schema ab. Nach Formulierung des Untersuchungsziels und dem Abgleich mit vergangenen Untersuchungen kommt die wichtige Phase der Festlegung des Erhebungsdesigns und des Modells. Diese beiden Komponenten müssen zusammenpassen. Will man beispielsweise ein Logit-Modell anwenden, muss der Erhebungsfragebogen derart gestaltet sein, dass er eine endliche Anzahl von Alternativen enthält, welche (i) *alle* möglichen Antworten enthalten (Vollständigkeit), (ii) jeweils nur eine wählbar ist (Exklusivität) und (iii) wesentlich voneinander verschieden sind (sonst müsste man ein verallgemeinertes Entscheidungsmodell anwenden). Ferner müssen die Anzahl der Variablen und ihre Skalierung (siehe unten) zur Modellspezifikation kompatibel sein.

Nach der dann folgenden Durchführung der Datenerhebung (Abschnitt 3) werden die Modellparameter anhand der Daten geschätzt, also ihre wahrscheinlichsten Werte und ihr Streubereich festgestellt. Die eigentliche Erkenntnis der ökonometrischen Untersuchung ergibt sich neben der deskriptiven Auswertung der Erhebung vor allem anhand der Schätzung der Modellparameter. Damit lassen sich unter Anderem wichtige von unwichtigen Einflussfaktoren trennen sowie Maßzahlen verkehrswirtschaftlicher Sachverhalte gewinnen wie eine Bevorzugung bestimmter Verkehrsmittel ("bei gleichem Zeit- und Kostenaufwand fahre ich lieber mit dem ÖV als dem Auto") oder implizite Zeitwerte ("eine Stunde weniger Reisezeit ist mir 20 Euro wert").

1 Allgemeines

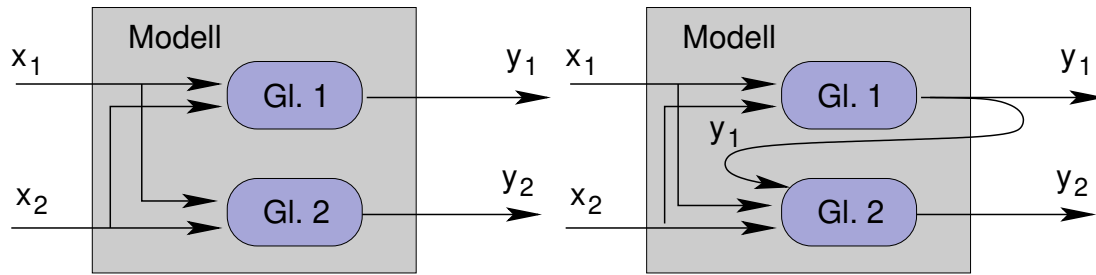


Abbildung 1.2: Flussdiagramm der exogenen und endogenen Variablen des allgemeinen ökonometrischen Modells (1.1). Gezeigt ist der Fall für je zwei exogene und endogene Variablen, $M = k = 2$ ohne (links) und mit (rechts) Kopplung der endogenen Variablen. Das Modell (eckige Box) besteht hier also aus zwei Gleichungen, welche zwei Eingleichungsmodellen (links) bzw. einem Mehrgleichungsmodell (rechts) entsprechen.

1.2 Modelle und Variablen der Verkehrsökonomie

Da die Ökonometrie wirtschaftliche Zusammenhänge quantitativ beschreibt, besteht ihre Grundlage aus mathematischen Gleichungen. Generell kann man ökonometrische Modelle (und mathematische Modelle allgemein) als eine mathematische Abbildung betrachten, die gewisse Eingabegrößen auf Ausgabegrößen abbildet, wobei die Abbildung an den konkreten Sachverhalt durch Modellparameter angepasst wird (Abb. 1.3). Wir werden im Folgenden Gleichungen folgender Struktur betrachten (vgl. Abb. 1.2):

$$Y_k = f_k(x_1, \dots, x_m, \dots, x_M, \beta_0, \dots, \beta_j, \dots, \beta_J) + \epsilon_k = f_k(\mathbf{x}, \boldsymbol{\beta}) + \epsilon_k. \quad (1.1)$$

1.2.1 Endogene Variablen

Die Y_k sind die **erklärten Variablen**. Eine solche Variable wird bisweilen auch als **Explanandum** (lat. "zu erklärende" bzw. erklärte Variable), als **endogene Variable** (Griechisch: von innen heraus, also aus dem Modell kommend) oder gemäß allgemeinen mathematischen Sprachgebrauch als **abhängige Variable** bezeichnet. In der Systemtheorie entsprechen die Y_k dem *Output* des jeweiligen Modells. Da viele Modelle stochastischer Natur sind, also Zufallsgrößen ϵ_k enthalten (vgl. Abschnitt 1.2.4), sind die Y_k im Allgemeinen Zufallsvariablen und werden, der allgemeinen Konvention der Statistik entsprechend, "groß" geschrieben.

Je nach der Anzahl und Kopplung der endogenen Variablen werden verschiedene Modellklassen unterschieden: von

- **Eingleichungsmodelle:** Nur eine endogene Variable $Y_1 = Y$,
- **Mehrgleichungsmodelle:** Mehrere endogene Variable Y_k , die miteinander gekoppelt sind (vgl. Abb. 1.2 rechts),

1 Allgemeines

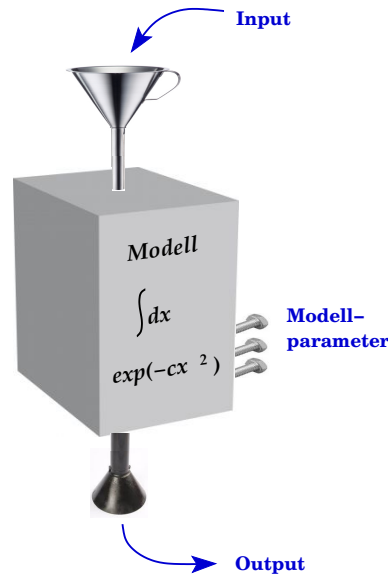


Abbildung 1.3: "Flussdiagramm" eines mathematischen Modells in einer anderen Perspektive: Oben kommen die exogenen Variablen hinein, unten die endogenen heraus. Die Box stellt das eigentliche Modell dar und die "Stellschrauben" die Modellparameter.

- **Mehreren Eingleichungsmodelle:** Mehrere ungekoppelte endogene Variable wie im Flussdiagramm 1.2 links. Dies entspricht mehreren separat zu lösenden Eingleichungsmodellen.

Beispiel Verkehrsmittelwahl: Die Y_k sind die absoluten Häufigkeiten der gewählten Verkehrsmittel, z.B. Y_1 =Zahl der Kfz-Fahrten, Y_2 =Zahl der ÖPNV-Fahrten und Y_3, Y_4 = Zahl der Wege mit dem Rad bzw. zu Fuß.

1.2.2 Exogene Variablen

Die Größen x_m , $m = 1, \dots, M$ sind die **erklärenden Variablen**, welche auch als **Explanans** (lat. "erklärend"), **exogene Variable** (Griechisch: von außen kommend) oder allgemein als **unabhängige Variable** bezeichnet werden. In der Systemtheorie entsprechen die x_m dem *Input* in das jeweilige Modell. In der Regel werden die x_m als deterministisch betrachtet und deshalb, der allgemeinen Konvention entsprechend, "klein" geschrieben.¹

Je nach Anzahl der exogenen Variablen definiert man

¹In der Praxis werden sowohl die x_m als auch die Y_k aus Erhebungen gewonnen, beide Variablenkategorien sind damit prinzipiell stochastischer Natur. Man kann aber ohne Einschränkung der Allgemeinheit die Stochastizität der Input-Größen auf die Zufallsanteile ϵ_k verschieben, da die Modelle ja stochastische *Funktionen* darstellen, also die exogenen Variablen sowieso variabel sind. Dies ist sogar geboten, da andernfalls die Stochastizität überbestimmt ist. Näheres in Abschnitt 1.2.4.

1 Allgemeines

- **univariate Modelle:** $M = 1$,
- **multivariate Modelle**²: $M > 1$.

Beispiel Verkehrsmittelwahl: Die x_m sind die Reisezeiten und andere, die Verkehrsmittelwahl beeinflussende Faktoren (Kosten, Zuverlässigkeit etc)

1.2.3 Modellparameter

Die Größen β_j , $j = 0, \dots, J$ sind die **Modellparameter**. Im Gegensatz zu den exogenen und endogenen Variablen, welche sich bei jedem System bzw. bei jeder Anwendung des Modells ändern, sollten die Modellparameter nach ihrer **Schätzung** bzw. **Modellkalibrierung** und einer solchermaßen bewirkten Anpassung des Modellverhaltens an die Wirklichkeit (vgl. Abschnitt 2.4) für alle Anwendungen einen konstanten Wert besitzen. *Diese Eigenschaft ist ein entscheidendes Kriterium für die Qualität eines Modells und seiner Aussage- und Prognosekraft!*

Beispiel Verkehrsmittelwahl: Modellparameter bestimmen die relative Gewichtung einzelner Einflussfaktoren (z.B. für jedes Verkehrsmittel den Wert der Zeit in €/h) sowie eine a-Priori-Bevorzugung bestimmter Verkehrsmittel gegenüber anderen (in € oder Minuten) bei eigentlich gleichem Nutzwert.

1.2.4 Zufallsanteile

Die ϵ_k beschreiben unbestimmte oder **nicht erklärte Anteile** welche meist durch Zufallsvariablen mit Erwartungswert $E(\epsilon_k) = 0$ modelliert werden³ Insbesondere gibt es folgende Gründe für die Notwendigkeit eines Zufallsanteils ϵ_k :

- Das Modell kann nicht alle Einflussfaktoren berücksichtigen. Wichtig ist aber, zumindest alle Faktoren mit *systematischen Einfluss* zu berücksichtigen, da das Modell sonst **fehlspezifiziert** ist und falsche Aussagen liefert (vgl. Abschnitt 2.3).
- Die zur Modellkalibrierung verwendeten Messwerte sind fehlerbehaftet. Dieser Anteil von ϵ gibt direkt die kumulierten Messfehler wider.
- Der Mensch ist keine Maschine. Der entsprechende Anteil von ϵ spiegelt die Abweichung des in Wirklichkeit oft nichtrationalen Verhaltens vom Idealbild des *Homo Oeconomicus* wider.

Korrekt spezifizierte und eindeutig definierte Zufallsterme spielen eine wesentliche Rolle bei der Parameterschätzung (Abschnitt 2.4)

²Streng genommen sind die Termini Einfach- und Mehrfachregression in diesem Zusammenhang nicht korrekt, da sie eine spezielle Schätzmethode (die Regression) und nicht die Modelle als solches bezeichnen. Dies wird aber in der Literatur häufig ungenauerweise gleichgesetzt.

³Man sagt, dass "Zufallselemente nichts anderes als das Eingeständnis von Unwissen" seien. Die Bedingung $E(\epsilon_k) = 0$ ist keine Einschränkung, da man einen eventuellen Erwartungswert auf den deterministischen Teil verlagern kann.

1 Allgemeines

Beispiel Verkehrsmittelwahl: ϵ_k entspricht dem *Zufallsnutzen*. Dieser sorgt dafür, dass – wie in der Wirklichkeit – mit geringerer Wahrscheinlichkeit auch ein Verkehrsmittel mit nichtmaximalen deterministischen, d.h. modellierten Nutzen gewählt wird.

1.2.5 Modellfunktionen

Die Funktionen $f_k(\dots)$ charakterisieren schließlich das eigentliche ökonometrische Modell. Wie bei den Parametern ist es für die Güte und Prognosekraft eines Modells wichtig, dass es nach seiner Entwicklung unverändert auf neue Instanzen des zu beschreibenden Sachverhalts mit unterschiedlichen exogenen Variablen angewandt werden kann.

Beispiel Verkehrsmittelwahl bei N Entscheidungen insgesamt:

$$P_k = \text{Prob}(\text{Altern } k \text{ gewählt}) = \frac{e^{V_k(\mathbf{x}, \boldsymbol{\beta})}}{\sum_{k'} e^{V_{k'}(\mathbf{x}, \boldsymbol{\beta})}}, \quad (1.2)$$

$$Y_k \sim \text{Multinom}(N, \mathbf{P}) \text{ - Verteilung.} \quad (1.3)$$

Dies ist ein verkettetes Modell (s. weiter unten) mit 2 Stufen:

- 1. Stufe: Deterministisches Modell (1.2) für die Auswahlwahrscheinlichkeiten
- 2. Stufe: Stochastisches Modell (1.3), welches die Zufallsvariablen “Zahl der Entscheidungen für Alternative k ” mit Hilfe der Multinomialverteilung modelliert.

Die Nutzenfunktionen $V_k(\mathbf{x}, \boldsymbol{\beta})$ enthalten die Einflussfaktoren (exogenen Variablen) \mathbf{x} (Reisezeit, Kosten,...) und die Modellparameter $\boldsymbol{\beta} = \{\beta_j\}$ (relative Gewichtungen und Bevorzugen), während die zweite Stufe keine weiteren exogenen Variablen oder Parameter enthält. Näheres dazu im Abschnitt 4.

1.2.6 Das Prinzip der Sparsamkeit

Bei der Modellformulierung gilt das Sparsamkeitsprinzip (*principle of parsimony*), welches man, frei nach Einstein, folgendermaßen formulieren kann:

Das Modell sollte so einfach wie möglich sein, aber nicht einfacher. Es sollte so wenig Parameter wie möglich enthalten, aber nicht weniger:

Mach' es so einfach wie möglich, aber nicht einfacher (Albert Einstein)

1.3 Modellgleichungen

Die das ökonometrische Modell beschreibenden mathematischen Gleichungen kann man nach zwei Kriterien klassifizieren: Bezüglich der *Inhaltlichen* bzw. *semantischen* Struktur und bezüglich der *mathematischen* Struktur.

1 Allgemeines

1.3.1 Inhaltliche Struktur

Hier unterscheidet man nach zwei Kategorien:

- Das **ökonomische Modell im engeren Sinn** wie Gl. (1.1) beschreibt den *abstrakten funktionalen Zusammenhang*. Manchmal wird noch zwischen dem *allgemeinen Modell* (nicht spezifizierte Werte der Modellparameter) und dem *geschätzten Modell* (nach der Parameterschätzung) unterschieden. Im einfachstmöglichen nicht-trivialen Fall eines linearen univariaten Eingleichungsmodells (eine exogene und eine endogene Variable) lautet das allgemeine ökonomische Modell

$$Y(x, \beta) = \beta_0 + \beta_1 x + \epsilon \quad (1.4)$$

und das geschätzte Modell

$$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x \quad (1.5)$$

Im *allgemeinen Modell* haben die Parameter feste, aber noch nicht bestimmte Werte und die Unsicherheit wird durch die Zufallsgröße ϵ ausgedrückt. Im *geschätzten Modell* gibt es keinen expliziten Zufallsterm mehr. Vielmehr sind die Zufallseinflüsse auf die geschätzten Parameterwerte $\hat{\beta}_0$ und $\hat{\beta}_1$ übergegangen, welche nun *ihrerseits* Zufallsvariablen darstellen.

Im Falle eines linearen multivariaten Eingleichungsmodells (mehrerer exogene und eine endogene Variable) lautet das allgemeine ökonomische Modell

$$Y(x) = \beta_0 + \sum_{m=1}^M \beta_m x_m + \epsilon \quad (1.6)$$

- Die **Systemgleichungen**, manchmal auch als **Messgleichungen** bezeichnet, beschreiben die *Anwendung* des ökonomischen Modells auf konkrete Systeme bzw. auf die von diesen Systemen gemessenen Werte der exogenen und endogenen Variablen. Im Gegensatz zum abstrakten Modell hängen die Systemgleichungen vom konkreten System ab. Die zum ökonomischen Modell (1.4) gehörigen Systemgleichungen lauten

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n \quad (1.7)$$

und werden auf n Datensätze $\{(x_i, y_i)\}$ angewandt. Die Systemgleichungen zum Mehrfachregressionsmodell (1.6) lauten

$$Y_i = \beta_0 + \sum_{m=1}^M \beta_m x_{mi} + \epsilon_i \quad (1.8)$$

und werden auf n Datensätze $\{(x_{1i}, x_{2i}, \dots, x_{Mi}, y_i)\}$ angewandt. Mit Hilfe der Systemgleichungen kann man die Modellparameter derart schätzen, dass eine Funktion $F(\{\epsilon_i\})$ der Modellierungsfehler ϵ_i , z.B. die Fehlerquadratsumme, minimal wird (Kapitel 2.4)

1 Allgemeines

Merke: Wichtig ist es, bei einfach indizierten exogenen Variablen zu unterscheiden, ob es sich um ein abstraktes Mehrfachregressionsmodell der Art (1.6) oder um Systemgleichungen eines Einfachregressionsmodells der Art (1.7) handelt.

Verständnisfrage:

Warum sollte es bei den Systemgleichungen immer mehr Sätze von Messwerten geben als es der Zahl der Parameter entspricht ($n > J + 1$)? Was passiert, wenn $n = J + 1$ oder gar $n < J + 1$?

1.3.2 Mathematische Struktur

Unterscheidung nach Linearität

Die Unterscheidung ist in Hinblick auf die Lösungsmethoden wichtig. Man kann folgende Kategorien unterscheiden:

- **Lineare Modelle:** Hier hängen die endogenen Variablen linear von den exogenen Variablen x und den Parametern β ab: Bei Eingleichungsmodellen haben sie also die Form

$$Y(x) = \beta_0 + \sum_{m=1}^M \beta_m x_m + \epsilon. \quad (1.9)$$

In vielen ökonomischen Lehrbüchern werden ausschließlich lineare Modelle behandelt, in der Verkehrsökonomie sind jedoch auch nichtlineare Modelle wichtig.

Beispiel: Einfaches Modell für die mittlere Fahrleistung Y pro Person in einer Region:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (1.10)$$

mit den erklärenden Variablen

- x_1 : Mittleres Einkommen (Euro pro Jahr)
- x_2 : Treibstoffpreis (Euro/Liter)
- x_3 : Ausbau des Straßennetzes (Kilometer pro Einwohner)

Als Aussagen lassen sich hier zum Beispiel Elastizitäten gewinnen wie

$$\epsilon_2 = \frac{x_2}{Y} \frac{dY}{dx_2} = \frac{\beta_2 x_2}{Y}$$

Ein Wert $\epsilon_2 = -0.2$ sagt z.B. aus, dass bei einer Erhöhung der Treibstoffpreise um 10% nur 2% weniger Auto gefahren wird.

- **Reduzible nichtlineare Modelle:** Zwar hängt hier die endogene Variable nichtlinear von sowohl den exogenen Variablen x als auch den Parametern β ab, man kann die Modellgleichungen jedoch durch Transformation der exogenen und/oder endogenen Variablen in eine lineare Form bringen.

1 Allgemeines

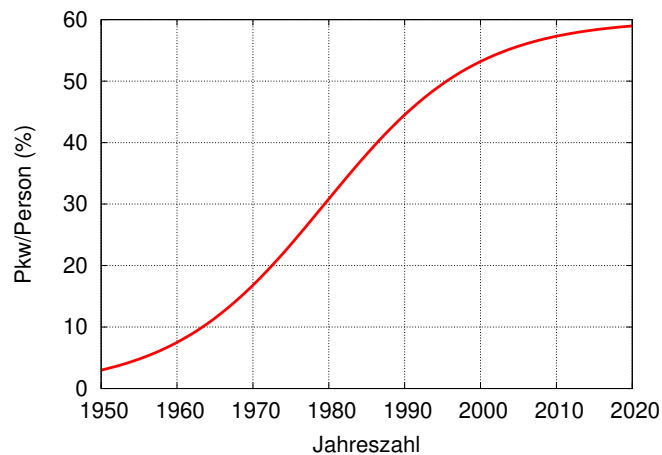


Abbildung 1.4: Graph des Modells (1.13).

Beispiel: Modell für das *unbeschränkte Wachstum* eines ökonomischen Prozesses (z.B. die anfänglichen Verkaufszahlen eines neu eingeführten Produkts). Mit der Zeit als exogenen Variablen lautet dieses

$$y(t) = y_0 e^{t/\tau + \epsilon} \quad (1.11)$$

Durch die Transformation $y = e^w$ und anschließende Logarithmierung wird dieses in ein lineares Eingleichungsmodell transformiert:

$$w(t) = \ln(y_0) + t/\tau + \epsilon = \beta_0 + \beta_1 t + \epsilon \quad (1.12)$$

Verständnisfrage:

Warum funktioniert dies bloß, wenn die Zufallsanteile im ursprünglichen Modell *multiplikativ* wirken?

- **Irreduzible nichtlineare Modelle:** Hier hängt die endogene Variable nichtlinear sowohl von den exogenen Variablen \mathbf{x} als auch von den Parametern ab und es ist keine Transformation auf eine lineare Form möglich. Dies ist zwingend immer dann der Fall, wenn die endogenen Variablen *diskreter Natur* sind (z.B. Y = Zahl der Wege, die mit dem Rad zurückgelegt werden) oder wenn es sich gar um **qualitative** bzw. **nominalskalierte Variablen** handelt, z.B. Y = gewählter Beruf mit den Ausprägungen Maurer, Schreiner, Physiker etc. Aber auch im Bereich der stetigen Modelle gibt es manchmal die Notwendigkeit von nichtlinearen Modellen, wie im folgenden Beispiel.

1 Allgemeines

Beispiel: Modell für das *Sättigungsverhalten* eines ökonomischen Prozesses (z.B. die Verkaufszahlen von Autos oder Mobiltelefonen seit Erfindung der jeweiligen Produkte) beschreiben will. Das klassische **Modell beschränkten Wachstums** mit der Zeit x als exogenen Variablen hat die Form (vgl. Abb. 1.4)

$$\hat{y}(x) = \frac{y_s}{1 + \left(\frac{y_s}{y_0} - 1\right) e^{-x/\tau}} \quad (1.13)$$

Verständnisfrage:

Diskutieren Sie die Bedeutung der Parameter τ , y_0 und y_s im Modell (1.13). Kann man das Modell auch in der Form $\hat{y}(x) = y_s/[1 + e^{-(x-x_0)/\tau}]$ schreiben? Was ist dann die Bedeutung des neuen Parameters x_0 und wie hängt er mit den Parametern der Formulierung (1.13) zusammen?

- **Quasilineare** bzw. **parameterlineare Modelle:** In dieser Modellklasse sind die Modellgleichungen linear bezüglich der Parameter β , aber nichtlinear bezüglich der exogenen Variablen \mathbf{x} . Solche Modelle kann man immer durch eine Transformation der exogenen Variablen linearisieren,⁴ so dass sie *völlig äquivalent* zu den linearen Modellen sind.

Alle Methoden linearer Modelle sind identisch auch auf die transformierten quasilinearen Modelle anwendbar. Im weiteren Verlauf dieses Skriptes sind deshalb bei linearen Modellen immer auch quasilineare Modelle mit eingeschlossen.

Beispiel: Einfaches Modell für die mittlere PKW-Fahrleistung Y eines PKW-Besitzers pro Jahr:

$$Y(\mathbf{x}, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \quad (1.14)$$

mit den exogenen Variablen

- x_1 : Mittleres Einkommen (Euro pro Jahr)
- x_2 : Treibstoffpreis (Euro/Liter)

Mit den Transformationen

$$z_1 = x_1, \quad z_2 = x_2, \quad z_3 = x_1 x_2$$

wird dieses Modell in ein lineares Modell mit *drei* exogenen Variablen transformiert:

$$Y(\mathbf{z}, \beta) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \epsilon \quad (1.15)$$

⁴Im Gegensatz zu den reduzierbaren nichtlinearen Modellen gibt es hier auch keine Probleme mit den Zufallsanteilen.

1 Allgemeines

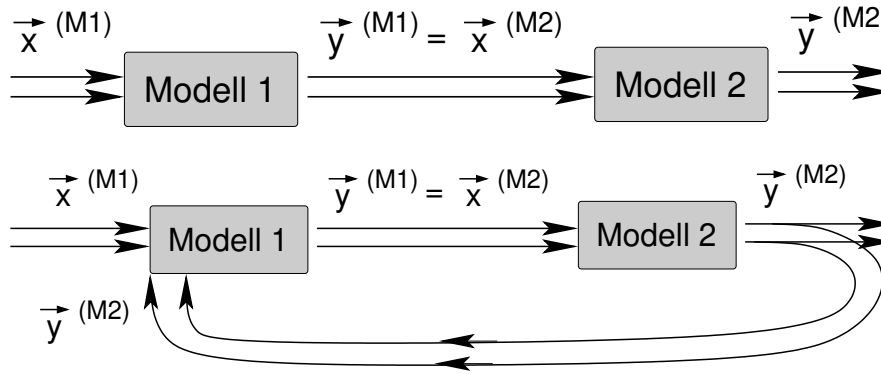


Abbildung 1.5: Flussdiagramm der exogenen und endogenen Variablen für verkettete (oben) und rückgekoppelte (unten) Modell-Systeme. Das ‐Innenleben‐ der Modelle (also die je zwei Gleichungen, vgl. Abb. 1.2) ist nicht mehr gezeigt. Die Vektorpfeile symbolisieren den ganzen Satz jeweiliger endogener und exogener Variablen.



Abbildung 1.6: Flussdiagramm der Verketten im Falle dynamischer Modelle. Das Modell selbst ist in der Regel in allen Schritten dasselbe, oft mit unveränderten Parametern (*autonomes* dynamisches Modell), manchmal mit von der Zeit abhängigen (*nichtautonomes* Modell).

Die zugehörigen Anstiegsparameter bedeuten

- β_1 : Anstieg der Fahrleistung mit dem Einkommen (üblicherweise $\beta_1 > 0$)
- β_2 : Preissensitivität der PKW-Nutzung (üblicherweise $\beta_2 < 0$)
- β_3 : Reduktion der Preissensitivität mit dem Einkommen (üblicherweise $\beta_3 > 0$)

Verständnisfrage:

Machen Sie sich die Vorzeichen der drei Parameter in diesem Beispiel klar. Warum ist bei Berücksichtigung auch extrem hoher Einkommen ein irreduzibel nichtlineares Modell notwendig? Zeigen Sie dies anhand der dann nicht plausiblen Aussagen des Modells (1.15).

1 Allgemeines

Unterscheidung nach weiteren mathematischen Kriterien

Weitere in der Ökonometrie relevante Unterscheidungsmerkmale sind

- Existenz eines Zufallsanteils: **Deterministische** vs. **stochastische** Modelle.
- Zahl der exogenen Variablen: Eine bei univariaten und mehrere bei multivariaten Modellen.
- Zahl der endogenen Variablen: Eine bei Eingleichungsmodellen, mehrere bei Mehrgleichungsmodellen.
- Skalierung der endogenen Variablen: **Entscheidungsmodelle** haben diskrete bzw. qualitative/nominalskalierte endogene Variablen, während **kontinuierliche Modelle** stetige sowie verhältnisskalierte endogene Variable aufweisen. Diese beiden Kategorien bedingen grundsätzlich verschiedene Herangehensweisen bei der Modellierung, beispielsweise sind Entscheidungsmodelle immer nichtlinear.

Hingegen ist die Skalierung der exogenen Variablen nicht so bedeutsam, da man qualitative bzw. nominalskalierte exogene Variable durch kardinalskalierte **Pseudovariablen** der Art 1=Maurer, 2=Schreiner, 3=Physiker etc ausdrücken kann.

- Existenz von Verkettung oder Rückkopplung (siehe Abb. 1.5 und Abb. 1.7):
 - Bei **verketteten Modellen** sind die endogenen Variablen einer Modellstufe die exogenen der nächsten. Wichtigstes Beispiel dafür ist das Vier-Stufen-Schema der klassischen Verkehrsplanung (Abb. 1.7).
 - Bei **rückgekoppelten Modellen** koppeln die endogenen Variablen eines Modells einer späteren Verkettungsstufe auf die exogenen Variablen eines Modells einer früheren Verkettungsstufe zurück. Beim Vier-Stufen-Modell der Abb. 1.7 können als Folge der Routenwahl einzelne Netzelemente überlastet, also verstaут werden. Die damit verbundenen Resiezeitverlängerungen beeinflussen wiederum die Ziel- und ggf. die Verkehrsmittelwahl.
- Dynamische vs. statische Modelle: **dynamische Modelle**, also Modelle mit expliziten dynamischen Zeitbezug, stellen einen Sonderfall der Verkettung im Zeitbereich dar: Die Entwicklung des aktuellen Zeitschrittes hängt von den Ergebnissen des vergangenen Zeitschrittes bzw. der vergangenen Zeitschritte ab (Abb. 1.6). Im Rahmen der Vorlesung werden wir einen einfachen Typ dieser Modelle kennen lernen, das **dynamische Input-Output-Modell** (Abschnitt 5).

1.4 Ein Anwendungsbeispiel: Verkehrsmittelwahl

Zur Veranschaulichung der in diesem Kapitel beschriebenen Konzepte soll das folgende Beispiel der Modellierung einer Verkehrsmittelwahl dienen.

1 Allgemeines

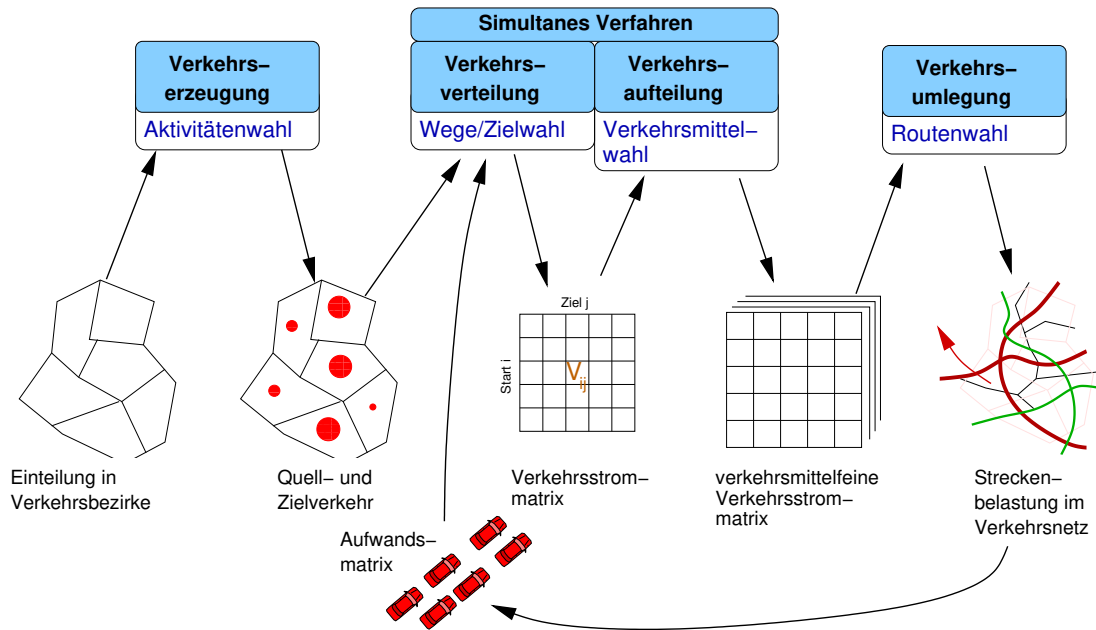


Abbildung 1.7: Das klassische Vierstufenmodell der Verkehrsplanung als Beispiel der Verkettung und Rückkopplung von Modellen. Die endogenen Variablen der ersten Modellstufe “Aktivitätenwahl”, die im Verlauf eines Tages anfallenden Verkehrsströme von und nach jedem Bezirk (Quell- bzw. Zielsummen), spielen gleichzeitig die Rolle von exogenen Variablen bei der Modellierung der “Zielwahl”. Die endogenen Variablen der Zielwahl (Quelle-Ziel-Relationen bzw. Wege zwischen jeweils zwei Bezirken) sind wiederum die exogenen Variablen der Verkehrsmittelwahl und in weiterer Folge der Routenwahl. Das Ergebnis der Routenwahl führt ggf. zu einer Überlastung einzelner Netzelemente und zu einer Erhöhung der Reisezeit auf einigen Relationen. Diese sogenannte “Reisezeitmatrix” bzw. “Aufwandsmatrix” ist nun wiederum eine weitere exogene Variable der Zielwahl (“wenn der Weg von Bezirk i nach j immer verstopft ist, gehe ich vielleicht doch woanders einkaufen”), so dass die Routenwahl auf die Zielwahl rückkoppelt.

1.4.1 Modellstruktur

Das Beispiel enthält als Verkettung die wichtigsten beiden der in dieser Vorlesung behandelten Modellklassen (Abb. 1.8):

- In der ersten Stufe gibt es als Teilmodelle vier stetige Eingleichungsmodelle, welche aus den exogenen Variablen
 - Zeitbedarf X_{1k} des Verkehrsmittels k ,
 - Kosten X_{2k} des Verkehrsmittels k ,

1 Allgemeines

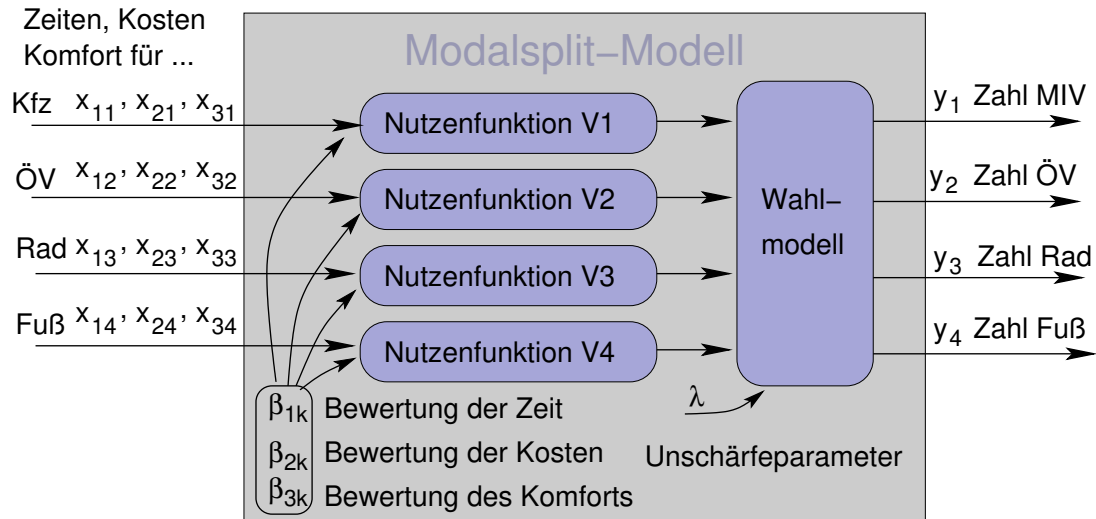


Abbildung 1.8: Beispiel eines zweistufigen ökonomischen Wahlmodells. Zunächst werden in vier voneinander unabhängigen stetigen Eingleichungsmodellen die deterministischen Nutzenfunktionen V_k der verschiedenen Verkehrsmittel k modelliert. Diese wiederum spielen die Rolle von exogenen Variablen im eigentlichen Wahlmodell. Dieses diskrete Mehrgleichungsmodell liefert letztendlich als endogene Variablen die Zahlen Y_k der Entscheidungen für Verkehrsmittel k .

– Komfort X_{3k} des Verkehrsmittels k

den deterministischen Nutzen V_k des jeweiligen Modus als endogene Variable ausgeben.

- Das eigentliche *Wahlmodell* enthält als verkettete exogene Variablen die deterministischen Nutzenfunktionen V_k der vier Alternativen und liefert die prognostizierten Entscheidungszahlen \hat{Y}_k für jede Alternative. In seiner Eigenschaft als diskretes Mehrgleichungsmodell ist das Wahlmodell notwendigerweise nichtlinear.

1.4.2 Modellgleichungen und Parameter

Die vier Modelle für die Nutzenfunktionen werden als voneinander unabhängige lineare (oder auch quasilineare) Eingleichungsmodelle formuliert:

$$U_k(\mathbf{x}_k, \boldsymbol{\beta}_k) = \beta_{0k} + \sum_{m=1}^3 \beta_{mk} x_{mk} + \epsilon_k \quad (1.16)$$

Hierbei bedeuten:⁵

⁵Die Parametrisierung ist nicht vollständig, es fehlen z.B. die sozioökonomischen Variablen (siehe Kapitel 4). Als einführendes Beispiel dient das vereinfachte Modell aber hier seinen Zweck.

1 Allgemeines

- β_{0k} : Globaler Bonus des Verkehrsmodus k gegenüber Modus 1 (damit ist offensichtlich $\beta_{01} = 0$),
- β_{1k} : Zeitbewertung bei Verkehrsmodus k (beispielsweise kann man die Zeit zu Fuß unangenehmer empfinden als die Zeit im Kfz, dann gilt $0 > \beta_{11} > \beta_{14}$),
- β_{2k} : Geldbewertung bei Verkehrsmodus k ,
- β_{3k} : Bewertung des Komforts bei Verkehrsmodus k .

Von diesen Nutzenfunktionen wird der deterministische Anteil $\hat{U}_k = V_k$ sowie die *Standardabweichung* β_0 des Zufallsnutzens an das eigentliche Wahlmodell weitergegeben. Bei insgesamt N vorzunehmenden Entscheidungen hat das einfachste und am weitesten verbreitete Wahlmodell, das Multinomial-Logit-Modell, die Modellgleichung

$$Y_k(\mathbf{V}, \beta_0) = NP_k(\mathbf{V}, \beta_0) = N \frac{e^{\beta_0 V_k}}{\sum_{k'} e^{\beta_0 V_{k'}}}. \quad (1.17)$$

1.4.3 Systemgleichungen

Da die verketteten intermediären Variablen V_k nicht messbar sind, muss man die Verkettung der Teilmodelle (1.16) und (1.17) als Black-Box betrachten. Damit gelangt man zu einem Modell der Form (1.3) mit messbaren exogenen Variablen x_{mk} und endogenen Variablen Y_k .

Um dieses Modell zu schätzen, benötigt man n Messungen jeweils aller exogenen und endogenen Variablen. Eine ‘‘Messung’’ ist hier der technische Ausdruck einer Befragung einer Person: Von jeder Person i müssen also folgende Werte erfragt werden:⁶

$$\left\{ \underbrace{x_{mki}}_{\substack{\text{Eigenschaften} \\ \text{der Modi}}}, \underbrace{y_{ki}}_{\substack{\text{Zahl Entscheidungen} \\ \text{für Alternative } k}} \right\}$$

Nach einer erfolgreichen Befragung hat der Interviewer eine Tabelle, welche folgendermaßen aussehen kann:

⁶Die Werte der exogenen Variablen haben einen Dreifachindex, da die Menge der exogenen Variablen bereits doppelt indiziert ist.

1 Allgemeines

Größe	Zeit- bedarf Auto	Kosten Auto	Subj. Komfort Auto (1-6)	Kompl. Reisezeit ÖPNV	...	Wahl- entsch. Auto	Wahl- entsch. ÖPNV	...
Variable	x_{11i}	x_{21i}	x_{31i}	x_{12i}	...	y_{1i}	y_{2i}	...
Person 1	20 min	2.50 €	3	30 min	...	0	1	...
Person 2	11 min	2.00 €	1	20 min	...	1	0	...
Person 3	34 min	4.00 €	1	15 min	...	0	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Anhand der Tabelle und der Zahl N_i der Entscheidungen, welche Person i insgesamt trifft, werden die Systemgleichungen formuliert,

$$\hat{Y}_{ki} = N_i \frac{e^{V_k(\mathbf{x}_i, \boldsymbol{\beta})}}{\sum_{k'} e^{V_{k'}(\mathbf{x}_i, \boldsymbol{\beta})}}, \quad i = 1, \dots, n. \quad (1.18)$$

Hierbei fasst der Vektor $\mathbf{x}_i = \{x_{11i}, \dots, x_{34i}\}$ die für Person i relevanten exogenen Variablen zusammen. Anhand der Systemgleichungen wird der (für alle Personen gleiche!) Parametervektor $\boldsymbol{\beta}$ durch Vergleich der modellierten endogenen Variablen \hat{Y}_{ki} mit den gemessenen (erfragten) endogenen Variablen Y_{ki} geschätzt.

Aggregation

Grundsätzlich sind Modelle der Verkehrsmittelwahl **mikroskopisch**, betrachten also jede Entscheidung einzeln. Damit kann auch nur eine der endogenen Variablen y_k für Alternative k gleich 1 sein, während alle anderen $y_{k'}$, $k' \neq k$, gleich 0 sein müssen. In drei Fällen ist eine *Aggregation*, also eine **makroskopische Herangehensweise** möglich, sinnvoll oder sogar notwendig:⁷

- *personenbezogene Aggregation*: Ein und dieselbe Person wird in vergleichbaren Situationen (gleiche exogene Variable) mehrfach befragt. Geht es z.B. um die Verkehrsmittelwahl zur Arbeit und wurden die Entscheidungen den letzten zwei Wochen erhoben, kann man die 10 Entscheidungen zusammenfassen.
- *situationsbezogene Aggregation*: Wählt man die exogenen Variablen so, dass sie für mehrere Personen *und alle Alternativen* etwa gleiche Werte besitzen, kann man diese Personen zu Gruppen bzw. Klassen zusammenfassen. In der Praxis funktioniert dies meist nur, wenn man nur Variable wählt, welche für alle Alternativen etwa gleich sind, also entweder sozioökonomische Variable (Alter, Geschlecht, Autoverfügbarkeit), externe Variable (Wetter) oder spezielle generische Variable wie die Entfernung.

⁷Hier dient die Aggregation hauptsächlich zu "pädagogischen" Zwecken, da sie die Darstellung vereinfacht. In Realität wird man bei Vorliegen mikroskopischen Datenmaterials immer das mikroskopische Vorgehen wählen.

1 Allgemeines

- *datenerzwungene Aggregation*: Liegen Daten nur aggregiert vor, kann man natürlich nur makroskopisch vorgehen.

Beispiel: mikroskopische Daten, welche eine Aggregation zulassen (situationsbezogene Aggregation)

	Entf. Fuß	Entf. Rad	Entf. ÖPNV	Entf. MIV	Wahl- entsch. Fuß	Wahl- entsch. Rad	Wahl- entsch. ÖPNV	Wahl- entsch. MIV
Variable	x_{11}	x_{12}	x_{13}	x_{14}	y_1	y_2	y_3	y_4
Person 1	2 km	2 km	3 km	3 km	0	1	0	0
Person 2	4 km	4 km	4.5 km	∞	0	0	1	0
Person 3	3 km	3 km	∞	3.5 km	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Aggregation dieser Daten in Entfernungsklassen

Entfernungsklasse	Wahl- entsch. Fuß	Wahl- entsch. Rad	Wahl- entsch. ÖPNV	Wahl- entsch. MIV	Pers. in Klasse
$x_1 =$ Klassenmitte	y_1	y_2	y_3	y_4	n_i
0-1 km	5	1	0	0	6
1-2 km	2	0	5	0	7
2-3 km	0	2	11	0	13
3-5 km	0	0	7	1	8
5-10 km	0	0	5	0	5
10-20 km	0	0	1	1	2
> 20 km	0	0	3	0	3
Σ	7	3	32	2	44

1 Allgemeines

1.5 Verwendete Symbole

Y	Abhängige bzw. erklärte bzw. endogene Variable. Bei Mehrgleichungsmodellen bekommen diese den Index k .
$x_m,$ $m = 1, \dots, M$	Unabhängige bzw. erklärende bzw. exogene Variablen
\mathbf{x}	Gesamtheit der exogenen Variablen, als Spaltenvektor geschrieben
$y_i, x_{mi},$ $i = 1, \dots, n$	Wert der abhängigen bzw. der m -ten unabhängigen Variablen bei der i -ten Messung bzw. dem i -ten Element der Stichprobe.
β_0	Achsenabschnitt (engl. <i>intercept</i>)
$\beta_j,$ $j = 1, \dots, J$	Lineare Anstiegsparameter
β	Gesamtheit der Modellparameter, als Spaltenvektor geschrieben
$\hat{\beta}_j, \hat{Y}$	Das "Dach" ist ein Symbol für geschätzte Größen, also geschätzte Parameterwerte, exogene Variable des geschätzten Modells, etc.
ϵ	Ein additiver i.i.d. Zufallsterm mit Erwartungswert 0 und der Varianz σ^2
z_m, w	Transformierte exogene und endogene Variablen (mit dem Ziel, ein lineares Modell in den transformierten Variablen zu erhalten)
V_k	Deterministische Nutzenfunktion der Alternative k bei Wahlmodellen
U_k	Gesamtnutzen $V_k + \epsilon_k$ der Alternative k bei Wahlmodellen
n	Zahl der Stichprobenelemente bzw. Messungen bzw. befragte Personen (Laufindex $i = 1, \dots, n$)
N	Zahl der Entscheidungen bei Wahlmodellen, $N = \sum_k Y_k$
N_i	Zahl der Entscheidungen von Person i bei Wahlmodellen, $N_i = \sum_k y_{ki}$

2 Stetige ökonomische Modelle

In diesem Kapitel werden Sachverhalte bzw. entsprechende ökonomische Modelle betrachtet, in denen die *endogene* bzw. erklärte Variable Y *stetiger* Natur ist. Solche Modelle auch als **Regressionsmodelle** bezeichnet.¹ Diskrete endogene Variable, welche im Verkehrskontext typischerweise die Ergebnisse von Wahlentscheidungen darstellen, werden im Kap. 4, **diskrete Wahlmodelle**, beschrieben. Aus prinzipiellen Gründen sind Modelle mit diskretwertigen endogenen Variablen immer nichtlinearer Natur, häufig sind auch mehrere endogene Variable als Mehrgleichungsmodell gekoppelt. Im Falle stetiger exogener Variablen hingegen ergeben (quasi-)lineare Eingleichungsmodelle einen Sinn und werden weitaus häufiger angewandt als nichtlineare Modelle oder Mehrgleichungsmodelle. Häufig kann man nämlich nichtlinear wirkende *exogene Variable* durch eine oder mehrere lineare *Faktoren* ersetzen. Abgesehen von den für alle stetigen Modelle gültigen Betrachtungen in den Abschnitten 2.1 - 2.3 werden deshalb nur lineare bzw. quasilineare Modelle betrachtet, welche mathematisch äquivalent behandelt werden können (vgl. Abschnitt 1.3.2). Einen Sonderfall stellt die **logistische Regression** (Kap. 2.8) dar, die Problemstellungen mit nominalskalierten endogenen Variablen, die eigentlich der Regression nicht zugänglich sind, in eine Regressionsstruktur bringt. Diese Modelle sind äquivalent zu den bereits erwähnten Modellen der diskreten Wahltheorie (Kap. 4), was bedeutet, dass insbesondere die Parameterschätzung eine andere als die hier angewandte **Least-Squared Errors**-Schätzung ist. Da die Formulierung als diskretes Wahlmodell anschaulicher und besser verallgemeinerbar ist, wird die logistische Regression nur kurz erwähnt.

2.1 Struktur der Gleichungen

Wir betrachten *eine* endogene bzw. erklärte Variable Y , welche von i. A. mehreren exogenen bzw. unabhängigen bzw. erklärenden Variablen x_m , $m = 1, \dots, M$, beschrieben wird. Wir schließen eine Rückkopplung aus, so dass die Beschränkung auf eine abhängige Variable, also auf sogenannte **Eingleichungsmodelle**, keine Einschränkung der Allgemeinheit darstellt.² Die obige Annahmen ergeben als Spezialisierung der allgemeinen Gleichungen (1.1) eine Modellgleichung folgender Form:

$$Y = \hat{y}(\mathbf{x}, \boldsymbol{\beta}) + \epsilon = \hat{y}(x_1, \dots, x_M, \beta_0, \beta_1, \dots, \beta_J) + \epsilon. \quad (2.1)$$

¹Den Name "Regression", d.h. Rückkehr (zum Mittelwert), hat im 19. Jahrhundert Francis Galton, ein Cousin Charles Darwins, geprägt, um das Phänomen der *Regression zur Mitte* zu beschreiben: Nachfahren großer Eltern tendieren dazu, nur durchschnittlich groß zu werden. Heute hat sich die Methode von dieser speziellen Anwendung emanzipiert.

²Gegebenenfalls definiert man für jede abhängige Variable ein eigenes Modell.

2 Stetige ökonomische Modelle

Die exogenen Variablen wirken im Allgemeinen nichtlinear auf die endogene Variable. Insbesondere ist, wie bereits in der Einleitung erläutert, bei diskretwertigen und nominalskalierten exogenen Variablen eine Nichtlinearität zwingend. Durch geeignete parameterfreie nichtlineare Transformationen, *die sich aus dem Sachverhalt ergeben müssen*, kann man eine nichtlinear wirkende oder nicht-metrische exogene Variable in eine oder mehrere **lineare Faktoren** transformieren. Beispiele dazu finden sich in den Abschnitten 1.3.2, 2.3.1 und 2.7. Bei der *linearen* Regression werden also aus den M ursprünglichen exogenen Variablen J lineare Faktoren, wobei i.A. $J \geq M$. Da aus dem Kontext (lineare oder nichtlineare Modellierung) klar ist, was gemeint ist, werden wir im weiteren auch die linearen Faktoren synonym als exogene Variable x_j bezeichnen, auch wenn sie, streng genommen, unterschieden werden müssten. Bei linearen Modellen (oder quasilinearen Modellen nach Transformation der exogenen Variablen in Faktoren) kann man das allgemeine ökonomische Eingleichungsmodell (2.1) schreiben als

$$Y(\mathbf{x}, \boldsymbol{\beta}) = \hat{y}(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^J \beta_j x_j + \epsilon. \quad (2.2)$$

Diese Gleichungen enthalten die folgenden Größen:

- Y ist die endogene (erklärte) Variable. Um ihre Natur als Zufallsvariable zu verdeutlichen, wird sie gemäß allgemeiner Konvention “groß” geschrieben.
- Die Funktion $\hat{y}(\cdot) = f(\cdot)$ beschreibt den durch das Modell erklärten Zusammenhang zwischen der endogenen und den exogenen Variablen. Da die Funktion den wahren Wert von Y schätzt (Y und $\hat{y}(\cdot)$ unterscheiden sich durch den nichterklärten Anteil), bekommt sie, wie alle Schätzfunktionen, ein “Dach”
- $\mathbf{x} = \{x_1, \dots, x_m, \dots, x_M\}$ bezeichnet die exogenen (erklärenden) Variablen bzw. die daraus extrahierten linearen Faktoren. Diese werden als bekannt bzw. kontrollierbar angenommen, so dass sie keinen Zufallsanteil besitzen und konsequenterweise “klein” geschrieben werden. Im multivariaten Fall werden sie oft kompakt als *Variablenvektor* \mathbf{x} geschrieben.
- $\boldsymbol{\beta} = \{\beta_0, \dots, \beta_j, \dots, \beta_J\}$ bezeichnet die Modellparameter. Auch diese werden oft kompakt als “Parametervektor” geschrieben. Im linearen Fall ist $J = M$, d.h. die Zahl ($J + 1$) der Parameter ist um eines größer als die Zahl der exogenen Variablen. Die **Schätzung** der Parameter bzw. **Kalibrierung** des Modells ist ein wesentliches Thema dieses Kapitels.
- Schließlich beschreibt der **Residualanteil** bzw. **nichterklärte Anteil** alle Anteile bzw. Zufallseinflüsse, welche nicht durch $\hat{y}(\dots)$ erklärt werden können.

Der nichterklärte Teil ϵ resultiert im Wesentlichen aus folgenden Gründen:

- Der Sachverhalt enthält explizite Zufallselemente.

2 Stetige ökonomische Modelle

- Die zu beschreibenden statistischen Einheiten sind heterogener Natur, müssten also eigentlich durch individuelle Parametersätze beschrieben werden. Dabei darf allerdings kein Strukturbruch auftreten, da das Problem sonst **fehlspezifiziert** wäre (vgl. Abschnitt 2.3.2).
- Nicht alle exogenen (bzw. unabhängigen bzw. erklärenden) Variablen sind bekannt oder im Modell enthalten. Allerdings sollten die nicht enthaltenen Einflüsse nicht zu *systematischen* Fehlern führen. Andernfalls wäre das Modell wieder fehlspezifiziert.
- Das mathematische Modell selbst ist nicht exakt zutreffend. Auch hier sollten die Fehler allerdings nicht zu systematischen Verzerrungen führen. Sonst ist das Modell wieder fehlspezifiziert.

Zu beachten ist, dass die mit einer Fehlspezifikation einhergehenden systematischen Verzerrungen zu falschen Schlussfolgerungen bis hin zur Unbrauchbarkeit des jeweiligen Modells führen.

2.2 Vorgehen bei der ökonomischen Untersuchung

2.2.1 Ablauf bei vorgegebenem Modell

Meist wird man kein eigenes Modell entwickeln, sondern ein bewährtes Modell anwenden, welches nicht weiter hinterfragt wird (bei Softwarepaketen kann man an den Modellen sowieso nicht “drehen”). In diesem Fall verläuft die ökonomische Untersuchung gemäß dem in Abb. 2.1 gezeigten Schema ab:

1. **Spezifikation** (Abschnitt 2.3): Gemäß dem Sachverhalt wählt man den Modelltyp aus (z.B. linear oder nichtlinear) und bestimmt die Anzahl und Bedeutung der exogenen und endogenen Variablen. Wird der Sachverhalt durch eine bekannte mathematische Beziehung beschrieben (z.B. quadratische Abhängigkeit des Bremsweges von der Geschwindigkeit, exponentielles Verhalten bei Wachstumsprozessen ohne Sättigung, die logistische Funktion bei Wachstumsprozessen mit Sättigung) oder will man auf ein **ökonomisches Modell** (z.B. das nachfragebasierte Keynes'sche Modell) zurückgreifen, ist das Modell dadurch vorgegeben. Häufig wird man aber auch *phänomenologisch* vorgehen und dann meist das einfachstmögliche Modell, also ein lineares Eingleichungsmodell, wählen. Ferner gehört zur Spezifikation, gewisse Bedingungen an die exogenen und endogenen Variablen und ihrer Messwerte, sowie am nichterklärten Anteil zu überprüfen. Erst das korrekt spezifizierte mathematische bzw. ökonomische Modell wird als **ökonomisches Modell** bezeichnet.
2. **Parameterschätzung** bzw. **Modellkalibrierung** (Abschnitt 2.4): Zunächst werden Datensätze des zu untersuchenden Sachverhalts (Werte für alle exogenen und endogenen Variablen) erhoben, meist in Form einer Stichprobe (Kap. 3). Anhand der Datensätze werden die Modellparameter so angepasst, dass das Modell die Daten so gut wie möglich beschreibt, also die nichterklärten Anteile nach bestimmten

2 Stetige ökonomische Modelle

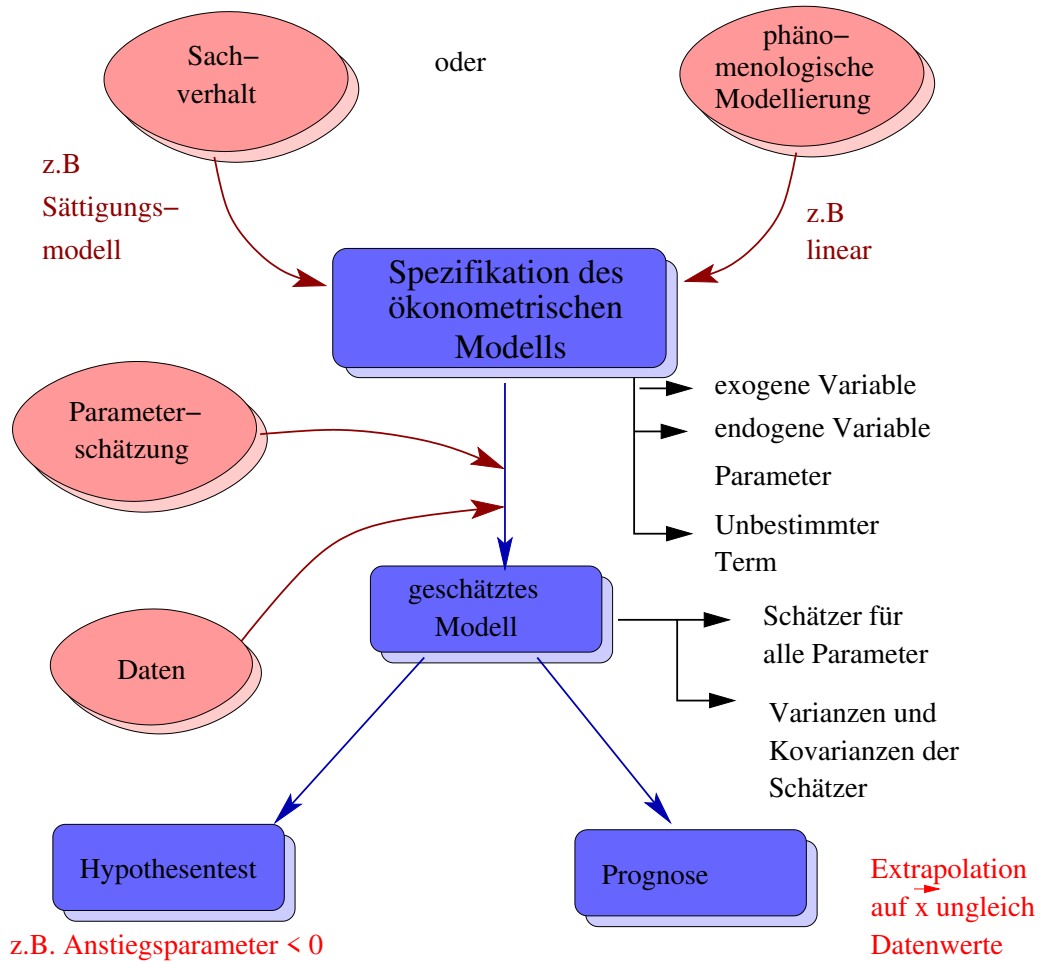


Abbildung 2.1: Flussdiagramm des ökonomischen Vorgehens bei Eingleichungsmodellen mit kontinuierlicher endogener Variable (Regressionsmodelle).

2 Stetige ökonomische Modelle

Kriterien minimiert werden. Das Ergebnis wird als **geschätztes ökonomisches Modell** bezeichnet.

3. **Statistische Eigenschaften der Schätzung** (Abschnitt 2.5): Hierbei wird mit statistischen Mitteln (Konfidenzintervalle der Parameterwerte sowie Signifikanztests) untersucht, ob die im Wesentlichen in den Beträgen und Vorzeichen der geschätzten Modellparameter enthaltenen Aussagen **signifikant** sind, also nicht durch den Zufall entstanden sein können.
4. **Modellaussagen** (Abschnitt 2.6): Das Ergebnis einer ökonomischen Untersuchung besteht zum Einen in den vom Modell im Anwendungs- bzw. Prognosefall vorausgesagten endogenen Variablenwerten. Zum Anderen können auch die Konfidenzintervalle der Parameterwerte (also die Ergebnisse des Belastbarkeitstests) sowie abgelehnte bzw. nicht ablehnbare Nullhypothesen der statistischen Tests eine Aussage darstellen, beispielsweise der Art “Der Zeitwert bei Benutzung des ÖPNV liegt im Freizeitverkehr zwischen 6 €/h und 10 €/h, im Berufsverkehr aber bei mindestens 15 €/h” oder “man kann die Annahme eines konstanten Reisezeitbudgets nicht verwerfen”.
5. **Prognosen** (Abschnitt 2.6.3): Schließlich kann man mit dem geschätzte Modell *mit unveränderten Parameterwerten* Prognosen durchführen, insbesondere
 - (i) Voraussage der Auswirkungen geänderter Einflussfaktoren bzw. exogener Variablen (“eine Verteuerung der ÖPNV-Tickets um 10% wird zu einem Einbruch der Nachfrage zwischen 6 % und 9 % führen“),
 - (ii) Prognosen für die Zukunft (“ein Jahr nach der Preiserhöhung werden zwischen 30% und 60% der Abwanderer wieder zum ÖPNV zurückgekehrt sein”). Dies ist ein Spezialfall von (i) für den Fall, dass für die Entwicklung der relevanten exogenen Variablen separate Modelle vorliegen bzw. eine der exogenen Variablen die Zeit selbst darstellt.

2.2.2 Modellentwicklung und Validierung

Liegt kein verlässliches Modell vor oder hat man Zweifel an der Gültigkeit der infrage kommenden Modelle, muss man die Untersuchung mit einer *Modellentwicklung* koppeln und die Modelle *selbst* (also die *Modellstruktur*, nicht nur die Modellspezifikationen und die aus den Modellen gewonnenen Aussagen) testen. Dies wird auch als **Modellvalidierung** bezeichnet.

Hierbei wird eine Aussage bzw. Prognose *simuliert*, also das kalibrierte Modell mit unveränderten Modellparametern auf neue Systeme (ähnliche Sachverhalte bzw. derselbe Sachverhalt für andere Zeitpunkte bzw. Zeiträume) angewandt. Im Gegensatz zur eigentlichen Anwendung/Prognose sind im Zielsystem allerdings alle Daten bekannt, insbesondere die der endogenen Variablen. Man unterscheidet zwei Vorgehensweisen:

- **Holdout-Validation.** Der verfügbare Datensatz wird segmentiert und dem Modell ein Segment bei der Schätzung vorenthalten. Anschließend wird die Aussage-

2 Stetige ökonomische Modelle

und Prognosefähigkeit dieses geschätzten Modells durch das “Test-Segment” überprüft. Hat man beispielsweise für n Städte Datensätze, werden $n - 1$ Städte zum Kalibrieren und die Daten einer “Test-Stadt” zur Validierung verwendet.

- **Cross-Validation.** Hier wird der Reihe nach eine Holdout-Validierung mit jeweils einem Daten-Segment als Test-Datensatz durchgeführt. Im Städtebeispiel dient also reihum jeweils eine Stadt als “Test-Stadt”.

Bei der Validierung werden die vorausgesagten und tatsächlichen Werte verglichen und getestet, ob sich die Abweichungen in vertretbaren Grenzen hält bzw. ob sie sich gegenüber der Kalibrierung stark erhöht haben. Ein wichtiges Kriterium für gute Modelle ist außerdem ihre **Robustheit**: Hierbei wird das Modell auf die neuen Daten kalibriert und die geschätzten Modellparameter mit denen der “alten” Kalibrierung verglichen. Je weniger Abweichungen man findet, desto robuster und aussagekräftiger ist das Modell.

Ist man mit dem Ergebnis der Validierung nicht zufrieden, muss das Modell entsprechend modifiziert werden, worauf sich ein neuer Zyklus aus Kalibrierung/Validierung und ggf. neuer Datenbeschaffung anschließt.

Ein Spezialfall der Validierung stellt der sogenannte **Back-Test** dar, bei welchem zum selben Sachverhalt Test-Daten (Stichproben) für mehrere Zeitpunkte bzw. Zeiträume vorliegen. Nur ein Teil dieser Daten wird zur Kalibrierung verwendet, der andere zur Validierung.

Beispiel:

Liegen die Mobilitätsdaten einer Stadt jährlich für die Jahre 2001-2008 vor, kann man die Daten von 2001-2007 für die Kalibrierung und das Jahr 2008 für die Validierung heranziehen (Back-Test). Man kann zur Validierung auch die Daten einer anderen Stadt derselben Grundgesamtheit (z.B. westdeutsche Großstadt) verwenden.

2.3 Modellspezifikation

Hier und im Rest dieses Kapitels werden **lineare Eingleichungsmodelle** mit einer oder mehreren exogenen Variablen bzw. quasilineare Modelle betrachtet. Zwar sind viele Aussagen auch auf nichtlineare Modelle anwendbar, diese werden aber hier nicht explizit betrachtet.

Als **Modellspezifikation** bezeichnet man die Gesamtheit aller Bedingungen, welche das lineare Modell (2.2), also

$$Y(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^J \beta_j x_j + \epsilon,$$

2 Stetige ökonometrische Modelle

und die dazugehörigen Systemgleichungen

$$Y_i = \beta_0 + \sum_{j=1}^J x_{ij}\beta_j + \epsilon_i \quad (2.3)$$

erfüllen müssen, damit valide statistische Aussagen möglich sind. Dies betrifft die Modellfunktion selbst (funktionale Spezifikation), die Eigenschaften des nichterklärten Anteils (statistische Spezifikation) und Anforderungen an die Datensätze zur Parameterschätzung (Datenspezifikation).

Ist das ökonometrische Modell nicht korrekt spezifiziert, sind die daraus gewonnenen Aussagen im besten Fall unnötig unscharf, meist aber verzerrt und damit falsch, auch wenn die ökonometrische Software noch so kleine Konfidenzintervalle und noch so signifikante Aussagen vorspiegelt. Es gilt dann das Prinzip *“Junk in - Junk out”!*.

2.3.1 Funktionale Spezifikation

Hier gibt es drei Bedingungen:

1. **Exogene Variable:** *Keine relevante exogene Variable (=Einflussfaktor) fehlt und keine überflüssige ist im Modell enthalten.*

Man könnte beispielsweise folgendes Modell für den Normverbrauch von PKW mit Verbrennungsmotoren aufstellen:

$$Y = \beta_0 + \sum_{j=1}^4 \beta_j x_j + \epsilon$$

mit

- x_1 : Gewicht
- x_2 : Leistung des Motors
- x_3 : Farbe
- x_4 : Kraftstoffart

In diesem Modell ist die Farbe eine irrelevante Variable, während das Herstellungsjahr als relevante Variable fehlt: Neuere Autos verbrauchen unter *ceteris paribus*-Bedingungen, also bei gleicher Leistung, gleichem Gewicht und gleicher Kraftstoffart, tendenziell weniger. Die Tatsache, dass eine **nominalskalierte** bzw. **qualitative** Variable (die Kraftstoffart mit den Ausprägungen Benzin und Diesel) auftritt, hat hingegen auf die Spezifikation keine negative Auswirkung, da man sie durch

2 Stetige ökonomische Modelle

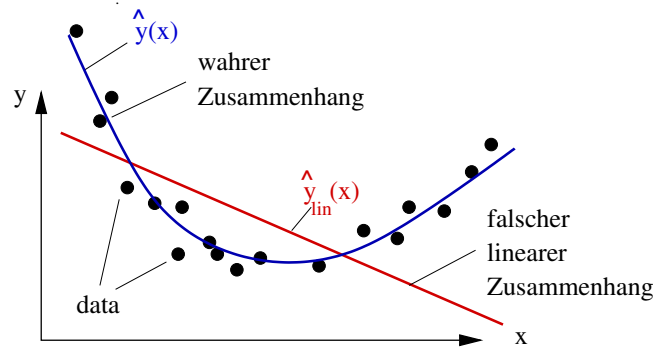


Abbildung 2.2: Linearer und nichtlinearer Zusammenhang.

eine Pseudo-Kardinalskalierung (0=Benzin, 1=Diesel) in eine numerische Variable verwandeln kann. Näheres in Abschnitt 2.7.³

- *Folgen einer Spezifikationsverletzung:* Das Fehlen relevanter Einflussfaktoren ist verheerend: Es gilt *Junk in - junk out*. Hingegen schwächt das Vorhandensein überzähliger irrelevanter Variablen nur die statistische Schärfe der Aussagen.
- *Abhilfe:* Hier gilt: “Gefahr erkannt, Gefahr gebannt”, da man danach einfach die Fehlenden zum Modell hinzufügen und die Überflüssigen eliminieren kann. Im Modell ggf vorhandene irrelevante Variablen lassen sich erkennen, indem man nacheinander alle Variablen mit dem in Abschnitt 2.6.9 besprochenen F-test auf Relevanz checkt. Das Überprüfen auf fehlende relevante Einflussfaktoren ist schwieriger⁴. Hat man einen Kandidaten, lässt sich dessen Relevanz ebenfalls mit dem F-Test feststellen.

2. **Linearität:** *Der wahre Sachverhalt ist, zumindest in geeignet transformierten Variablen, linear*

- *Folgen einer Spezifikationsverletzung:* Bei nichtlinearen Sachverhalten muss ein lineares Modell versagen (Abb. 2.2).
- *Abhilfe:* Sind die verschiedenen durch die x_i beschriebenen Einflüsse jedoch separierbar (Quasilinearität), kann man das Modell linearisieren (Abb. 2.3).

Dabei kann durchaus die Anzahl der exogenen Variablen verändert werden. Beispiel ergibt sich aus kinematischen Überlegungen für den Anhalteweg Y

³Bei mehr als zwei Ausprägungen (z.B. Erdgas als 3. Ausprägung) wird die Sache komplizierter. Allgemein benötigt man bei M Ausprägungen $M - 1$ binäre Variable, welche jeweils bei einer bestimmten Ausprägung den numerischen Wert 1 und ansonsten null annehmen. Der zugehörige Modellparameter beschreibt die Veränderung des erklärten Anteils der endogenen Variable bezüglich der Ausprägung, welcher *keiner* binäre Variable zugeordnet ist. Näheres siehe Abschnitt 2.7.

⁴Man weiß ja nicht genau, nach was man sucht; das Gesuchte könnte überall und nirgends sein!

2 Stetige ökonomische Modelle

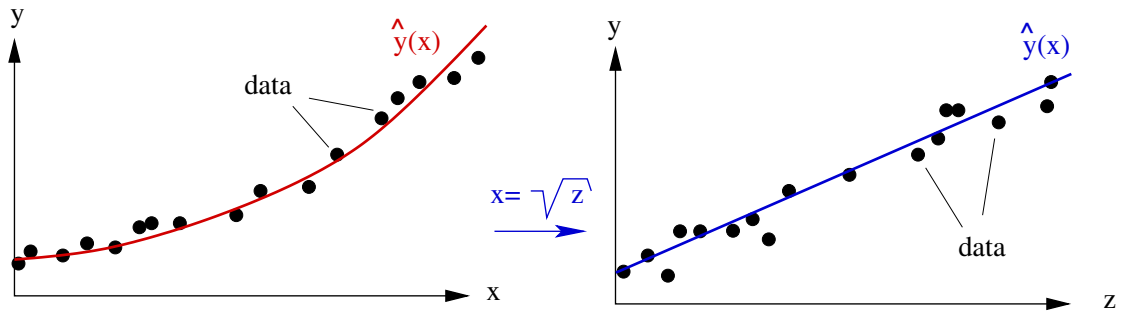


Abbildung 2.3: Linearisierung durch Transformation der exogenen Variablen.

eines Fahrzeugs (=Bremsweg zuzüglich der während der Reaktionszeit zurückgelegte Weg) in Abhängigkeit der Geschwindigkeit x das Modell

$$Y(x) = \beta_1 x + \beta_2 x^2 + \epsilon$$

Mit der Transformation $x = z_1$ und $x^2 = z_2$ erhält man daraus ein lineares Modell

$$Y(z) = \beta_1 z_1 + \beta_2 z_2 + \epsilon$$

mit *zwei* exogenen Variablen. Hierbei charakterisiert z_1 den Einfluss der Reaktionszeit (β_1 entspricht direkt der Reaktionszeit) und z_2 den Einfluss der Bremsverzögerung ($0.5/\beta_2$ entspricht der Bremsverzögerung).

Wie in Kapitel 4 gezeigt werden wird, sind Modelle mit diskreten endogenen Variablen, insbesondere Wahlmodelle, *immer* irreduzibel nichtlinear. Daraus folgt:

Modelle diskreter Wahlentscheidungen (Kapitel 4) erfüllen prinzipiell nicht die Spezifikationsbedingungen dieses Abschnitts. Damit kommt zur Parameterschätzung dieser Modelle die Regressionsmethode nicht in Frage.

3. Homogenität: *Die Parameterwerte sind für alle Systeme innerhalb des Anwendungsbereichs dieselben.* Es gibt keine Strukturbrüche wie den in Abb. 2.4. Strukturbrüche können unter anderem durch Änderungen des Erhebungsgebietes (wie in dieser Abbildung), aber auch durch Umstellung der statistischen Berechnungsmethode (wie bei den Arbeitslosenzahlen im Jahre 2008) hervorgerufen werden.

- *Folgen einer Spezifikationsverletzung:* Schwerwiegend, da zumindest für Teile des Datensatzes die Schätzer systematische Fehler aufweisen.
- *Abhilfe:*
 - (i) Im einfachsten Fall modelliert man die Verhältnisse vor und nach einem Strukturbruch separat, wobei natürlich die Stichprobengröße kleiner wird. Dazu muss die Abgrenzung des Strukturbruchs bekannt sein, z.B.

2 Stetige ökonometrische Modelle

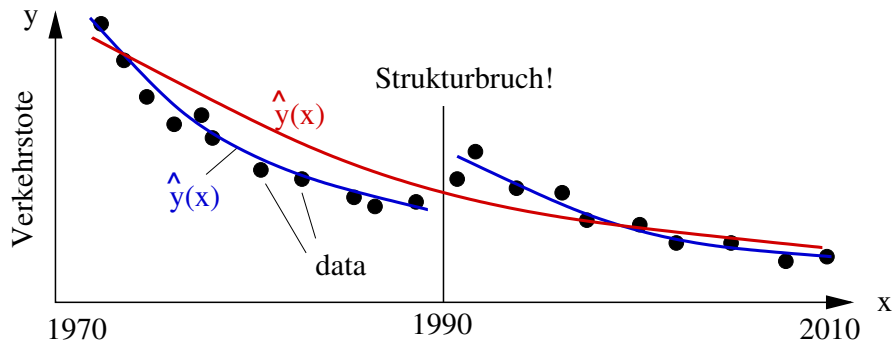


Abbildung 2.4: Strukturbruch, welcher beispielsweise bei der Zahl der Verkehrstoten in der Bundesrepublik (vor 1990) und Deutschland (ab 1990) auftreten könnte.

eine Zeitreihe mit Perioden vor und nach einer drastischen Änderung wie einer Revolution.

- (ii) Wird der Strukturbruch durch Veränderung der Grundgesamtheit wie in Abb. 2.4 verursacht, kann eine Neudefinition der endogenen Variablen als spezifische Größe (Verhältniszahl) helfen, beispielsweise könnte man bei der Analyse der Verkehrstoten nicht die Zahl selbst, sondern eine *Risikokennziffer* wie Verkehrstote/Jahr/Einwohner oder Verkehrstote pro gefahrenen Kilometer untersuchen.
- (iii) Bei bekannter Abgrenzung kann man den Strukturbruch auch durch einen neuen Einflussfaktor mit dazugehöriger binärwertiger exogener Pseudovariable (0=vor, 1=nach dem Strukturbruch) beschreiben. Dies formalisiert im Wesentlichen die bei (i) besprochene separate Modellierung der beiden Perioden, geht dabei aber sparsamer mit den Daten um.

2.3.2 Statistische Spezifikation

Die *unbestimmten Anteile* müssen ebenfalls mehrere Bedingungen erfüllen, damit kontrollierbare statistischer Aussagen möglich sind.

1. Verschwindender Erwartungswert:

$$E(\epsilon) = E(\epsilon_i) = 0.$$

- *Folgen einer Spezifikationsverletzung*: Schwerwiegend, da sie zu systematischen Fehlern führen.
- *Abhilfe*: Im linearen Fall wird dieser Fehler (Abb. 2.5 links) automatisch durch die Parameterschätzung korrigiert. Auch allgemein ist diese Bedingung am einfachsten erfüllbar, da man einen nichtverschwindenden Anteil am Erwartungswert als Konstante auf den erklärten Anteil $\hat{y}(x)$ transferieren kann.

2 Stetige ökonomische Modelle

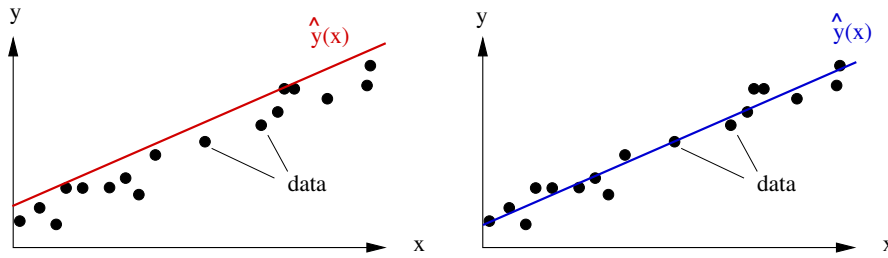


Abbildung 2.5: Der Erwartungswert der Störgröße muss verschwinden.

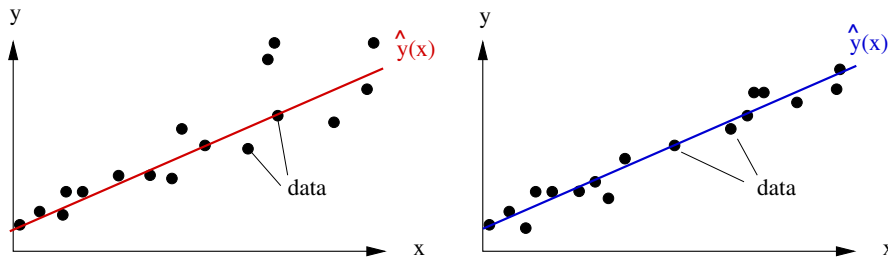


Abbildung 2.6: Heteroskedastizität (links) vs. Homoskedastizität (rechts).

2. **Homoskedastizität:** Die Varianz der Störgröße darf weder von den exogenen Variablen (wie in Abb. 2.6 links) noch vom Messindex i abhängen,

$$V(\epsilon) = V(\epsilon_i) = \sigma_\epsilon^2 = \text{const.} \quad (2.4)$$

Diese Bedingung ist gewissermaßen die Entsprechung der Homogenitätsbedingung der funktionalen Spezifikation. Ein klassischer Sachverhalt, bei dem diese Bedingung nicht erfüllt ist (*Heteroskedastizität*), sind Börsenkurse in Abhängigkeit der Zeit als exogenen Variablen. Solche Kurse weisen “Volatilitätspeaks” (Maxima der Varianz) in Boom- und Crashphasen auf.

- *Folgen einer Spezifikationsverletzung:* relativ harmlos, da keine systematischen Fehler auftauchen. Lediglich die Unschärfe der Aussagen ist suboptimal.
- *Abhilfe:* Entweder durch Transformation der endogenen Variablen (z.B. durch Logarithmierung in Abb. 2.6 links) oder durch spezielle Analysemethoden (die in dieser Vorlesung nicht behandelt werden).

3. **Unabhängigkeit:** Die Störgrößen dürfen weder untereinander noch von exogenen oder endogenen Variablen abhängen. Als notwendige Bedingung folgt daraus $E(\epsilon_i \epsilon_j) = 0$ falls $i \neq j$. Die Bedingungen 2 und 3 kann man zusammen formulieren durch

$$E(\epsilon_i \epsilon_j) = \begin{cases} V(\epsilon) = \sigma_\epsilon^2 & i = j \\ 0 & \text{sonst.} \end{cases} \quad (2.5)$$

2 Stetige ökonomische Modelle

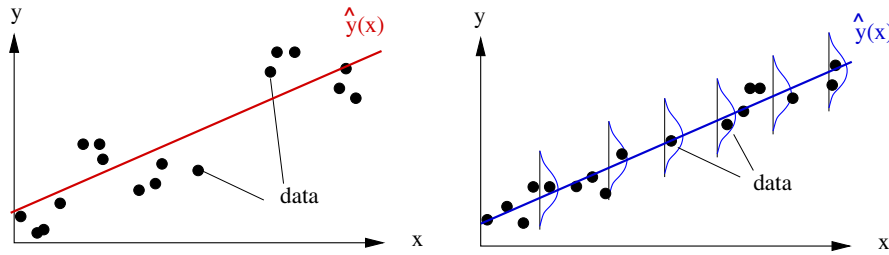


Abbildung 2.7: Korrelierte Störgrößen (links) vs. unkorrelierte Störgrößen (rechts).

Insbesondere dürfen die Störgrößen keine systematische Abhängigkeit von den exogenen Variablen (Abb. 2.7 links) und auch nicht vom Messindex i haben. Diese Bedingung ist beispielsweise bei Arbeitslosenzahlen in Abhängigkeit der Zeit verletzt, wenn man diese durch ein einfaches lineares Modell beschreibt: In diesem Modell wäre der unbestimmte Anteil im Sommer systematisch negativ und im Winter positiv.

- *Folgen einer Spezifikationsverletzung:* Eine Abhängigkeit des Zufallsanteils von y oder x impliziert systematische Fehler und ist damit schwerwiegend!
- *Abhilfe:* Nur durch ein erweitertes Modell, beispielsweise durch eine zusätzliche saisonale Analyse bei den Arbeitslosendaten:

Bemerkung: Sind die statistischen Spezifikationsbedingungen 1-3 alle erfüllt, kann man die exogenen (nicht die endogenen!) Variablen beliebig transformieren, ohne dass sich daran etwas ändert. Das heißt, man kann ein statistisch korrekt spezifiziertes quasilineares Modell problemlos in ein lineares transformieren, ohne dass die Spezifikation zerstört wird.

4. **Normalverteilung:** Die Störgrößen sollten eine Gauß'sche Normalverteilung besitzen und nicht etwa eine bimodale wie in Abb. 2.8 links.
 - *Folgen einer Spezifikationsverletzung:* Gering. Die Schätzer bleiben auch bei Verletzung dieser Annahme unverzerrt und von optimaler Trennschärfe im Sinne einer minimalen Varianz (effektive Schätzer), zumindest, wenn die anderen statistischen Spezifikationen zutreffen). Lediglich Konfidenzintervalle von Schätzgrößen und Fehlerwahrscheinlichkeiten von Tests (Abschnitt 2.6) sind geringfügig verfälscht, da die in diesen Abschnitten zu besprechenden Standardverteilungen der Schätz- und Testgrößen (wie χ^2 und Student-t-Verteilung) dann nicht exakt zutreffen.
 - *Abhilfe:* Nicht möglich.

2 Stetige ökonometrische Modelle

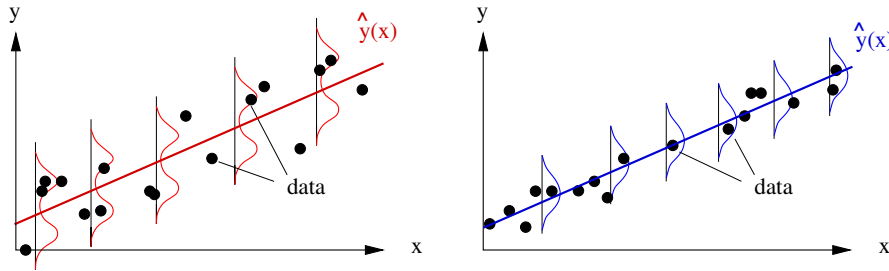


Abbildung 2.8: Bimodale Verteilung der Störgrößen (links) *vs.* gaußverteilte Störgrößen (rechts).

Zusammenfassung

Sind die Spezifikationsbedingungen 2 und 3 erfüllt, so haben die unbestimmten Anteile die sogenannte *i.i.d.-Eigenschaft*:

Zufallsgrößen haben die **i.i.d.-Eigenschaft**, wenn sie unabhängig voneinander sind (*independent*) und eine identische Verteilung aufweisen (*identically distributed*).

Alle vier statistischen Spezifikations-Bedingungen zusammen kann man kompakt durch folgende Bedingung schreiben:

$$\epsilon_i \sim i.i.d. N(0, \sigma_\epsilon^2) \quad (2.6)$$

Hierbei bedeutet die Tilde \sim "gehört einer" und $N(\mu, \sigma^2)$ bezeichnet die Gaußverteilung (Normalverteilung) mit Erwartungswert μ und Varianz σ^2 .

Verständnisfrage:

Machen Sie sich klar, dass die Bedingung (2.6) *alle vier* statistischen Spezifikations-Bedingungen enthält.

2.3.3 Datenspezifikation

1. **Die exogenen Variablen sind keine Zufallsvariablen.** Dies ist bei der ökonomischen Datenerhebung vordergründig meist *nicht* erfüllt. Beispielsweise sind bei einer Personenbefragung aus einer Zufallsstichprobe typische exogene Variable wie Geschlecht, Einkommen, Alter, Länge des Arbeitsweges etc Zufallsvariablen. Häufig kann man jedoch einen gerichteten Ursache-Wirkungs-Zusammenhang postulieren: Das Einkommen und das Geschlecht sind möglicherweise relevante Einflussfaktoren (=exogene Variablen) bei der Frage nach der Nutzungshäufigkeit von Fahrrädern wohingegen das Rad nicht das Einkommen (abgesehen vielleicht bei

2 Stetige ökonomische Modelle

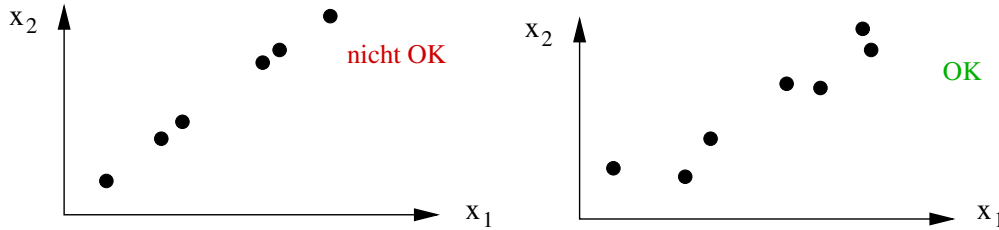


Abbildung 2.9: Zur korrekten Datenspezifikation.

Fahrradhändlern) oder das Geschlecht bestimmt. In diesem Fall kann man ausnutzen, dass das lineare Modell aufgrund der bisherigen Spezifikationen für alle Werte der exogenen Variablen im Anwendungsbereich gleichermaßen gilt, so dass der konkrete Wert von \mathbf{x} (ob zufällig oder deterministisch) über die Anstiegsparameter automatisch auch in y berücksichtigt ist.

2. **Keine Multikollinearität:** Jede exogene Variable x_j muss eingeständige Information liefern, darf also nicht für alle Datensätze als Linearkombination anderer exogener Variablen darstellbar sein. x_k wäre beispielsweise dann **multikollinear** zu allen $x_j, j \neq k$, wenn sich ihre Werte x_{ik} für alle Datensätze i als Linearkombination der anderen exogenen Variablen schreiben ließen, also

$$x_{ik} = \sum_{j \neq k} c_j x_{ij}, \quad \forall i = 1, \dots, n \quad (2.7)$$

mit von i unabhängigen Konstanten c_j . Mathematisch zeigt sich dies daran, dass die deskriptive Varianz-Kovarianz-Matrix

$$(\mathbf{S})_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (2.8)$$

einen Rang kleiner $J + 1$ hat bzw. $\text{Det } \mathbf{S} = 0$ ist. Nichtperfekte Korrelationen (Abb. 2.9 unten) sind aber ausdrücklich zugelassen!

Bei einer exogenen Variable bedeutet dies, dass nicht alle Datenpunkte denselben Wert dieser Variablen haben dürfen (Abb. 2.9 links oben). Bei zwei exogenen Variablen dürfen die exogenen Datenpunkte nicht perfekt auf einer Geraden liegen (Abb. 2.9 rechts oben), eine Korrelation ungleich null ist aber sehr wohl erlaubt (Abb. 2.9 unten).

3. Die zur Schätzung bzw. Modellkalibrierung herangezogenen Daten stammen aus einer Stichprobe vom Umfang $n > J + 1$
 - *Folgen einer Verletzung der Datenspezifikation 2 und 3:* Dann versagt automatisch die in Abschnitt 2.4.1 beschriebene Parameterschätzung. Man kommt also erst gar nicht in Versuchung, Ergebnisse falsch zu interpretieren. Bei sehr

2 Stetige ökonometrische Modelle

hoher Korrelation zweier exogenen Variablen werden die diesen Variablen entsprechenden Schätzer sehr unscharf.

- *Abhilfe:* Bei zu wenig Datenpunkten ist die Abhilfe offensichtlich: Mehr Daten!. Bei Multikollinearität ist das einfachste Mittel, abhängige exogene Variablen zu identifizieren und zu eliminieren. Beispielsweise ist bei Fußwegen in der Ebene die Entfernung x_1 und die Reisezeit x_2 direkt proportional, also eliminiert man eine der Variablen. Bei ÖPNV-Untersuchungen ausschließlich in einer Zone (z.B. Dresden, Zone 1) ist der Fahrpreis immer derselbe, also muss er eliminiert werden. Auch bei extrem hoher Korrelation zweier exogener Variablen j und k ($|r_{jk}| > 0.99$) lassen sich die Einflüsse nicht separieren und es ist besser, diese zu einem Verbundfaktor zusammenzufassen.

Verständnisfrage:

Warum impliziert die Bedingung 2 automatisch $n \geq J + 1$. Warum muss man aber auch $n = J + 1$ ausschließen?

Verständnisfrage:

Ökonometrische Modelle enthalten oft viele exogenen Variablen und es ist bei Multikollinearität nicht offensichtlich, welche davon der/die "Übeltäter" ist/sind, da man ihnen die Multikollinearität nach Gl. (2.7) nicht mehr ansieht. Wie könnte man das dennoch rein schematisch herausfinden?

2.3.4 Zusammenfassung

Alle drei Kategorien von Spezifikationen (Modellspezifikation, statistische Spezifikation und Datenspezifikation) bezeichnet man zusammen als **Gauß-Markow-Annahmen**. Wenn im Folgenden nichts weiter gesagt wird, wird angenommen, dass alle Gauß-Markow-Annahmen erfüllt sind.

2.4 Parameterschätzung

Eine Modellkalibrierung bzw. Parameterschätzung durch Vergleich der Modellaussagen mit den Daten, kann wegen der unvermeidlichen nichterklärten Elemente nicht zum Ziel haben, eine *exakte* Übereinstimmung des Modells mit den Daten zu liefern. Vielmehr gilt es, durch Variation der Modellparameter ein noch zu definierendes Fehlermaß zu minimieren (Regression) bzw. die Wahrscheinlichkeit für das Zutreffen bei gegebenen Beobachtungen zu maximieren (Maximum-Likelihood-Methode).

Zur Durchführung der Kalibrierung werden $n > J + 1$ Messungen bzw. Beispielsinstanzen des zu beschreibenden Sachverhalts als Stichprobe vorausgesetzt. Jedes Element i der Stichprobe bzw. Messung i enthält dabei folgenden Satz von Größen:

2 Stetige ökonomische Modelle

- den Wert y_i der abhängigen bzw. erklärten Variablen bei der Messung i ,
- die dazugehörigen Messwerte x_{mi} der erklärenden Variablen x_m ,
- möglichst auch noch Merkmalsausprägungen weiterer möglicherweise relevanter Variablen, mit denen man das Modell bei Bedarf erweitern kann.

Daraus ergeben sich aus dem linearen Modell (2.2) die Systemgleichungen

$$y_i = \beta_0 + \sum_{j=1}^J x_{ij}\beta_j + \epsilon_i. \quad (2.9)$$

Verständnisfrage:

Ohne die unbestimmten Terme ϵ_i hätte man ein überbestimmtes Gleichungssystem. Warum ist dies in diesem Fall erwünscht und notwendig?

Wegen der Überbestimmtheit lässt sich dieses System ohne Fehlerterme ϵ_i nicht lösen, sondern es können letztere nur minimiert werden. Darauf begründen die zwei wichtigsten Methoden zur Modellkalibrierung:

- Bei der **Methode der kleinsten Fehlerquadrate**, englisch auch als *Least Squares* (LQ)-Methode bzw. **Least Squared Errors (OLS)**-Methode bzw. **Regressionsmethode** bezeichnet, werden, wie der Name schon sagt, die Fehlerquadratsummen bezüglich der Modellparameter minimiert.
- Bei der **Maximum-Likelihood (ML-) Methode** wird in Abhängigkeit der Parameter die Wahrscheinlichkeit dafür bestimmt, dass das Modell alle Beobachtungen voraussagt und diese Wahrscheinlichkeitsfunktion, auch **Likelihoodfunktion** genannt, bezüglich der Parameter maximiert.

Beide Methoden stehen zueinander in Beziehung.⁵ Im allgemeinen multivariaten Fall mit korrelierten stochastischen Störgrößen (es sind also nicht alle Bedingungen der statistischen Spezifikation erfüllt) ist die Maximum-Likelihood-Methode dennoch anwendbar. Dafür wird für die OLS-Methode Bedingung 4 der statistischen Spezifikation (gaußverteilte Störgrößen) nicht benötigt. Zusammenfassend ist für Modelle mit diskreten endogenen Variablen (diskrete Wahlmodelle, Kapitel 4) nur die Maximum-Likelihood-Methode sinnvoll, während für lineare Modelle mit kontinuierlicher endogener Variable die OLS-Methode vorteilhaft ist und meist angewandt wird. Im Folgenden wird die OLS-Methode betrachtet, während die ML-Methode im Abschnitt 4.6 beschrieben werden wird.

⁵Sind die vier Bedingungen der statistischen Spezifikation erfüllt, so sind die Ergebnisse beim Schätzen der Modellparameter β_j (nicht aber etwa beim Schätzen der Varianz) dieselben.

2 Stetige ökonometrische Modelle

Beispiel zur Begleitung: Hotel-Auslastung

Die verschiedenen Statistischen Methoden der folgenden Abschnitte werden anhand folgendes Sachverhalts erläutert: Gegeben ist eine Untersuchung, bei der von $n = 12$ Hotels die Merkmale "Sternzahl" x_1 , Preis x_2 in Euro/Nacht und Bettenauslastung y in Prozent erhoben wurden:

Zahl der Sterne	1	1	1	2	2	2	3	3	3	4	4	4
Preis (Euro/Nacht)	15	31	40	34	50	58	67	72	84	82	98	116
Auslastung (%)	42	38	24	76	52	40	90	77	62	90	82	68

Bei der Untersuchung soll das lineare Zweifachregressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

zum Einsatz kommen.

2.4.1 Methode der kleinsten Fehlerquadrate

Gemäß dem Namen dieser Methode wird die Fehlerquadratsumme (SSE) als Funktion der Menge $\beta = \{\beta_0, \dots, \beta_J\}$ der Parameterwerte betrachtet,

$$S(\beta) = \sum_{i=1}^n \epsilon_i = \sum_{i=1}^n [y_i - \hat{y}(x_{i1}, \dots, x_{iJ}; \beta_0, \dots, \beta_J)]^2, \quad (2.10)$$

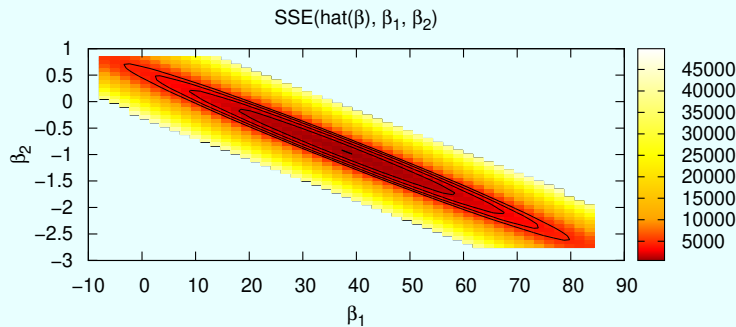
und bezüglich der *Parameter* minimiert:

$$\hat{\beta} = \arg \min_{\beta} S(\beta). \quad (2.11)$$

2 Stetige ökonomische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die SSE als Funktion der Parameter β_1 und β_2 für den bereits geschätzten Parameterwert $\hat{\beta}_0 = 25.5$ des Achsenabschnitts ist in folgender Abbildung zu sehen:



Als notwendige Bedingungen für das Minimum müssen alle Ableitungen verschwinden, $\frac{dS}{d\beta_j} = 0$. Aufgrund der speziellen Struktur von (2.10) (Summe von Quadraten) kann man in diesem Fall sogar zeigen, dass die Bedingungen verschwindender Ableitungen hinreichend sind und im linearen Fall darüberhinaus ein eindeutiges Ergebnis liefern. Dies ergibt $J + 1$ Gleichungen für die $J + 1$ unbekannt Parameterwerte, welche bei linearen und quasilinearen Modellen linear sind und sich daher leicht lösen lassen. Eine analytische Lösung im nichtlinearen Fall ist nur in Ausnahmefällen möglich. Da die quadratische Form aber meist zu "gutmütigen" Nichtlinearitäten führt, ist eine numerische Nullstellensuche, beispielsweise durch das später im Abschnitt 4.8 vorgestellte **Newton-Raphson-Verfahren**, meist problemlos möglich.

Mit Hilfe der dadurch minimierten Fehlerquadratsumme S_{\min} lässt sich anschließend auch die Stärke des stochastischen Beitrags, d.h. die **Residualvarianz** σ_ϵ^2 abschätzen:

$$\hat{\sigma}_\epsilon^2 = \frac{S(\hat{\beta})}{n - J} = \frac{S_{\min}}{n - J} \quad (2.12)$$

und damit auch ein Maß für die Güte des Modells gewinnen: Je näher das (deskriptive) **Bestimmtheitsmaß**

$$B = R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = 1 - \frac{s_\epsilon^2}{s_y^2}, \quad s_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.13)$$

2 Stetige ökonometrische Modelle

bzw. das **Korrigierte Bestimmtheitsmaß**⁶ (vgl. weiter unten)

$$\tilde{B} = \bar{R}^2 = 1 - \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_y^2}, \quad \hat{\sigma}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.14)$$

an den Maximalwert $B = 1$ herankommt, desto besser ist das Modell.

Aufgabe: Trivialmodell

In einem “Trivialmodell” ist die deterministische Teil der endogenen Variable durch die Gleichung

$$\hat{y} = a$$

mit dem Modellparameter a gegeben. Wie lautet der Ausdruck für den kalibrierten Wert von a als Funktion der y_i und x_{ij} ? Geben Sie auch die Ausdrücke für die Residualvarianz und das Bestimmtheitsmaß an. Welche Bedeutung hat die Residualvarianz in diesem Fall?

Lösung:

Die Minimierung der Fehlerquadratsumme führt auf $a = \bar{y}$. Daraus ergibt sich mit (2.12) und (2.13) $s_\epsilon^2 = s_y^2$ sowie $\hat{\sigma}_y^2 \hat{\sigma}_\epsilon^2 = \hat{\sigma}_y^2$, also die Bestimmtheitsmaße $B = 0$ und $\tilde{B} = 0$. Das Bestimmtheitsmaß gibt also an, in welchem Maße das Modell eine bessere Aussage liefert als das Trivialmodell. Dies ist analog zu den Modellen der Wettervorhersage, die sich auch am Trivialmodell (“Das Wetter bleibt wie es ist”) messen lassen müssen.

2.4.2 Parameterschätzung I: Formulierung mit Summen

Aus den Systemgleichungen (2.9) des linearen Modells errechnet sich direkt die Fehlerquadratsumme⁷

$$S(\beta) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{m=1}^J x_{im} \beta_m \right)^2, \quad (2.15)$$

Man beachte, dass die Fehlerquadratsumme (2.15) nur von den Modellparametern abhängt, da die y_i und x_{im} ja feste Messwerte darstellen! Ableiten nach β_0 und Nullsetzen liefert eine erste Gleichung für die OLS-Schätzer $\hat{\beta}_j$ der Parameter β_j , $j = 0, \dots, J$:

$$\hat{\beta}_0 = \bar{y} - \sum_{m=1}^J \hat{\beta}_m \bar{x}_m. \quad (2.16)$$

⁶Eigentlich ist dies die relevante Definition, da es, im Gegensatz zu B , auf erwartungstreuen Schätzern beruht. Häufig wird jedoch das einfache deskriptive Bestimmtheitsmaß B verwendet.

⁷Im Folgenden wird absichtlich nicht über j sondern über m summiert, da später nach β_j abgeleitet werden soll.

2 Stetige ökonomische Modelle

Die arithmetischen Mittel der endogenen und exogenen Variablen sind dabei definiert durch

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}. \quad (2.17)$$

Die Extremumsbedingungen $\frac{\partial S}{\partial \hat{\beta}_j} = 0$ liefern⁸

$$\sum_{i=1}^n \left(y_i x_{ij} - \hat{\beta}_0 x_{ij} - \sum_{m=1}^J \hat{\beta}_m x_{im} x_{ij} \right) = 0.$$

Setzt man nun (2.16) für $\hat{\beta}_0$ ein, erhält man als Kalibrierungsbedingungen der LQ-Methode ein lineares Gleichungssystem für die Koeffizienten $\hat{\beta}_1$ bis $\hat{\beta}_J$:

$$\sum_{m=1}^J s_{jm} \hat{\beta}_m = s_{jy} \quad (2.18)$$

mit den Elementen der bereits in Gl. (2.8) definierten deskriptiven Kovarianzmatrix der erklärenden Variablen,

$$s_{jm} = \frac{1}{n} \sum_{i=1}^n x_{ij} x_{im} - \bar{x}_j \bar{x}_m = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{im} - \bar{x}_m)$$

und der deskriptiven Kovarianz zwischen x_j und y ,

$$s_{jy} = \frac{1}{n} \sum_{i=1}^n x_{ij} y_i - \bar{x}_j \bar{y} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y}). \quad (2.19)$$

Die deskriptive Kovarianzmatrix \mathbf{S} ⁹ mit den Elementen nach Gl. (2.8) spielte bereits bei der Datenspezifikation (Abschnitt 2.3.3) eine Rolle. Dort wurde gefordert, dass $\text{Det} \mathbf{S} \neq 0$. Hier sieht man nun, dass andernfalls (2.18) nicht gelöst werden kann: Diese Formel stellt nämlich ein lineares inhomogenes Gleichungssystem dar, welches im Allgemeinen nur gelöst werden kann, wenn die Determinante der Koeffizientenmatrix, also hier \mathbf{S} , ungleich null ist. Dass \mathbf{S} die Koeffizientenmatrix ist, sieht man noch deutlicher, wenn man (2.18) in Matrix-Vektor-Notation schreibt:

$$\begin{pmatrix} s_{11} & \cdots & s_{1J} \\ \vdots & & \vdots \\ s_{j1} & \cdots & s_{jJ} \end{pmatrix} \cdot \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_J \end{pmatrix} = \begin{pmatrix} s_{1y} \\ \vdots \\ s_{jy} \end{pmatrix}. \quad (2.20)$$

Allerdings verwendet man bei Verwendung der Matrix-Vektor-Notation meist eine modifizierte Formulierung, welche im Abschnitt 2.4.5 beschrieben wird.

⁸Man beachte, dass man den Index j der Ableitung nach β_j vom Summationsindex m unterscheiden muss!

⁹Bitte nicht die matrix \mathbf{S} mit der skalaren Fehlerquadratsumme S verwechseln!

2 Stetige ökonomische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die Mittelwerte der deskriptiven Variablen x_1 (Zahl der Sterne), x_2 (Preis pro Nacht) und y (Auslastung in Prozent) betragen

$$\bar{x}_1 = 2.5, \quad \bar{x}_2 = 62.25, \quad \bar{y} = 61.75.$$

Damit sind die deskriptiven Varianzen und Kovarianzen

$$s_{11} = 1.25, \quad s_{12} = s_{21} = 29.63, \quad s_{22} = 818.2$$

sowie

$$s_{1y} = 19.54, \quad s_{2y} = 352.6.$$

und die OLS-Schätzer der Komponenten von β mit (2.18) (vgl. explizite Formel (2.25) weiter unten):

$$\begin{aligned} \hat{\beta}_1 &= \frac{s_{1y}s_{22} - s_{2y}s_{12}}{s_{11}s_{22} - s_{12}^2} = 38.2, \\ \hat{\beta}_2 &= \frac{s_{2y}s_{11} - s_{1y}s_{12}}{s_{11}s_{22} - s_{12}^2} = -0.953, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 = 25.5. \end{aligned}$$

2.4.3 Additionsregel, Residualvarianz und Bestimmtheitsmaß

Auch für mehrere exogene Variablen gilt die allgemeine Additionsregel der deskriptiven Varianzen der endogenen Variable:

$$\underbrace{s_y^2}_{\text{Gesamtvarianz}} = \underbrace{s_{\hat{y}}^2}_{\text{Erklärte Varianz}} + \underbrace{s_\epsilon^2}_{\text{Residualvarianz}} \quad (2.21)$$

Nachdem man die Modellparameter durch Lösen des linearen Gleichungssystems geschätzt hat, kann man mit ihrer Hilfe und der Additionsregel die deskriptive, durch das Modell erklärte Varianz $s_{\hat{y}}^2$ als Summe von gewichteten deskriptiven Kovariationen der endogenen mit den exogenen Variablen ausdrücken:¹⁰

$$s_{\hat{y}}^2 = \sum_{j=1}^J \hat{\beta}_j s_{jy} \quad (2.22)$$

¹⁰Die Herleitung der Additionsrelation und dieser Beziehung ist allerdings nur in der Formulierung mit Matrizen einigermaßen einfach, siehe Abschnitt 2.4.5.

2 Stetige ökonomische Modelle

Man beachte, dass $\hat{\beta}_0$ in dieser Summe nicht vorkommt! Mit der Additionsregel (2.21) folgt daraus für das deskriptive Bestimmtheitsmaß (2.13) direkt

$$B = \frac{\sum_{j=1}^J \hat{\beta}_j s_{jy}}{s_y^2}, \quad (2.23)$$

während man für die entsprechenden erwartungstreuen Größen der induktiven Statistik, Gl. (2.14), den Umweg über die Residualvarianz gehen muss (zur Frage des Nenners $n - J - 1$ folgender Formel siehe Abschnitt 2.6.1):

$$\tilde{B} = 1 - \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_y^2}, \quad \hat{\sigma}_\epsilon^2 = \frac{n}{n - J - 1} (s_y^2 - s_y^2) = \frac{S_{\min}}{n - J - 1} \quad \hat{\sigma}_y^2 = \frac{n}{n - 1} s_y^2. \quad (2.24)$$

Hierbei ist $S_{\min} = S(\hat{\beta})$ das Minimum der Fehlerquadratsumme.

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Minimale Fehlerquadratsumme:

$$S_{\min} = S(\hat{\beta}) = 498.2$$

deskriptive und induktive Residualvarianzen ($n = 12$ Hotels):

$$s_\epsilon^2 = \frac{S_{\min}}{12} = 41.5, \quad \hat{\sigma}_\epsilon^2 = \frac{S_{\min}}{12 - 2 - 1} = \frac{S_{\min}}{9} = 55.4.$$

deskriptive und induktive Gesamtvarianzen:

$$s_y^2 = \frac{1}{12} \sum_{i=1}^{12} (y_i - \bar{y})^2 = 452, \quad \hat{\sigma}_y^2 = \frac{12}{11} s_y^2 = 493$$

Damit ergibt sich für die deskriptiven und korrigierten Bestimmtheitsmaße

$$B = 1 - \frac{s_\epsilon^2}{s_y^2} = 0.908, \quad \tilde{B} = 1 - \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_y^2} = 0.888.$$

2.4.4 Spezialfälle: Eine und zwei exogene Variable

Aus der linearen Regression (2.18) bzw. (2.20) und Gl. (2.16) ergeben sich im Falle zweier exogener Variablen ($J = 2$) die Schätzer der linearen Anstiegsparameter und der Konstante (*intercept*) zu

2 Stetige ökonometrische Modelle

$$\begin{aligned}\hat{\beta}_1 &= \frac{s_{1y}s_{22} - s_{2y}s_{12}}{\det(\mathbf{S})}, \\ \hat{\beta}_2 &= \frac{s_{2y}s_{11} - s_{1y}s_{21}}{\det(\mathbf{S})}, \\ \hat{\beta}_0 &= \bar{y} - \beta_1\bar{x}_1 - \beta_2\bar{x}_2.\end{aligned}\tag{2.25}$$

Aufgabe:

Zeigen Sie, dass (2.18) bei einer einzigen erklärenden Variablen $x_1 = x$ sowie der Gleichsetzung $\hat{\beta}_0 = \hat{a}$ und $\hat{\beta}_1 = \hat{b}$ die aus den einführenden Statistikvorlesungen bekannten OLS-Schätzer

$$\hat{b} = \hat{\beta}_1 = s_{1y}/s_{11} = s_{xy}/s_{xx}, \quad \hat{a} = \hat{\beta}_0 = \bar{y} - \hat{b}\bar{x}\tag{2.26}$$

der Einfachregression liefert. Kann man nun auch sehen, warum B manchmal als R^2 in der Bedeutung “Quadrat der Korrelation zwischen x und y ” bezeichnet wird?

2.4.5 Parameterschätzung II: Formulierung mit Vektoren und Matrizen

Die abstrakte Formulierung mit Vektoren und Matrizen ist zunächst etwas ungewohnt, sie hat aber, wenn man einmal die wichtigsten Matrizenregeln kennt, durchaus Vorteile:

- Die Formulierung ist kompakter und einheitlicher,
- die induktive Statistik lässt sich mit dieser Formulierung einfacher und eleganter durchführen, insbesondere die Berechnung der statistischen Eigenschaften der Parameterschätzer.
- die Formulierung entspricht den Standard-Softwarepaketen. Die numerische Lösung ist einfacher zu programmieren.

Dem stehen aber auch Nachteile gegenüber:

- Die Formulierung ist abstrakt und unanschaulich,
- Die Rechnung “per Hand” ist komplizierter. Beispielsweise haben die Gleichungssysteme eine “Unbekannte” mehr

Man sollte also beide Herangehensweisen kennen. Sowohl konzeptionell als auch praktisch ist mal die eine, mal die andere Formulierung effizienter.

2 Stetige ökonomische Modelle

Systemgleichungen

Der Trick bei dieser Darstellung ist, den konstante Achsenabschnitt β_0 formal zu einem Anstiegsparameter zu transformieren, indem man die *Pseudo-Variable* $x_0 = 1$ einführt. Das lineare Modell wird dann zu

$$Y(\mathbf{x}) = \sum_{j=0}^J \beta_j x_j + \epsilon = \boldsymbol{\beta}' \cdot \mathbf{x} + \epsilon \quad (2.27)$$

und die Systemgleichungen zu

$$\mathbf{Y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2.28)$$

bzw. ausgeschrieben

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{10} & \cdots & x_{1J} \\ \vdots & & \vdots \\ x_{n0} & \cdots & x_{nJ} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_J \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (2.29)$$

Der “Trick” mit der Pseudovariablen führt also zu einer (zumindest formal) einfacheren Repräsentation der Modell- und Systemgleichungen. Hierbei bedeuten

- \mathbf{Y} und $\boldsymbol{\epsilon}$ sind Spaltenvektoren mit n Elementen. Y_i bzw. ϵ_i geben die Werte der endogenen Variablen Y bzw. der Restabweichung ϵ bei der i -ten Messung an.
- $\boldsymbol{\beta}$ ist ein Spaltenvektor mit $J + 1$ Elementen; β_j mit $j = 0, 1, \dots, J$ gibt den Wert des j -ten Parameters an.
- Die **Datenmatrix** \mathbf{X} ist eine $n \times (J + 1)$ -Matrix. Das Element x_{ij} gibt den Wert der j -ten exogenen Variablen bei der i -ten Messung an. Da die Pseudovariablen $x_0 = 1$, haben automatisch auch deren “Messwerte” x_{i0} den Wert 1.¹¹

Fehlerquadratsumme

Auch diese lässt sich kompakt in Vektor-Matrix-Notation schreiben:

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}' \cdot \boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\beta})' \cdot (\mathbf{Y} - \mathbf{X} \cdot \boldsymbol{\beta}). \quad (2.30)$$

Hierin bedeutet das Superskript T die Transposition, welche auf Vektoren und Matrizen angewandt werden kann (vgl. die Matrixregeln in Abschnitt 2.11 dieses Kapitels):

- Transponiert man einen Spaltenvektor, erhält man den zugehörigen Zeilenvektor und umgekehrt,
- Bei einer Matrix werden die Zeilen und Spalten vertauscht: $(\mathbf{A}')_{ij} = (\mathbf{A})_{ji}$.

¹¹Verwechseln Sie bitte nicht den Vektor \mathbf{x} der exogenen Variablen im Modell (2.27) und die Matrix \mathbf{X} der Werte der exogenen Variablen in der Systemgleichung (2.28).

2 Stetige ökonomische Modelle

Minimierung der Fehlerquadratsumme

Zur Minimierung bezüglich β multipliziert man S zunächst aus:¹²

$$\begin{aligned} S &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}'\mathbf{Y} - (\mathbf{X}\beta)'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + (\mathbf{X}\beta)'(\mathbf{X}\beta) \\ &= \mathbf{Y}'\mathbf{Y} - \beta'(\mathbf{X}'\mathbf{Y}) - (\mathbf{Y}'\mathbf{X})\beta + \beta'(\mathbf{X}'\mathbf{X})\beta \end{aligned}$$

und mit der Skalarproduktregel $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$ die Fehlerquadratsumme (*sum of squared errors, SSE*)

$$S(\beta) = \mathbf{Y}'\mathbf{Y} - 2\beta'(\mathbf{X}'\mathbf{Y}) + \beta'(\mathbf{X}'\mathbf{X})\beta \quad (2.31)$$

Hierbei wurden die Distributivität aller Operationen (man kann wie bei Zahlen ausmultiplizieren), die Assoziativität (man kann beliebig klammern) und folgende Transpositionsregeln ausgenutzt (vgl. die Zusammenfassung zur Matrixrechnung in Abschnitt 2.11):

$$(\mathbf{A}\mathbf{b})' = \mathbf{b}'\mathbf{A}', \quad \mathbf{a}' \cdot \mathbf{b} = \mathbf{b}' \cdot \mathbf{a} \quad (2.32)$$

Also erhält man letztendlich

$$S(\beta) = \text{const.} - 2\beta' \cdot \mathbf{a} + \beta'\mathbf{A}\beta \quad \text{mit} \quad \mathbf{a} = \mathbf{X}'\mathbf{y}, \quad \mathbf{A} = \mathbf{X}'\mathbf{X}. \quad (2.33)$$

Auch das Ableiten nach allen Parametern kann man formal als (Spalten-)Vektoroperator schreiben:

$$\frac{\partial}{\partial \beta} \equiv \left(\frac{\partial}{\partial \beta_0}, \dots, \frac{\partial}{\partial \beta_J} \right)' \quad (2.34)$$

Mit den Ableitungsregeln für beliebige konstante Vektoren \mathbf{a} und Matrizen \mathbf{A} (vgl. wieder Abschnitt 2.11),

$$\frac{\partial}{\partial \beta} (\beta' \cdot \mathbf{a}) = \mathbf{a}, \quad (2.35)$$

$$\frac{\partial}{\partial \beta} (\beta'\mathbf{A}\beta) = (\mathbf{A} + \mathbf{A}')\beta, \quad (2.36)$$

erhält man, wenn man \mathbf{a} mit $\mathbf{X}'\mathbf{y}$ und \mathbf{A} mit der symmetrischen Matrix $\mathbf{X}'\mathbf{X}$ identifiziert,

$$\frac{\partial S}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta \stackrel{!}{=} 0$$

und schließlich nach Multiplikation mit der inversen Matrix $(\mathbf{X}'\mathbf{X})^{-1}$ von links den finalen OLS-Schätzer $\hat{\beta}$ der Modellparameter

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}. \quad (2.37)$$

Die Inversion der Matrix $\mathbf{A} = \mathbf{X}'\mathbf{X}$ entspricht der Lösung eines linearen Gleichungssystems mit $(J + 1)$ Unbekannten.

¹²Wo Eindeutigkeit besteht, werden die Punkte, welche das Skalarprodukt bzw. die Multiplikationen von Matrizen mit Vektoren und von zwei Matrizen beschreiben, weggelassen.

2 Stetige ökonomische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Man erhält

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 15 \\ 1 & 1 & 31 \\ 1 & 1 & 40 \\ 1 & 2 & 34 \\ 1 & 2 & 40 \\ 1 & 2 & 58 \\ 1 & 3 & 67 \\ 1 & 3 & 72 \\ 1 & 3 & 84 \\ 1 & 4 & 82 \\ 1 & 4 & 98 \\ 1 & 4 & 116 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 42 \\ 38 \\ 24 \\ 76 \\ 52 \\ 40 \\ 90 \\ 77 \\ 62 \\ 90 \\ 82 \\ 68 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 \\ n\bar{x}_1 & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ n\bar{x}_2 & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix} = \begin{pmatrix} 12 & 30 & 747 \\ 30 & 90 & 2223 \\ 747 & 2223 & 56319 \end{pmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.506 & -0.116 & -0.00215 \\ -0.116 & 0.470 & -0.0170 \\ -0.00215 & -0.0170 & 0.000718 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{pmatrix} 25.5 \\ 38.2 \\ -0.953 \end{pmatrix}$$

Aufgabe:

Zeigen Sie, dass für den Spezialfall eines Modells einer exogenen Variablen, $\hat{y} = a + bx$, die Matrixformulierung (2.37) auf die aus den einführenden Statistikvorlesungen (bzw. aus (2.26)) bekannten OLS-Schätzer $\hat{b} = s_{1y}/s_{11} = s_{xy}/s_{xx}$ und $\hat{a} = \bar{y} - \hat{b}\bar{x}$ der Einfachregression führt.

Hinweis: Die erste Spalte von \mathbf{X} besteht (wie immer) aus Einsen und die zweite Spalte aus den Messwerten x_i , $i = 1, \dots, n$.

2.4.6 Anschauliche Interpretation des linearen Modells

Das multivariate lineare Modell und die Bedeutung seiner Modellparameter wird nun anhand des einfachsten nichttrivialen Falls (zwei unabhängige Variablen) anhand der Nutzungshäufigkeit der ÖPNV-Systeme vergleichbarer Städte veranschaulicht (Abb. 2.10): Die wichtigsten Einflussparameter und damit geeignete unabhängige Variablen sind der Preis x_1 und die Reisezeit bzw., um unabhängig von der Größe des Systems zu werden, die mittlere Geschwindigkeit x_2 . Die abhängige Variable Y gibt das Maß der Nutzung,

2 Stetige ökonomische Modelle

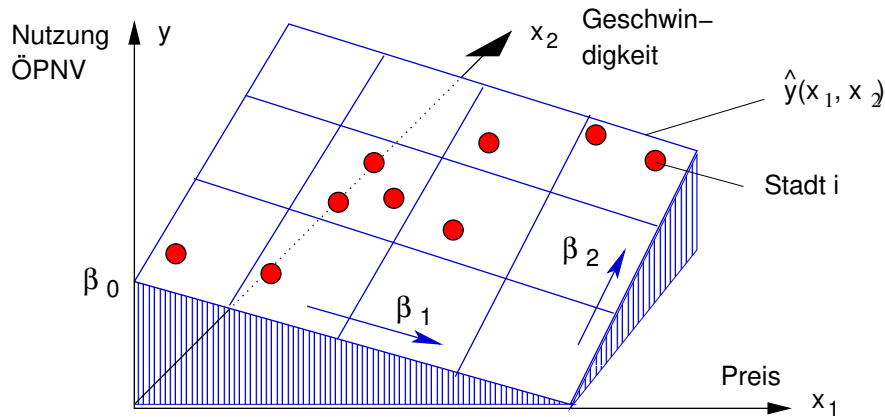


Abbildung 2.10: Bedeutung des Achsenabschnittes β_0 und der Anstiegsparameter β_1 und β_2 im bivariaten linearen Modell zur Beschreibung der Einflussfaktoren der ÖPNV-Nutzung. Die roten Punkte sind die Daten (Streudiagramm), während die blaue Ebene das bivariate Modell darstellt. Aus Gründen der Darstellung ist hier der Residualfehler=0 angenommen, so dass die roten Punkte auf der blauen Ebene liegen. Das lineare Modell beschreibt hier die Daten also exakt.

z.B. in Form von verkauften Tickets oder der spezifischen Verkehrsleistung in ÖPNV-Kilometern pro Einwohner und Jahr an. Die einzelnen Punkte in Abb. 2.10 geben die Beobachtungswerte in verschiedenen Städten i an.

Plausiblerweise fällt die Nutzung mit dem Preis, also wird ein negativer Steigungsparameter $\beta_1 = \frac{\partial \hat{y}}{\partial x_1}$ erwartet. Hingegen sollte die Nutzung mit der mittleren Geschwindigkeit unter *ceteris-paribus*-Bedingungen (gleiche Preise, gleiche sonstige Umstände) steigen, also $\beta_2 = \frac{\partial \hat{y}}{\partial x_2} > 0$. Schließlich gibt der Achsenabschnitt β_0 die Nutzungshäufigkeit bei kostenlosem jedoch stillstehendem ÖPNV an (offensichtlich ist hier der Anwendungsbeereich des linearen Modells überschritten).

Wichtig ist, sich anschaulich zu machen, dass die Vernachlässigung einer der beiden erklärenden Variablen zu falschen Schlussfolgerungen führen würde. Dies wird wesentlich durch die im Allgemeinen vorhandene *Korrelation* der Werte der unabhängigen Variablen bei den verschiedenen Stichprobenelementen bewirkt: Städte mit gut ausgebauten und schnellen ÖPNV haben tendenziell höhere Ticketpreise (pro Streckeneinheit) als Städte mit geringerwertigen ÖPNV-Netz. Die Korrelation $r_{12} = s_{12} / \sqrt{s_1^2 s_2^2}$ bzw. die Kovarianz s_{12} ist also in unserem Beispiel positiv: Die auf die x_1 - x_2 -Ebene projizierte Punktwolke ist also tendenziell längs einer ansteigenden Geraden angeordnet (Abb. 2.11).

Vernachlässigt man nun entweder x_1 oder x_2 , projiziert also die Datenpunkte auf die x_2 - y - bzw. auf die x_1 - y -Ebene, so hebt die Korrelation s_{12} zusammen mit der Projektion die Wirkung des noch verbleibenden wahren Anstiegsparameters β_2 bzw. β_1 weitgehend auf bzw. kann sogar zu gegenteiligen Effekten führen: Erhöht man die Geschwindigkeit, verbessert also das Angebot, reduziert sich die Nachfrage (Abb. 2.12). Dies führt zu

2 Stetige ökonometrische Modelle

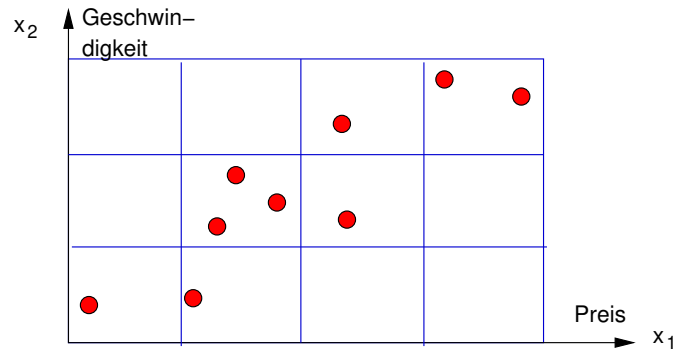


Abbildung 2.11: Projektion der Datenpunkte der Abbildung 2.10 auf die x_1 - x_2 -Ebene.

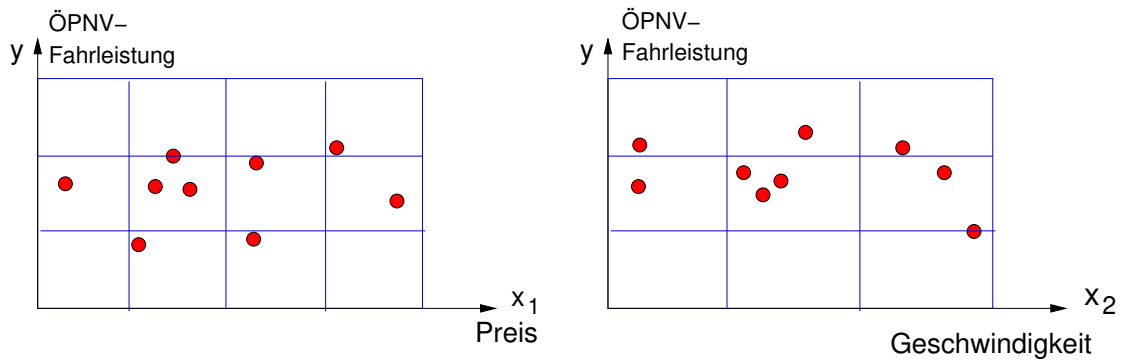


Abbildung 2.12: Streudiagramm der Datenpunkte, falls nur x_1 (links) bzw. x_2 (rechts) als erklärende Variable berücksichtigt wird. Die Daten entsprechen denen von Abbildung 2.10 und 2.11.

2 Stetige ökonometrische Modelle

folgender wichtigen Feststellung (vgl. auch Abschnitt 2.3):

Vernachlässigt man in einem linearen ökonometrischen Modell relevante erklärende Variable, ist das Modell **fehlspezifiziert** und die Modellkalibrierung durch lineare Regression oder der Maximum-Likelihood-Methode führt zu systematischen Fehlern. Dies resultiert in falschen Schlussfolgerungen und selbst der anschließende Belastungstest liefert oft irreführende bzw. falsche Aussagen. Das Modell ist dann, im besten Fall!, unbrauchbar. Eine unnötig mitgeführte irrelevante erklärende Variable hingegen verschlechtert zwar die Aussageschärfe, verfälscht das Ergebnis aber wenigstens nicht.

Dieser scheinbare Widerspruch entspricht einer der beiden Varianten des **Simpson'schen Paradoxes**:

- eine zusätzliche Variable im Modell kann den Effekt einer bereits vorhandenen Variablen umdrehen oder diese signifikant bzw. nichtsignifikant machen (dies ist im Hotelbeispiel der Fall, wenn man zum Preis auch noch die Sternenzahl hinzufügt, wodurch die vorher unplausible Preissensitivität wie erwartet negativ wird),
- auch zusätzliche Daten bei ein- und denselben Modell können die Effekte drehen. Typischerweise haben dann die neue Daten eine andere Ausprägung einer nicht berücksichtigten relevanten Variable, die mit den bereits vorhandenen korreliert (engl. *confounding variable*). Beispiel: Wertet man das Hotel-Modell mit nur dem Preis als exogene Variable zuerst allein mit den Daten der 3-Sterne Hotels aus, ist die Preissensitivität wie erwartet negativ. Fügt man dann zu den Daten die anderen Kategorien hinzu und wertet erneut aus, dreht die Preissensitivität auf positiv.

2.5 Statistische Eigenschaften der OLS-Schätzer quasilinearere Modelle

Da die Systemgleichungen des multivariaten lineare Modells (2.28).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Zufallsanteile $\boldsymbol{\epsilon}$ enthalten und die Schätzer $\hat{\boldsymbol{\beta}}$ der Parameter $\boldsymbol{\beta}$ von \mathbf{y} und damit von $\boldsymbol{\epsilon}$ abhängen, sind auch die Schätzer Zufallsgrößen. In diesem Abschnitt werden deren statistische Eigenschaften, also

- Erwartungswerte,
- Varianzen und Kovarianzen, und
- Verteilungsfunktionen

2 Stetige ökonometrische Modelle

ermittelt. Es wird die Vektor-Matrix-Notation angewandt, da diese hier viel effizienter ist. Für den Fall der Einfachregression werden die statistischen Eigenschaften in Abschnitt 2.10.1 zum Vergleich auch “zu Fuß” berechnet. Da quasilineare (parameterlineare) Modelle durch Transformation der exogenen Variablen in lineare Modelle umgeformt werden können, gilt das Folgende auch für letztere, sofern diese wohlspezifiziert sind.

Der Vektor der Parameterschätzer ist nach (2.37) gegeben durch

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

und \mathbf{y} nach (2.28) durch

$$\mathbf{y} = \mathbf{X} \cdot \beta + \epsilon.$$

Wichtig hierbei ist, dass in letzterer Gleichung die *wahren* Werte von β gemeint sind, während die Gleichung davor die (OLS-) *Schätzer* $\hat{\beta}$ definiert. Einsetzen der obigen Gleichungen (bzw. (2.28) in (2.37)) liefert

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X} \cdot \beta + \epsilon)$$

und mit den Matrixregeln von Abschnitt 2.11

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon. \quad (2.38)$$

2.5.1 Erwartungswert

Da $\hat{\beta}$ linear von ϵ abhängt und $E(\epsilon)$ aufgrund der Gauß-Markow-Annahmen =0 ist, gilt

$$E(\hat{\beta}) = \beta \quad (2.39)$$

Der Erwartungswert des Schätzers ist also gleich dem wahren Wert (zumindest, wenn die Gauß-Markow-Annahmen erfüllt sind, das Problem also korrekt spezifiziert ist). Damit erfüllt der OLS-Schätzer das wichtige Kriterium der *Unverzerrtheit*:

Eine notwendige Bedingung an alle sinnvoll verwendbaren Schätzer ist deren Unverzerrtheit. Ein **unverzerrter** bzw. **erwartungstreuer** Schätzer liefert “im Mittel” den wahren Wert. Der OLS-Schätzer linearer oder quasilinearer Modelle ist erwartungstreu.

2.5.2 Varianzen und Kovarianzen der OLS-Schätzer

Es soll nun die allgemeine **Varianz-Kovarianz-Matrix** $\mathbf{V} = \mathbf{V}_{\hat{\beta}}$ der Parameterschätzer mit den Elementen

$$V_{jk} = E \left[(\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k) \right] \quad (2.40)$$

bestimmt werden. Diese Matrix enthält

- in den Diagonalelementen die Varianzen der Schätzer, $V_{jj} = V(\hat{\beta}_j) = \sigma_{\beta_j}^2$,

2 Stetige ökonometrische Modelle

- in den Nichtdiagonalelementen die Kovarianzen $V_{jk} = \text{Cov}(\hat{\beta}_j, \hat{\beta}_k)$ zwischen verschiedenen Koeffizienten $k \neq j$.

Anhand der in Abschnitt 2.11.2 zusammengefassten Definitionen sieht man, dass man die Varianz-Kovarianz-Matrix kompakt als **Tensorprodukt** bzw **dyadisches Produkt** der Abweichungen der Schätzer vom wahren Wert schreiben kann:

$$\mathbf{V} = \mathbf{V}_{\hat{\beta}} = E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right). \quad (2.41)$$

Einsetzen von (2.38) unter Verwendung der Matrixregeln von Abschnitt 2.11 liefert

$$\begin{aligned} \mathbf{V} &= E \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \epsilon (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \epsilon' \right) \\ &= E \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \epsilon \epsilon' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \right) \end{aligned}$$

(beachten Sie, dass $[(\mathbf{X}'\mathbf{X})^{-1}]' = (\mathbf{X}'\mathbf{X})^{-1}$). Aus den Gauß-Markow-Annahmen folgt, dass \mathbf{X} deterministisch ist, so dass man den Erwartungswert $E(\dots)$ allein auf $\epsilon\epsilon'$ anwenden kann. Außerdem folgt wegen der i.i.d-Eigenschaft der Elemente von ϵ , dass sich der Erwartungswert auf eine Einheitsmatrix reduziert:¹³

$$E(\epsilon\epsilon') = \sigma_\epsilon^2 \mathbf{1} \quad (2.42)$$

Damit ergibt sich $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \sigma_\epsilon^2 \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$ und schließlich

$$\mathbf{V}_{\hat{\beta}} = \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1} \quad (2.43)$$

Daraus ergeben sich die im Weiteren wichtigen Größen Varianz, Kovarianz und Korrelationen¹⁴ der Parameterschätzer:

Varianzen:	$V(\hat{\beta}_j) = \sigma_{\hat{\beta}_j}^2 = V_{jj},$	(2.44)
Kovarianzen:	$\text{Cov}(\hat{\beta}_j, \hat{\beta}_k) = V_{jk},$	
Korrelationen:	$\text{Corr}(\hat{\beta}_j, \hat{\beta}_k) = \frac{V_{jk}}{\sqrt{V_{jj}V_{kk}}}$	

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die Varianz-Kovarianzmatrix der β -Schätzer ist proportional zu

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 0.506 & -0.116 & -0.00215 \\ -0.116 & 0.470 & -0.0170 \\ -0.00215 & -0.0170 & 0.000718 \end{pmatrix}$$

¹³Um Verwechslungen mit dem Erwartungswertoperator zu vermeiden, wird die Einheitsmatrix hier mit $\mathbf{1}$ bezeichnet.

¹⁴Beachten Sie, dass die Bezeichnung r_{jk} bereits für die deskriptiven Korrelationen zwischen den exogenen Variablen vergeben ist.

2 Stetige ökonomische Modelle

Die Varianz-Kovarianz-Matrix der Schätzer hängt also nur von der (wahren) Residualvarianz σ_ϵ^2 und von der Datenmatrix der exogenen Variablen ab, hingegen *nicht* von den gemessenen Werten der endogenen Variable. Da die Residualvarianz ihrerseits aber geschätzt werden muss (Näheres siehe bei der Diskussion von Formel (2.63)), kennt man die wahre Varianz-Kovarianz-Matrix nicht, sondern nur die **geschätzte** Varianz-Kovarianz-Matrix¹⁵

$$\hat{\mathbf{V}} = \hat{\sigma}_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1}, \quad \hat{\sigma}_\epsilon^2 = \frac{S_{\min}}{n-1-J}. \quad (2.45)$$

Hier ist $S_{\min} = S(\hat{\boldsymbol{\beta}})$ die bei der Parameterschätzung minimierte Fehlerquadratsumme (vgl. auch Formel (2.62)). Offensichtlich ist ein erwartungstreuer Schätzer umso besser, je geringer seine Varianz ist. Der “beste” Schätzer, also der mit der kleinsten Varianz, wird als **effizient** bezeichnet. Im Allgemeinen hängt die Form des effizienten Schätzers sowohl vom deterministischen Modell als auch von den statistischen Eigenschaften der Zufallsanteile ab. Für die OLS-Schätzer kann man (schwierig) folgendes **Gauß-Markow-Theorem** zeigen:

Für quasilineare Modelle mit i.i.d. normalverteilten additiven Zufallsvariablen ist der OLS-Schätzer **effektiv**, also unter allen *erwartungstreuen* Schätzern der mit der kleinsten Varianz. Dies wird auch als **BUE-Eigenschaft** (*best unbiased estimator*) bezeichnet. Ferner ist bei additiven i.i.d. Zufallsanteilen mit beliebiger Verteilung der OLS-Schätzer effektiv unter allen *linearen Schätzern*. Dies wird auch als **BLUE-Eigenschaft** (*best linear unbiased estimator*) bezeichnet.

Ein linearer Schätzer ist dabei, wie der OLS-Schätzer, in den Zufallsanteilen linear.

Die bei der Schätzung der Kovarianzmatrix zu invertierende Matrix $\mathbf{X}'\mathbf{X}/\hat{\sigma}_\epsilon^2$ bezeichnet man auch als **Fisher’sche Informationsmatrix**. Dass der OLS-Schätzer die BLUE-Eigenschaft hat, folgt aus dem allgemeinen mathematischen Satz, dass die minimale Streuung der Schätzer durch die **Kramér-Rhao-Schranke** $\hat{\sigma}_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1}$, also genau durch (2.45), nach unten beschränkt ist.

Für die Anwendung bedeutet die BLUE-Eigenschaft des OLS-Schätzer, dass man in der Regel bei linearen und quasilinearen Modellen ausschließlich diesen verwendet, obwohl es bei nicht-gaußverteilten i.i.d. Zufallsvariablen einen *nichtlinearen* Schätzer geben *könnte*, der einen kleinere Varianz besitzt. Ferner ist bei nichtlinearen Modellen, insbesondere bei den in Kapitel 4 behandelten Modellen der diskreten Wahltheorie, der OLS-Schätzer im Allgemeinen nicht effektiv (oder sogar nicht einmal unverzerrt) und es werden andere Schätzer, sogenannte *Maximum-Likelihood-Schätzer*, verwendet.

¹⁵Man beachte, dass sowohl der Parameter geschätzt wird (Dach über β_j), als auch dessen Varianz (Dach über V).

2 Stetige ökonometrische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die geschätzte Residualvarianz lautet

$$\hat{\sigma}_\epsilon^2 = \frac{S_{\min}}{9} = \frac{1}{9}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) = 55.4.$$

Die geschätzte Standardabweichung $\hat{\sigma}_\epsilon$ der unbestimmten Anteile entspricht einer Schwankung von etwa $\sqrt{55.4}\% = 7.5\%$ bei der Belegung. Ferner ergibt sich der Schätzer der Varianz-Kovarianz-Matrix der Parameter-Unbestimmtheiten zu

$$\hat{\mathbf{V}} = \hat{\sigma}_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 28.0 & -6.40 & -0.119 \\ -6.40 & 26.0 & -0.942 \\ -0.119 & -0.942 & 0.0397 \end{pmatrix}$$

und daraus die geschätzten Varianzen

$$\begin{aligned} \hat{V}(\hat{\beta}_0) &= \hat{V}_{00} = 28.0, \\ \hat{V}(\hat{\beta}_1) &= \hat{V}_{11} = 26.0, \\ \hat{V}(\hat{\beta}_2) &= \hat{V}_{22} = 0.0397 \end{aligned}$$

und die geschätzten Korrelationen

$$\begin{aligned} \text{Corr}(\beta_0, \beta_1) &= \frac{\hat{V}_{01}}{\sqrt{\hat{V}_{00}\hat{V}_{11}}} = -0.24, \\ \text{Corr}(\beta_1, \beta_2) &= \frac{\hat{V}_{12}}{\sqrt{\hat{V}_{11}\hat{V}_{22}}} = -0.93, \\ \text{Corr}(\beta_0, \beta_2) &= \frac{\hat{V}_{02}}{\sqrt{\hat{V}_{00}\hat{V}_{22}}} = -0.11. \end{aligned}$$

2.5.3 Verteilungs- und Dichtefunktionen der OLS-Schätzer

Allgemein gilt (vgl. Statistik-Formelsammlung), dass eine Linearkombination von normalverteilten Zufallsvariablen wieder normalverteilt ist.¹⁶ Da der OLS Schätzer in seiner Eigenschaft als linearer Schätzer eine Linearkombination der Zufallsanteile darstellt, stellt auch der Schätzer selbst eine gaußverteilte Zufallsvariable dar.

¹⁶Dies gilt auch, wenn die Zufallsvariablen korreliert sind.

2 Stetige ökonomische Modelle

Sind alle Gauß-Markow-Annahmen erfüllt (also insbesondere die Zufallsanteile i.i.d normalverteilt), so gehorcht die Verteilungsfunktion des Schätzvektors $\hat{\beta}$ einer multivariaten Normalverteilung mit Erwartungswert β und Kovarianzmatrix $\sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$.

Allgemein ist die Dichtefunktion einer multivariaten Gaußverteilung mit $J + 1$ exogenen Variablen, Erwartungswert $\mathbf{0}$ und Varianz-Kovarianz-Matrix Σ definiert durch (vgl. Abschnitt 2.11.4)

$$f(\mathbf{x}) = \sqrt{\frac{1}{(2\pi)^{J+1}\text{Det}\Sigma}} \exp\left[-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}\right]. \quad (2.46)$$

Vergleicht man die Varianz-Kovarianz-Matrix (2.43) mit der Dichte der allgemeinen multivariaten Gaußverteilung (vgl. Abschnitt 2.11.4), kann man die Dichtefunktion der Parameterschätzer explizit als Funktion der Abweichungen $\Delta\hat{\beta} = \hat{\beta} - \beta$ des OLS-Schätzvektors vom wahren Wert darstellen:

$$f_{\hat{\beta}}(\Delta\hat{\beta}) = \sqrt{\frac{\text{Det}(\mathbf{X}'\mathbf{X})}{(2\pi\sigma_\epsilon^2)^{J+1}}} \exp\left[-\frac{\Delta\hat{\beta}'\mathbf{X}'\mathbf{X}\Delta\hat{\beta}}{2\sigma_\epsilon^2}\right]. \quad (2.47)$$

Sind, wie in der Praxis, die wahren Werte von β nicht bekannt, aber (entgegen der Praxis) die wahre Varianz σ_ϵ^2 der Störvariablen, so stellt (2.47), nun in Abhängigkeit von β anstelle $\hat{\beta}$, die **Likelihoodfunktion** der wahren Werte von β dar.¹⁷

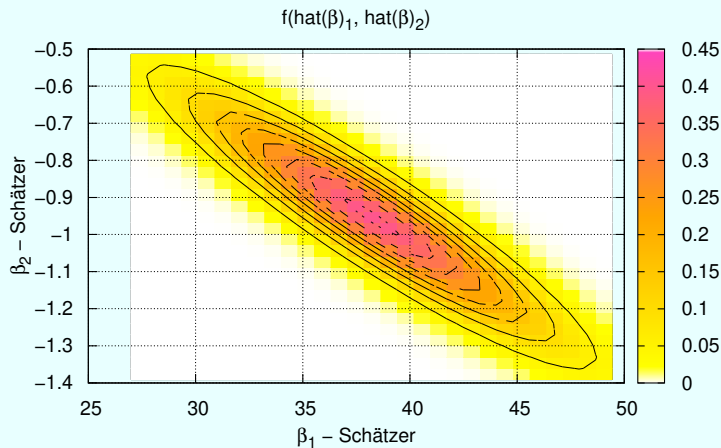
Ist man nur an der nichtnormierten Dichtefunktion interessiert, z.B. beim Darstellen der Verteilung als Kontourplot (Abb. 2.22) zum grafischen Test von verbundenen Nullhypothesen (Abschnitt 2.6.8), entfällt die Determinante und man muss keine komplizierte Matrixoperation (Matrixinverse oder Berechnung der Determinante) sondern nur eine Matrixmultiplikation durchführen.

¹⁷Man würde diese Verteilung mit Hilfe der im nächsten Kapitel behandelten *Maximum-Likelihood-Methode* bekommen.

2 Stetige ökonometrische Modelle

Beispiel zur Begleitung: Hotel-Auslastung

Im Folgenden ist der Schnitt der dreidimensionalen Dichtefunktion $f_{\hat{\beta}}(\hat{\beta})$ durch die Ebene $\beta_0 = \hat{\beta}_0$ gezeigt, wenn man annimmt, dass der Residualvarianz-Schätzer $\hat{\sigma}_\epsilon^2 = 55.4$ den wahren Wert trifft:



Varianz und Dichtefunktion in Abhängigkeit der Quadratsumme SSE

Wir schreiben zunächst die Fehlerquadratsumme (vgl. den Ausdruck vor (2.31)) als Funktion der Abweichung $\Delta\beta = \beta - \hat{\beta}$ eines beliebigen (i.A. nicht des wahren) Parametervektors vom Schätzer $\hat{\beta}$, welcher diese Summe minimiert. Während der Rechnung identifizieren wir $\mathbf{y} - \mathbf{X}\hat{\beta} = \epsilon$ mit den Residualabweichungen:

$$\begin{aligned}
 S(\beta) &= (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\
 &= (\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{X}\Delta\beta)'(\mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{X}\Delta\beta) \\
 &= (\epsilon - \mathbf{X}\Delta\beta)'(\epsilon - \mathbf{X}\Delta\beta) \\
 &= \epsilon'\epsilon - 2\epsilon'\mathbf{X}\Delta\beta + (\Delta\beta)'\mathbf{X}'\mathbf{X}\Delta\beta \\
 &= S_{\min} + (\Delta\beta)'\mathbf{X}'\mathbf{X}\Delta\beta.
 \end{aligned}$$

In der letzten Zeile wurde ausgenutzt, dass die Residualabweichung ϵ senkrecht auf den Abweichungsvektor $\Delta\mathbf{y} = \mathbf{X}\Delta\beta$ der endogenen Variablen steht, was allgemein bei der Minimierung (oder Maximierung) einer Quadratsumme gilt.¹⁸ Damit ist Differenz der Fehlerquadratsumme vom minimalen Wert eine quadratische Form der Abweichungen

¹⁸Man kann das auch mühsam “zu Fuß” zeigen, indem man in diesem Ausdruck $\epsilon = \mathbf{y} - \mathbf{X}\hat{\beta}$ und $\Delta\beta = \beta - \hat{\beta} = \beta - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ setzt und ausmultipliziert.

2 Stetige ökonomische Modelle

vom OLS-Schätzer:

$$S(\boldsymbol{\beta}) - S_{\min} = (\Delta\boldsymbol{\beta})' \mathbf{X}' \mathbf{X} \Delta\boldsymbol{\beta}. \quad (2.48)$$

Andererseits ergibt eine Taylorentwicklung von $S(\boldsymbol{\beta})$ um $\hat{\boldsymbol{\beta}}$ bis zur zweiten Ordnung (die exakt ist, da es bei parameterlinearen Modellen keine höheren Ordnungen gibt):

$$S(\boldsymbol{\beta}) = S_{\min} + \frac{1}{2}(\Delta\hat{\boldsymbol{\beta}})' \mathbf{H} \Delta\hat{\boldsymbol{\beta}} \quad (2.49)$$

mit der symmetrischen Matrix der zweiten Ableitungen (**Hessematrix**)

$$(\mathbf{H})_{ij} = \left[\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j} \right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}. \quad (2.50)$$

Vergleicht man (2.48) mit (2.49), ergibt sich $\mathbf{X}' \mathbf{X} = \frac{1}{2} \mathbf{H}$. Da aber andererseits nach (2.43) die Varianz-Kovarianzmatrix durch $\mathbf{V}_{\hat{\boldsymbol{\beta}}} = \sigma_\epsilon^2 (\mathbf{X}' \mathbf{X})^{-1}$ gegeben ist, gilt

$$\mathbf{V}_{\hat{\boldsymbol{\beta}}} = 2\sigma_\epsilon^2 \mathbf{H}^{-1}. \quad (2.51)$$

Die Varianz-Kovarianz-Matrix der Parameterschätzer ist also direkt proportional zur Inversen der Hessematrix der Fehlerquadratsumme am Kalibrierungspunkt. Formal dieselbe Beziehung ergibt sich bei der Maximum-Likelihood-Schätzung, wenn man $S(\boldsymbol{\beta})$ durch die negative Log-Likelihood $-2\sigma_\epsilon^2 \ln L(\boldsymbol{\beta})$ ersetzt (vgl. Abschnitt 4.6). Schließlich kann man unter Beschriftung von $\Delta\hat{\boldsymbol{\beta}} = -\Delta\boldsymbol{\beta}$ die quadratische Form (2.48) mit der Dichtefunktion (2.47) der allgemeinen multivariaten Gaußverteilung vergleichen:

$$f_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) \propto \exp\left(\frac{-(S(\boldsymbol{\beta}) - S_{\min})}{2\sigma_\epsilon^2}\right). \quad (2.52)$$

Dieses Ergebnis werden wir auch später mit Hilfe der Maximum-Likelihood-Schätzung herleiten. Da die Residualvarianz σ_ϵ^2 im Allgemeinen durch $\hat{\sigma}_\epsilon^2 = S_{\min}/(n - J - 1)$ (vgl. (2.45)) geschätzt werden muss, erhält man die *geschätzte* Verteilungsfunktion durch

$$\hat{f}_{\hat{\boldsymbol{\beta}}}(\boldsymbol{\beta}) \propto \exp\left[-\frac{(n - J - 1)}{2} \left(\frac{S(\boldsymbol{\beta})}{S_{\min}} - 1\right)\right]. \quad (2.53)$$

2.5.4 Niedrigdimensionale Spezialfälle

Zur Veranschaulichung werden nun die bisherigen Ergebnisse dieses Abschnitts für die Spezialfälle keine exogene Variable (Trivialmodell) bis zwei exogene Variable (Zweifachregression) betrachtet.

Trivialmodell

Das Trivialmodell ohne exogene Variable lautet einfach $y = \beta_0 + \epsilon$. Damit ist die Datenmatrix eine $n \times 1$ Matrix (also ein Spaltenvektor) aus Einsen:

$$\mathbf{X}' = (1 \ 1 \ \dots \ 1)$$

2 Stetige ökonometrische Modelle

und man erhält mit $\mathbf{X}'\mathbf{X} = n$ (eine Zahl bzw 1×1 Matrix) und $(\mathbf{X}'\mathbf{X})^{-1} = 1/n$ (das "Matrixinverse" einer Zahl ist das normale Inverse bzw. der Kehrwert)

$$\hat{\beta}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}, \quad (2.54)$$

mit den statistischen Eigenschaften

$$E(\hat{\beta}_0) = \beta_0, \quad V(\hat{\beta}_0) = \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma_\epsilon^2}{n}. \quad (2.55)$$

Der Schätzer ist in diesem Fall einfach der Mittelwertschätzer und die Varianz ist durch die schon aus der Statistikvorlesung bekannten Varianzformel gegeben. Insbesondere ist die Standardabweichung $\propto 1/\sqrt{n}$.

Eine exogene Variable

Hier gilt (vgl. die Inversionsformel einer 2×2 -Matrix am Ende dieses Kapitels)

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad (\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{ns_x^2} \begin{pmatrix} \frac{1}{n} \sum_i x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

und damit

$$\begin{aligned} V(\hat{\beta}_0) = V_{00} &= [(\mathbf{X}'\mathbf{X})^{-1}]_{00} = \frac{\sigma_\epsilon^2 \sum_i x_i^2}{n^2 s_x^2} \\ &= \frac{\sigma_\epsilon^2}{n^2 s_x^2} \left(\sum_i x_i^2 - n\bar{x}^2 + n\bar{x}^2 \right) = \frac{\sigma_\epsilon^2}{n^2 s_x^2} (ns_x^2 + n\bar{x}^2) \end{aligned}$$

Die anderen Elemente der Kovarianzmatrix ergeben sich direkt ohne langes Umrechnen, so dass man zusammenfassend folgende Komponenten erhält:

$$\begin{aligned} V(\hat{\beta}_0) &= V_{00} = \frac{\sigma_\epsilon^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2} \right), \\ V(\hat{\beta}_1) &= V_{11} = \frac{\sigma_\epsilon^2}{ns_x^2}, \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= V_{01} = \frac{-\sigma_\epsilon^2 \bar{x}}{ns_x^2}. \end{aligned} \quad (2.56)$$

Insbesondere ist (2.56) konsistent zur viel umständlicheren Rechnung "zu Fuß" in Abschnitt 2.10.1.

Zwei exogene Variable

Analog erhält man, nach etwas längerer Rechnerei, folgende Ausdrücke für die Varianzen der Anstiegsparameterschätzer und die Kovarianz V_{12} der Anstiegsparameter:

2 Stetige ökonomische Modelle

$$V_{jj} = \sigma_{\hat{\beta}_j}^2 = \frac{\sigma_\epsilon^2}{ns_{jj}(1-r_{12}^2)}, \quad j = 1, 2, \quad V_{12} = \frac{-\sigma_\epsilon^2 r_{12}}{ns_1 s_2 (1-r_{12}^2)} \quad (2.57)$$

mit der deskriptiven Korrelation zwischen den exogenen Variablen

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}. \quad (2.58)$$

Für die Varianzen von $\hat{\beta}_0$ und die weiteren Kovarianzen ist eine explizite Formel nicht mehr sinnvoll und man wertet direkt den allgemeinen Ausdruck (2.43) aus.

Diese Beziehungen zeigen, dass es im multivariaten Fall wichtig ist, dass die erklärenden Variablen x_j nicht direkt voneinander abhängig bzw. vollständig kollinear sind. Andernfalls ist ihre Korrelation $r_{12} = 1$ oder $r_{12} = -1$ und die Parameter und ihre Varianzen sind wegen der Nenner $(1-r_{12}^2)$ nicht definiert. Hängen die unabhängigen Variablen x_1 und x_2 nicht exakt aber doch weitgehend voneinander ab, was auch als **partielle Kollinearität** bezeichnet wird, wird die Varianz $V(\hat{\beta}_j)$ für $j = 1$ oder 2 wegen des Nenners $(1-r_{12}^2)$ sehr groß, was zu großen Konfidenzintervallen bzw. nicht ablehnbaren Teststats, also zu schwachen Aussagen führt. Ein Beispiel für eine solche Kollinearität ist häufig das Paar Reisezeit und Reiseweite. Dazu noch einige Bemerkungen:

- Eine nichtverschwindende Korrelation vom Betrag kleiner eins zwischen den erklärenden Variablen (also eine nichtexakte Kollinearität) ist ausdrücklich zugelassen, vgl. Abb. 2.11.
- Im mehrdimensionalen Fall “verdirbt” eine starke Kollinearität zwischen zwei exogenen Variablen nicht automatisch die Aussagekraft bezüglich anderer Variablen. Seien z.B. x_1 die Reisezeit, x_2 die Reiseweite und x_3 die Kosten und seien x_1 und x_2 stark kollinear, so sind β_1 und β_2 nur sehr ungenau zu bestimmen. Den Anstiegspareparameter β_3 kann man jedoch mit normaler Genauigkeit bestimmen. Dasselbe gilt für die *Summe* $\beta_1 + \beta_2$.

2.6 Gewinnung einer statistisch belastbaren Modellaussage

Die zwei Standardmethoden, um zu belastbaren statistischen Aussagen zu kommen, sind die bereits aus der [Grundvorlesung in Statistik II](#) bekannten Konfidenzintervalle und statistischen Tests:

- Bei der **Intervallschätzung** eines Parameters wird um einen geschätzten Parameterwert $\hat{\beta}_j$ ein **Konfidenzintervall** bestimmt, in dem sich der wahre Wert β_j mit einer gewissen Sicherheit $1 - \alpha$ befindet (Abschnitt 2.6.2). Hierbei wird die **Fehlerwahrscheinlichkeit** α meist mit $\alpha = 5\%$ oder 1% angenommen. Je kleiner α , desto sicherer, aber auch desto unschärfer werden die Aussagen. Das Konzept kann man auch auf das Schätzen von Parameterkombinationen sowie auf simultanes (verbundenes) Schätzen mehrerer Parameter bzw. Parameterkombinationen durch **Konfidenzregionen** erweitern (Abschnitt 2.6.7).

2 Stetige ökonomische Modelle

- Beim **statistischen Test** wird eine gewisse **Nullhypothese** über Parameter (z.B. $\beta_1 = \beta_{10}$ oder $\beta_2 < 0$) oder auch Parameterkombinationen (z.B. $\beta_1/\beta_2 < 1$) getroffen und getestet, ob man diese bei einer Fehlerwahrscheinlichkeit α ablehnen kann (Abschnitt 2.6.4). Nur im Falle einer Ablehnung bekommt man eine signifikante Aussage, nämlich “Die Nullhypothese trifft nicht zu”. Kann man den Test hingegen nicht ablehnen, erhält man keine Aussage, da daraus *nicht* die Gültigkeit der Nullhypothese folgt. Auch hier gilt: Je höher α , desto häufiger erreicht man die Ablehnung, desto unsicherer ist jedoch die damit verbundene Aussage. Eine direkte quantitative Formulierung über die Stärke der Aussage ist möglich, indem man die Argumentation umkehrt und den **p-Wert** als die minimale Fehlerwahrscheinlichkeit bestimmt, bei der H_0 gerade noch abgelehnt werden kann. Schließlich ist eine Erweiterung auf kompliziertere Nullhypothesen sowie auf simultane Tests mehrere Parameter(kombinationen) möglich (Abschnitt 2.6.8).

2.6.1 Schätzer für die Residualvarianz

Nachdem die Varianz-Kovarianzmatrix der Parameterschätzer und sogar deren Verteilung (vgl. den vorhergehenden Abschnitt) bekannt sind, könnte man im Prinzip Konfidenzintervalle und Hypothesentests bereits formulieren. Leider hängt aber die Varianz-Kovarianzmatrix von der Residualvarianz σ_ϵ^2 der unbestimmten Anteile ab, welche ebensowenig wie die wahren Werte von β bekannt ist. Die Residualvarianz σ_ϵ^2 wird aber benötigt: Konfidenzintervalle werden als Vielfaches der Standardabweichung der Schätzer formuliert; die Ablehnbarkeit von Hypothesentests hängt davon ab, ob die Nullhypothesen-Menge ein gewisses Mindest-Vielfaches der Standardabweichung des Schätzers außerhalb des geschätzten Wertes liegt.

Anders ausgedrückt: Nicht nur die wahren Parameterwerte, sondern auch die *Verteilung* der Parameterschätzer ist nicht bekannt. Man weiß zwar, dass die Schätzer (wenn die Gauß-Markow-Annahmen erfüllt sind) gaußverteilt sind, dass die maximale Dichte beim wahren Wert liegt (da unverzerrt), aber man kennt nicht die durch \mathbf{V} gegebenen wahren Streubreiten.

Um ein Konfidenzintervall oder einen Test anwenden zu können muss man also nicht nur die Parameter β selbst durch $\hat{\beta}$ schätzen, sondern *zusätzlich* die durch \mathbf{V} gegebenen Varianzen und Kovarianzen dieses Schätzers. Man benötigt also einen unverzerrten (und effizienten) Schätzer $\hat{\mathbf{V}}$ für \mathbf{V} . Die Bedingung der Unverzerrtheit lautet

$$E[\hat{\mathbf{V}}] = \mathbf{V}. \quad (2.59)$$

Da $\mathbf{V} = \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ als einzigen Zufallsanteil die Residualvarianz σ_ϵ^2 enthält, läuft dies auf das Auffinden eines unverzerrten Schätzers $\hat{\sigma}_\epsilon^2$ der Residualvarianz σ_ϵ^2 hinaus, denn wenn

$$E(\hat{\sigma}_\epsilon^2) = \sigma_\epsilon^2 \quad (2.60)$$

dann ist der naheliegende Schätzer

$$\hat{\mathbf{V}} = \hat{\sigma}_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1} \quad (2.61)$$

2 Stetige ökonomische Modelle

erwartungstreu:

$$E[\hat{\mathbf{V}}] = (\mathbf{X}'\mathbf{X})^{-1}E(\hat{\sigma}_\epsilon^2) = (\mathbf{X}'\mathbf{X})^{-1}\sigma_\epsilon^2 = \mathbf{V}.$$

Ein naheliegender Schätzer für σ_ϵ^2 wird mit dem Minimalwert $S_{\min} = S(\hat{\boldsymbol{\beta}})$ der bei der OLS-Schätzung bereits benötigten Fehlerquadratsumme S formuliert. Man könnte denken, dass das arithmetische Mittel der Summanden von S_{\min} ,

$$s_\epsilon^2 = \frac{S_{\min}}{n} = \frac{1}{n} \sum_i (\hat{y}_i - y_i)^2 = \frac{1}{n} (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) \quad (2.62)$$

bereits ein solcher erwartungstreuer Schätzer für σ_ϵ^2 ist. Dies ist aber nicht so, da ein Teil der in den y_i enthaltenen Information bereits bei der Parameterschätzung und damit bei der Definition von \hat{y}_i “verbraucht” wird, so dass die Summanden $(\hat{y}_i - y_i)^2$ nicht von y_i unabhängig sind. Um einen erwartungstreuen Schätzer zu bekommen, wird im Abschnitt 2.10.3 für Interessierte der Erwartungswert von S_{\min} berechnet. Das Ergebnis ist

$$E(S_{\min}) = (n - J - 1)\sigma_\epsilon^2$$

und damit letztendlich mit (2.22) folgenden erwartungstreuen (und auch effizienten) Schätzer der Residualvarianz:

$$\hat{\sigma}_\epsilon^2 = \frac{S_{\min}}{n - 1 - J} = \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{n - 1 - J} = \frac{(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})}{n - 1 - J}. \quad (2.63)$$

2.6.2 Konfidenzintervalle

Mit den Schätzern $\hat{\beta}_j$ für den Parameter β_j und deren Varianzschätzern $\hat{V}(\hat{\beta}_j) = \hat{V}_{jj}$ kann man nun den für die Konfidenzintervalle nötigen Abstand des Schätzers vom wahren Wert in Einheiten der *geschätzten* Standardabweichung angeben (vgl. Abb. 2.13):

$$T_{\beta_j} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} := \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim T(n - 1 - J). \quad (2.64)$$

Der Zähler ist dabei nach den bisherigen Überlegungen gaußverteilt, während die statistischen Eigenschaften der Varianz-Kovarianz-Matrix und damit von $\hat{\sigma}_{\hat{\beta}_j}$ nur von $\hat{\sigma}_\epsilon^2$ bestimmt werden. Aus der Herleitung des Varianzschätzers in Abschnitt 2.10.3 folgt, dass $\hat{\sigma}_\epsilon^2$ effektiv eine Summe aus $n - J - 1$ Quadraten von unabhängigen standardnormalverteilten Zufallsvariablen ist. Eine solche Summe ist, geeignet skaliert, **Chi-Quadratverteilt** mit $n - J - 1$ Freiheitsgraden.¹⁹ Des Weiteren ist der Quotient einer standardnormalverteilten Zufallsgröße und der Wurzel aus einer unabhängigen $\chi^2(m)$ verteilten Zufallsgröße (nach Skalierung der Chi-quadrat verteilten Größe auf den Erwartungswert 1) per Definition **Student-t-verteilt** mit m “Freiheitsgraden”. Die Skalierungen der

¹⁹Die χ^2 Verteilung mit m Freiheitsgraden ist die Verteilung der Summe von m unabhängigen standardnormalverteilten Zufallsvariablen.

2 Stetige ökonomische Modelle

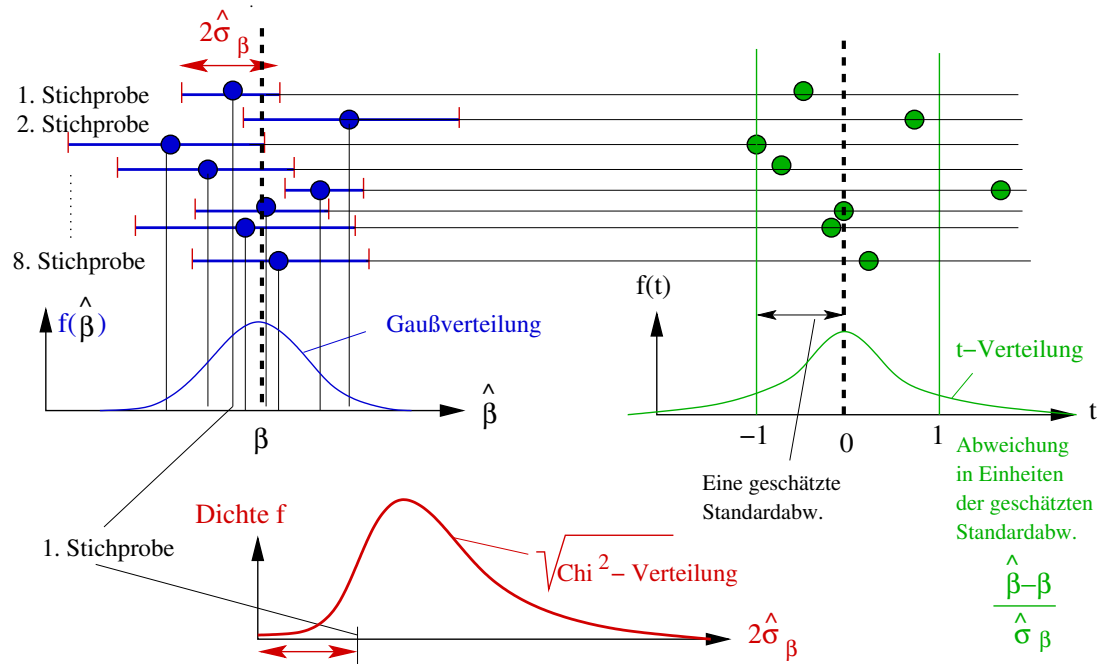


Abbildung 2.13: Veranschaulichung der Konfidenzintervalle und der Entstehung einer Student-t-Verteilung: Gezeigt sind 8 Stichproben eines Sachverhalts, welcher einen wahren Anstiegskoeffizienten β (z.B. β_1) aufweist. Die erwartungstreuen Schätzer $\hat{\beta}$ weisen einen i.i.d. gaußverteilten Schätzfehler mit wahrer Varianz σ_{β}^2 auf. Jede dieser 8 Stichproben liefert andere Schätzwerte $\hat{\beta}$ für β (Punkte links oben) und auch andere Schätzwerte $\hat{\sigma}_{\beta}$ für die Standardabweichung des Schätzers von β (Intervalle links oben). Da man die wahre Varianz nicht kennt, kann man die Abweichungen $\hat{\beta} - \beta$ (bzw. die Abweichungen vom Grenzwert β_0 einer Nullhypothese) nur in Einheiten der *geschätzten* Standardabweichung $\hat{\sigma}_{\beta}$ angeben, also die Größe $T = (\hat{\beta} - \beta_0) / \hat{\sigma}_{\beta}$. Dabei kann es durchaus vorkommen, dass ein Schätzwert, welcher näher am wahren Wert liegt als ein anderer (hier Stichprobe 5 gegenüber 2), in Einheiten der geschätzten Standardabweichungen weiter entfernt ist, die Größe T also betragsmäßig größer ist (Punkte rechts oben). Insbesondere kann bei wenig Datenpunkten die geschätzte Standardabweichung und damit der Nenner von T deutlich kleiner sein als die wahre Standardabweichung (Grafik links unten; bei der minimalen Zahl $n = J + 2$ von Datenpunkten ist der wahrscheinlichste Wert von $\hat{\sigma}_{\beta}$ sogar gleich null). Durch die resultierenden kleinen Nenner sind in der Größe T sehr große oder sehr kleinen Werte weitaus wahrscheinlicher als bei der Gaußverteilung, was man auch als **fat tails** bezeichnet (siehe die grüne Dichtefunktion auf der rechten Seite der Grafik). Insbesondere gehorcht T damit, im Gegensatz zu $\hat{\beta}$, keiner Gaußverteilung, sondern vielmehr einer **Student-t-Verteilung**.

2 Stetige ökonomische Modelle

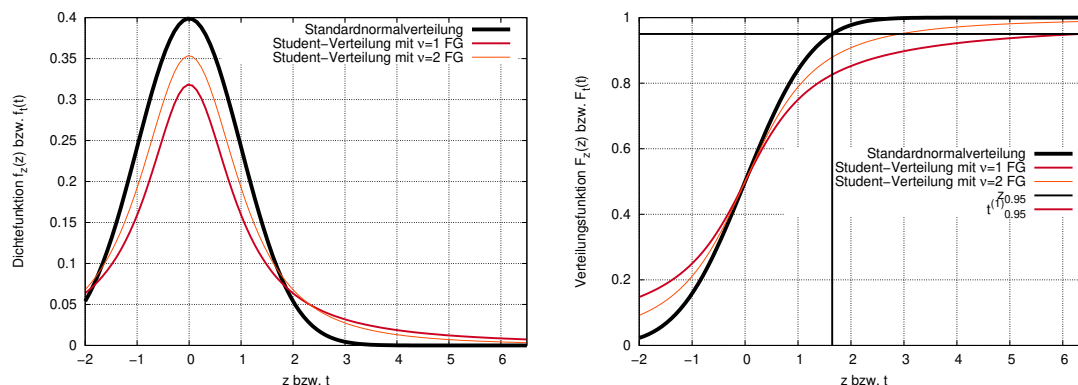


Abbildung 2.14: Dichte- und Verteilungsfunktion der Student-t-Verteilung mit ein und zwei Freiheitsgraden und der Limes (Standardnormalverteilung) für unendlich viele Freiheitsgrade. Bei den Verteilungsfunktionen sind als vertikale Linien auch die Quantile $t_{0,95}^{(1)}$ und $z_{0,95}$ der $T(1)$ - bzw. der Standardnormalverteilung eingezeichnet.

Zähler und Nenner (auf eine Standardnormalverteilung bzw. auf den Erwartungswert eins) heben sich hier gerade weg, so dass die Größe T direkt student-t - verteilt mit $n - 1 - J$ Freiheitsgraden ist. Die Student-t - Verteilung sieht wie eine Standardnormalverteilung aus, hat aber auf beiden Seiten mit wesentlich höherer Wahrscheinlichkeit Ausreißer, sogenannte *fat tails*. Dadurch ist das 95. Perzentil der Verteilung mit den ausgeprägtesten *fat tails*, der $T(1)$ -Verteilung, fast viermal so groß ($t_{0,95}^{(1)} \approx 6.3$) wie das entsprechende Quantil der Standardnormalverteilung ($z_{0,95} = 1.69$), vgl. (Abb. 2.14). In der Unterschrift zu Abb. 2.13 wird erklärt, woher diese Ausreißer kommen.

Gleichung (2.64) gibt die Verteilung des Abstandes T des Schätzers $\hat{\beta}_j$ von β_j in Einheiten der geschätzten Standardabweichung an. Dreht man nun die Argumentation um, so gilt, bei Gültigkeit der Gauß-Markow-Annahmen (insbesondere Homoskedastizität), dass $-T$ die bedingte Wahrscheinlichkeitsverteilung des Abstandes des wahren Wert β_j vom Schätzwert in Einheiten der geschätzten Standardabweichung bei gegebenen Messdaten angibt. Da T symmetrisch ist, ist die *Wahrscheinlichkeits-Dichte* des Abstandes des Schätzers vom wahren Wert gleich der **Likelihood-Dichte** (bedingte Wahrscheinlichkeitsdichte) des Abstandes des unbekanntes wahren Wertes bei gegebenem Schätzer.

Dies begründet die Konfidenzintervalle: Ein **Konfidenzintervall** für den wahren Parameter β_j zur Fehlerwahrscheinlichkeit α ist gleich dem um $\hat{\beta}_j$ symmetrischen Intervall, in welchem der wahre Wert β_j mit der Wahrscheinlichkeit $1 - \alpha$ liegt, also dass Intervall, was die Fläche $1 - \alpha$ der Likelihoodfunktion über sich vereinigt. Damit steht auf beiden Seiten der Likelihood-Dichte eine Wahrscheinlichkeit (also Fläche unter der Kurve) von $\alpha/2$ für "Ausreißer" zur Verfügung (Abb. 2.19). Das Konfidenzintervall wird also durch die entsprechenden **Quantile** der Student-t Verteilung mit $n - J - 1$ Freiheitsgraden bestimmt. Die **Quantilsfunktion** einer Verteilungsfunktion gibt das Argument

2 Stetige ökonometrische Modelle

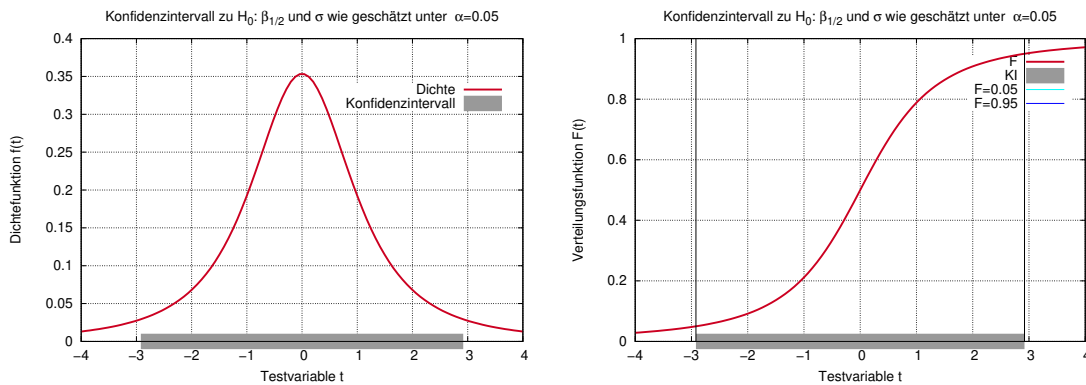


Abbildung 2.15: Konfidenzintervall bezüglich des auf die geschätzte Standardabweichung bezogenen Schätzwertes zu $\alpha = 5\%$ Fehlerwahrscheinlichkeit für $n - J - 1 = 2$ Freiheitsgrade. Das Intervall ist durch die Quantile $t_{0.05}^{(2)} = -t_{0.95}^{(2)}$ und $t_{0.95}^{(2)}$ begrenzt.

zu einer gegebenen kumulierten Wahrscheinlichkeit an, ist also die *Umkehrfunktion* der Verteilungsfunktion (Abb. 2.14). Damit ist das Konfidenzintervall gegeben durch (vgl. Abb. 2.15)

$$K_{\beta_j}^{(\alpha)} : \beta_j \in \left[\hat{\beta}_j - \Delta \hat{\beta}_j, \hat{\beta}_j + \Delta \hat{\beta}_j \right], \quad \Delta \hat{\beta}_j = t_{1-\alpha/2}^{(n-J-1)} \hat{\sigma}_{\hat{\beta}_j}. \quad (2.65)$$

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die Konfidenzintervalle der drei Parameter zu $\alpha = 5\%$ betragen (vgl. Abb. 2.16)

$$\begin{aligned} K_{\beta_0}^{(\alpha)} &= [13.5, 37.5], \\ K_{\beta_1}^{(\alpha)} &= [26.7, 49.7], \\ K_{\beta_2}^{(\alpha)} &= [-1.40, -0.50]. \end{aligned}$$

Insbesondere sind beide Anstiegsparameter signifikant von null verschieden und damit die durch die exogenen Variablen repräsentierten Faktoren “Sternezahl” und “Preis” bei einer Fehlerwahrscheinlichkeit $\alpha = 5\%$ beide relevant.

2.6.3 Extrapolation und Prognose

Nachdem man das Modell auf korrekte Spezifikation getestet hat und insbesondere alle nichtsignifikanten Einflussfaktoren (deren dazugehörige Anstiegsparameter nicht signi-

2 Stetige ökonometrische Modelle

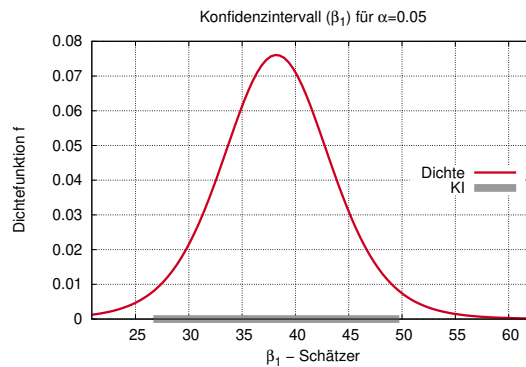


Abbildung 2.16: Likelihood-Dichte und Konfidenzintervall ($\alpha = 5\%$) der Sternenzahl-Sensitivität (Anstieg der Auslastung in Prozent pro Stern bei *ceteris-paribus*-Bedingungen) im Hotelbeispiel.

ifikant von null verschieden sind) eliminiert hat, kann man (mit Sorgfalt!) durch *Extrapolation* durch das geschätzte Modell auch Aussagen in Bereichen treffen, die nicht von den bisherigen Daten abgedeckt wurden. Dies wird am Beispiel der Einfachregression gezeigt.

Mit (2.56) erhält man für das geschätzte Modell der Einfachregression die Beziehung

$$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x = \beta_0 + \beta_1 x + \frac{1}{n} \sum_{i=1}^n \left(1 + \frac{(x_i - \bar{x})(x - \bar{x})}{s_{xx}} \right) \epsilon_i$$

und damit

$$E(\hat{Y}(x)) = \beta_0 + \beta_1 x. \quad (2.66)$$

Es sind also nicht nur die geschätzten Modellparameter sondern auch das gesamte geschätzte lineare Modell x erwartungstreu. Die Herleitung der Varianz als Funktion der Extrapolationsvariablen $x_1 = x$ geht analog wie die der Varianzen $V(\hat{\beta}_0)$ und $V(\hat{\beta}_1)$ in Abschnitt 2.10.1. Man muss nur immer beachten, dass sowohl die Messerte x_i als auch x als Argument der Funktion $\hat{Y}(x)$ *nichtstochastische* Größen darstellen. Das Ergebnis ist die sogenannte “Hyperbelformel” (vgl. Abb. 2.6.3):

$$V(\hat{Y}(x)) = \frac{\sigma_\epsilon^2}{n} \left(1 + \frac{(x - \bar{x})^2}{s_{xx}} \right). \quad (2.67)$$

2.6.4 Statistische Tests eines Parameters

Bei den hier betrachteten **Signifikanztests** wird eine bestimmte Annahme, die sogenannte **Nullhypothese** H_0 , daraufhin untersucht, ob sie mit den Messdaten in Übereinklang sein kann oder nicht.

2 Stetige ökonometrische Modelle

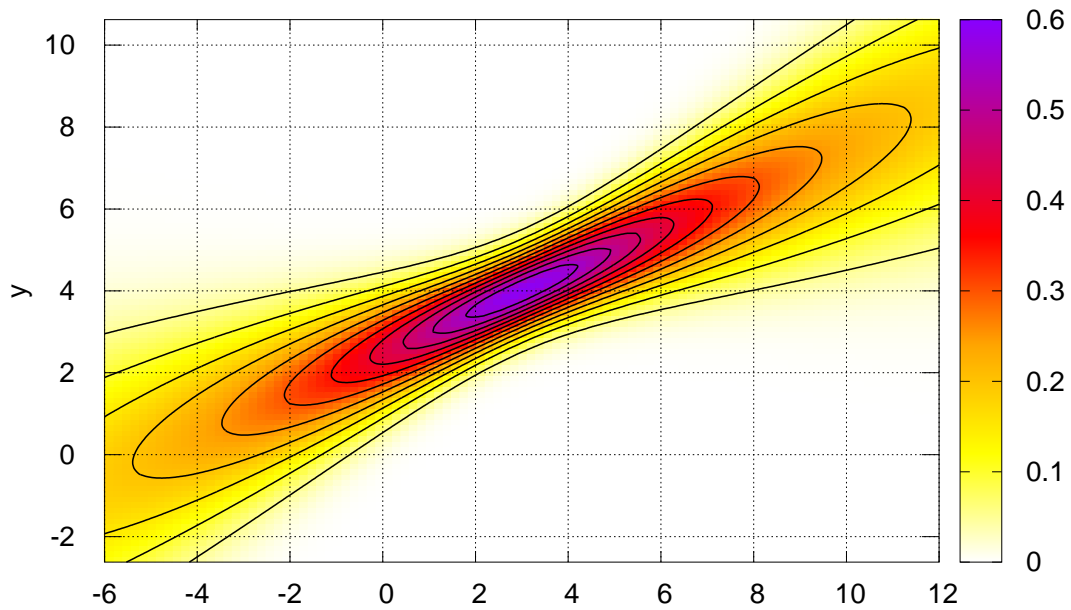


Abbildung 2.17: Wahrscheinlichkeitsdichte einer durch (2.67) beschriebenen Normalverteilung um den Erwartungswert $\beta_0 + \beta_1 x$ des geschätzten linearen Modells.

Die Nullhypothese H_0 beinhaltet eine Menge von Parameterkombinationen des verwendeten Modells. Liegt der dem wahren Sachverhalt (der Grundgesamtheit) entsprechende Parametervektor in dieser Menge, sagt man, H_0 sei erfüllt.

Hinweis: Sind die Gauß-Markow-Annahmen, insbesondere die Homogenitätsannahme, erfüllt, gibt es einen eindeutigen und unveränderlichen wahren Wert des Parametervektors in der Grundgesamtheit. H_0 ist also entweder objektiv erfüllt oder nicht.

Hinweis 2: Die Nullhypothese H_0 kann im Prinzip beliebig formuliert werden. Man sollte sie so wählen, dass deren *Ablehnung* einer relevanten Aussage entspricht, denn es gilt:

Eine signifikante Aussage gewinnt man bei Signifikanztests nur, wenn man die Nullhypothese H_0 verwerfen kann. Ist dies bei einer gewissen Fehlerwahrscheinlichkeit α nicht möglich, heißt das nicht, dass H_0 “angenommen” wird, sondern dass keine statistische Aussage möglich ist!

Diese Aussage wird dadurch klar, wenn man die in Abb. 2.18 dargestellten Konsequenzen einer Testentscheidung näher betrachtet. Die Zeilen stellen dabei den objektiven

2 Stetige ökonomische Modelle

	H_0 nicht abgelehnt	H_0 abgelehnt
H_0 trifft zu	✓	Fehler erster Art
H_0 trifft nicht zu	Fehler zweiter Art	✓

Abbildung 2.18: Definition der Fehler erster und zweiter Art bei Signifikanztests.

Sachverhalt dar: H_0 ist entweder wahr oder falsch (aufgrund der Homogenitätsannahme ist diese Unterscheidung, wie gesagt, immer eindeutig). Die Spalten stellen die Testentscheidung dar: H_0 annehmen oder nicht. Damit sind vier Fälle möglich:

- H_0 trifft zu und der Test wird nicht abgelehnt: Richtige Schlussfolgerung
- H_0 trifft nicht zu und der Test wird abgelehnt: Richtige Schlussfolgerung
- H_0 trifft zu und der Test wird abgelehnt. Dann hat man einen **Fehler erster Art** bzw. einen **α -Fehler** begangen. Dieser lässt sich mit der Fehlerwahrscheinlichkeit α (z.B. $\alpha = 5\%$) kontrollieren.
- H_0 trifft nicht zu und der Test wird nicht abgelehnt. Dann hat man einen **Fehler zweiter Art** bzw. einen **β -Fehler** begangen. Dieser lässt sich im Allgemeinen *nicht* kontrollieren. Dessen Wahrscheinlichkeit kann durchaus $1 - \alpha$ erreichen, also z.B. 95% , vgl. Abschnitt 2.6.6. Daher kommt die Regel, dass man Nullhypothesen nur ablehnen, nicht aber “annehmen” kann!

Die Vorgehensweise bei Signifikanztests verläuft immer nach demselben Schema:

1. Formulierung der Nullhypothese H_0 : Dies kann erfolgen als

- **Punkthypothese bzw. zweiseitiger Test:** $\beta_j = \beta_{j0}$, z.B. $\beta_j = 0$, und als
- **Intervallhypothese bzw. einseitiger Test:** $\beta_j < \beta_{j0}$, $\beta_j \leq \beta_{j0}$, $\beta_j > \beta_{j0}$, $\beta_j \geq \beta_{j0}$,
- **gekoppelte Nullhypothesen**, z.B. $\beta_1/\beta_2 < 1$.

Die Nullhypothese wird in Hinblick auf die gewünschte Aussage getroffen. Da man H_0 nicht ablehnen kann, behauptet man in H_0 immer das Gegenteil der zu untersuchenden Aussage, so dass man bei Ablehnung die Aussage selbst signifikant “bestätigt” hat. Will man beispielsweise die Klimaerwärmung nachweisen, stellt man sich auf den Standpunkt eines Klimaskeptikers: Die Nullhypothese lautet also, dass die globale Durchschnittstemperatur sinkt oder stagniert bzw. der entsprechende Anstiegparameter nichtpositiv ist. Dies versucht man dann zu widerlegen.

2 Stetige ökonometrische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Das betrachtete Modell lautete $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. Hierbei ist x_1 der Qualitätsindikator "Sternzahl" und x_2 der Übernachtungspreis in Euro pro Nacht. sinnvolle Nullhypothesen sind u.a.:

- $H_{01} : \beta_1 = 0$: "Die Qualität ist egal. Wenn überhaupt, dann entscheidet nur der Preis".
- $H_{02} : \beta_1 \leq 0$: "Ich liebe schlechte Hotels". Natürlich versucht man, diese Hypothese zu widerlegen, um damit signifikant nachzuweisen, dass die Qualität eben *nicht* egal ist.
- $H_{03} : \beta_2 = 0$: "Der Preis ist egal".
- $H_{04} : \beta_2 \leq -1$: "Jeder Euro Preiserhöhung vergrault so viele Gäste, dass die Auslastung um mindestens 1 % sinkt". Auch dies versucht man natürlich zu widerlegen ("der Schwund beträgt höchstens 1 % pro Euro Preisaufschlag"), was dem Hotelmanager als Entscheidungshilfe für die künftige Preisgestaltung dienen kann.
- $H_{05} : \beta_1 + 30\beta_2 \leq 0$: "Ein Stern mehr ist mir höchstens 30 Euro pro Nacht wert". Eine Widerlegung dieser Aussage beinhaltet, dass nach einer Renovierung, z.B. von einem Drei- in ein Viersternehotel, pro Nacht 30 Euro mehr verlangt werden kann, ohne einen Gästerückgang befürchten zu müssen.

2. **Formulierung der Testgröße ("Test-Statistik"):** Im realistischen Fall unbekannter Residualvarianz verwendet man bei Tests der Regressionsparameter direkt den Abstand T vom Rand des Parameterbereich, in welchem H_0 gilt (Rand der H_0 -Menge) und zwar *in Einheiten der geschätzten Standardabweichung*. Dabei rechnet man immer unter der Voraussetzung, dass H_0 tatsächlich *grenzwertig* zutrifft:

- Bei einer Punkthypothese ist dies eindeutig: Der wahren Wert β_j wird gleich dem Nullhypothesen-Wert β_{j0} gesetzt.
- Umfasst die Nullhypothesenmenge ein ganzes Intervall (Intervallhypothese), wird der *Worst Case* angenommen, also der Wert aus der Nullhypothesenmenge, bei dem ein Fehler erster Art die höchste Wahrscheinlichkeit hat. Dieser liegt immer auf der Grenze des H_0 -Intervalls bzw. am Rande des Parameterbereichs, in welchem H_0 zutrifft.
- Bei einer gekoppelten Nullhypothese wird eine Parameterkombination eingeführt, bezüglich derer der Test einer einfachen Punkt- oder Intervallhypothese

2 Stetige ökonometrische Modelle

entspricht, z.B. $\gamma = \beta_1 + 30\beta_2$ für die Nullhypothesen $\beta_1 < -30\beta_2$, $\beta_1 > -30\beta_2$ oder $\beta_1 = -30\beta_2$.

Daraus folgt:

Unabhängig von der Art der Nullhypothese ($=, \geq, >, \leq, <$) ist die Testvariable T ("Test-Statistik") des t -Tests immer der gerichtete Abstand des Parameterschätzers vom Rand der Nullhypothese in Einheiten der geschätzten Standardabweichung:

$$T_{\beta_j} = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\hat{V}(\hat{\beta}_j)}} := \frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}}. \quad (2.68)$$

Falls H_0 grenzwertig zutrifft, ist diese Testvariable Student-t verteilt mit $n - 1 - J$ Freiheitsgraden (Zahl der Datensätze abzüglich Zahl der zu schätzenden Parameter):

$$T_{\beta_j} \sim T(n - J - 1). \quad (2.69)$$

3. **Berechnen einer Instanz (Realisierung) t_{β_j} der Testvariablen** aus der Stichprobe.

Hotelbeispiel

Die Testgröße für die Nullhypothesen H_{01} und H_{02} und die aus den Daten berechnete Instanz lautet:

$$T_1 = T_2 = \frac{\hat{\beta}_1}{\sqrt{\hat{V}_{11}}} \sim T(9)$$

und die Realisierung

$$t_1 = t_2 = \frac{38.2}{\sqrt{26.0}} = 7.49.$$

Für H_{03} ergibt sich,

$$T_3 = \frac{\hat{\beta}_2}{\sqrt{\hat{V}_{22}}} \sim T(9), \quad t_3 = \frac{-0.952}{\sqrt{0.0397}} = -4.78$$

und für H_{04}

$$T_4 = \frac{\hat{\beta}_2 + 1}{\sqrt{\hat{V}_{22}}} \sim T(9), \quad t_4 = \frac{0.048}{\sqrt{0.0397}} = 0.237.$$

2 Stetige ökonometrische Modelle

Die fünfte kombinierte Nullhypothese (vgl. Abschnitt 2.6.7) kann man umformulieren zu

$$\gamma = \beta_1 + 30\beta_2 < 0$$

und damit

$$T_5 = \frac{\hat{\gamma}}{\sqrt{\hat{V}(\hat{\gamma})}} \sim T(9).$$

Mit

$$\begin{aligned}\hat{\gamma} &= \hat{\beta}_1 + 30\hat{\beta}_2 = 9.63, \\ \hat{V}(\hat{\gamma}) &= \hat{V}_{11} + 60\hat{V}_{12} + 900\hat{V}_{22} = 5.27\end{aligned}$$

ergibt dies die Realisierung

$$t_5 = \frac{9.63}{\sqrt{5.27}} = 4.20$$

4. **Auswertung des Tests** Die Testaussage “ H_0 kann abgelehnt werden” hängt wieder von der Nullhypothese ab:

- Bei *zweiseitigen Tests* teilt man, wie bei den Konfidenzintervallen, die Fehlerwahrscheinlichkeit auf beiden Seiten gleichmäßig auf (vgl. Abb. 2.19). Die Nullhypothese $\beta_j = \beta_{j0}$ kann bei einer Fehlerwahrscheinlichkeit von α abgelehnt werden, falls die Realisierung der Teststatistik der Bedingung

$$|t_{\beta_j}| > t_{1-\alpha/2}^{(n-J-1)} \quad (2.70)$$

genügt. Auf der rechten Seite steht dabei das (tabellierte) Quantil der T -Verteilung mit $n - 1 - J$ Freiheitsgraden. Mit der Nullhypothese $\beta_{j0} = 0$ findet dieser Test vor allem Anwendung, wenn man Einflussfaktoren (exogene Variable) auf Signifikanz oder speziell Zeitreihen auf das Vorhandensein eines Trends prüfen will.

- *Einseitige Tests*: Hier kann man die volle Fehlerwahrscheinlichkeit auf jeweils eine Seite der Test-Statistik T häufen:
 - $H_0 : \beta_j < \beta_{j0}$ und $H_0 : \beta_j \leq \beta_{j0}$ ist bei einer Fehlerwahrscheinlichkeit von α ablehnbar, falls

$$t_{\beta_j} > t_{1-\alpha}^{(n-J-1)}. \quad (2.71)$$

2 Stetige ökonometrische Modelle

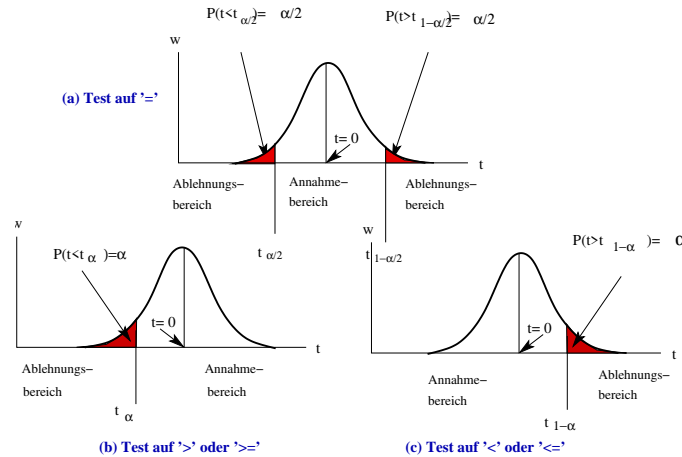


Abbildung 2.19: Annahme- und Ablehnungsbereiche bei (a) einseitigen Tests (Tests einer Punkt-Hypothese), (b,c) zweiseitigen Tests (Intervallhypothese).

- $H_0 : \beta_j > \beta_{j0}$ und $H_0 : \beta_j \geq \beta_{j0}$ ist bei einer Fehlerwahrscheinlichkeit von α ablehnbar, falls

$$t_{\beta_j} < t_{\alpha}^{(n-J-1)} = -t_{1-\alpha}^{(n-J-1)}. \quad (2.72)$$

Hotelbeispiel

Im folgenden wird versucht, die obigen Nullhypothesen bei $\alpha = 5\%$ Fehlerwahrscheinlichkeit zu widerlegen.

$$\begin{aligned} H_{01} : \beta_1 = 0, & \quad |t_1| = 7.49 > t_{0.975}^{(9)} = 2.26 & \Rightarrow \text{ja, also } H_{01} \text{ widerlegt!} \\ H_{02} : \beta_1 < 0, & \quad t_2 = 7.49 > t_{0.95}^{(9)} = 1.83 & \Rightarrow \text{ja, also } H_{02} \text{ widerlegt!} \\ H_{03} : \beta_2 = 0, & \quad |t_3| = +4.78 > t_{0.975}^{(9)} = 2.26 & \Rightarrow \text{ja, also } H_{03} \text{ widerlegt!} \\ H_{04} : \beta_2 < -1, & \quad t_4 = 0.24 > t_{0.95}^{(9)} = 1.83 & \Rightarrow \text{nein, } H_{04} \text{ nicht widerlegbar!} \\ H_{05} : \gamma < 0, & \quad t_5 = 4.20 > t_{0.95}^{(9)} = 1.83 & \Rightarrow \text{ja, also } H_{05} \text{ widerlegt!} \end{aligned}$$

Zweiseitige Tests bzw. Punkt-Nullhypothesen werden genau dann abgelehnt, wenn die Nullhypothese nicht im Konfidenzintervall (zu gleicher Fehlerwahrscheinlichkeit) liegt.

2.6.5 p-Werte

Eigentlich ist das Vorgehen, einen Test bei einer vorher festgelegten Fehlerwahrscheinlichkeit abzulehnen zu versuchen, umständlich und man verschenkt ggf. einiges an Schärfe

2 Stetige ökonomische Modelle

in den Aussagen. Beispielsweise könnte eine Nullhypothese, welche man bei $\alpha = 5\%$ ablehnen konnte, auch bei 0.1% ablehnbar sein und damit die damit verbundene Aussage nicht nur signifikant, sondern "hochsignifikant". Interessant ist deshalb eigentlich der *Grenzwert* der Fehlerwahrscheinlichkeit, bei dem man einen Test *gerade noch* ablehnen kann:

Der **p-Wert** (engl.: *p-value*) ist die Grenz-Fehlerwahrscheinlichkeit, bei der H_0 gerade noch abgelehnt werden kann:

$$p = \min(\alpha \mid H_0 \text{ ist ablehnbar bei Fehlerwahrscheinlichkeit } \alpha). \quad (2.73)$$

Eine allgemeinere Definition von p werden wir in Kap. 2.9 kennenlernen:

Der p -Wert zu einer Beobachtung b gibt die Wahrscheinlichkeit dafür an, die Beobachtung oder einen "extremere" Wert zu erhalten, falls die Nullhypothese *grenzwertig zutrifft*.

Nach (2.73) ergibt sich p aus den Test-Entscheidungsgleichungen, indem man α durch p und die Ungleichheitszeichen durch Gleichheitszeichen ersetzt, z.B. beim Test auf "kleiner" oder "kleiner gleich": $t_{\beta_j} > t_{1-\alpha}^{(n-J-1)} \rightarrow t_{\beta_j} = t_{1-p}^{(n-J-1)}$. Um dies nach dem p -Wert aufzulösen, nutzt man aus, dass die Quantile nichts anderes als die Umkehrfunktion der entsprechenden (kumulierten) Verteilungsfunktion darstellen, beispielsweise bei den Student-t und Standardnormalverteilungen:

$$F_T(t_\alpha) = \alpha, \quad \Phi(z_\alpha) = \alpha \quad (2.74)$$

mit $F_T(t)$ der Verteilungsfunktion der jeweiligen Student-t-Verteilung mit m Freiheitsgraden, t_α das jeweilige Quantil (der Index m für die Freiheitsgrade wurde der Übersichtlichkeit halber weggelassen), $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung und z_α die Quantilsfunktion der Standardnormalverteilung. Konkret führt dies auf folgende Ausdrücke für die p -Werte:

- Zweiseitige Tests bzw. Punkthypothesen bei gegebener Realisierung t der Test-Statistik $T \sim T(n-1-J)$:

$$\begin{aligned} p &= \min(\alpha \mid \text{so dass } |t| \geq t_{1-\alpha/2}) \\ \Leftrightarrow |t| &= t_{1-p/2} \\ F_T(|t|) &= 1 - \frac{p}{2}, \end{aligned}$$

also

$$p = 2(1 - F_T(|t|)). \quad (2.75)$$

- Bei einseitigen Tests auf $>$ bzw. \geq gilt analog

$$p = P(T < t) = F_T(t) \quad (2.76)$$

2 Stetige ökonometrische Modelle

- Bei einseitigen Tests auf $<$ bzw. \leq gilt

$$p = P(T > t) = 1 - F_T(t) \quad (2.77)$$

Die durch den p -Wert erhaltene Aussage teilt man häufig in drei Klassen ein:

- $p > 5\%$: Keine signifikante Aussage, da H_0 nicht abgelehnt werden kann
- $1\% \leq p < 5\%$: Durch die Ablehnung von H_0 erhält man eine *signifikante* Aussage. Ist dieses Signifikanzniveau bei einem Test eines Parameters auf den Wert 0 erreicht, so bekommt der Zahlenwert des Parameters per Konvention der meisten Statistikprogramme einen Stern, z.B. $\beta_1 = 2.1^*$
- $0.1\% \leq p < 1\%$: Durch die Ablehnung von H_0 erhält man eine *sehr signifikante* Aussage. Ist ein Parameter auf diesem Signifikanzniveau von null verschieden, so bekommt der Zahlenwert zwei Sterne, z.B. $\beta_2 = -3.14^{**}$
- $p < 0.1\%$: Durch die Ablehnung von H_0 erhält man eine *hochsignifikante* Aussage. Ist ein Parameter auf diesem Signifikanzniveau von null verschieden, so bekommt der Zahlenwert drei Sterne, z.B. $\beta_3 = 42^{***}$

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Die p -Werte der fünf Nullhypothesen auf S. 65 errechnen sich zu

$$\begin{aligned} H_{01} : \beta_1 = 0, \quad |t_1| = 7.49 \quad p &= 2(1 - F_T^{(9)}(|t_1|)) = 3.7 \times 10^{-5} \quad *** \\ H_{02} : \beta_1 < 0, \quad t_2 = 7.49 \quad p &= 1 - F_T^{(9)}(t_2) = 1.9 \times 10^{-5} \quad *** \\ H_{03} : \beta_2 = 0, \quad |t_3| = +4.78 \quad p &= 2(1 - F_T^{(9)}(|t_3|)) = 0.10\% \quad ** \\ H_{04} : \beta_2 < -1, \quad t_4 = 0.24 \quad p &= 1 - F_T^{(9)}(t_4) = 40\% \\ H_{05} : \gamma < 0, \quad t_5 = 4.20 \quad p &= 1 - F_T^{(9)}(t_5) = 0.12\% \quad ** \end{aligned}$$

Die kritischen Fehlerwahrscheinlichkeiten sind also meist sehr signifikant mit p -Werten häufig kleiner als 1%. Ein Test von beispielsweise H_{01} bei $\alpha = 5\%$ ergibt mit $t_{0.975}^{(9)} = 2.4$ eine Ablehnung, in Übereinstimmung mit den Konfidenzintervallen, aber man verschenkt die Aussage, dass dies nicht nur bei $\alpha = 5\%$ gilt, sondern hinunter bis zu $\alpha = p = 0.0037\%$.

2 Stetige ökonometrische Modelle

2.6.6 Gütefunktion und Fehler zweiter Art

Hier wird kurz erläutert, warum man den Fehler zweiter Art nicht bzw. nicht gleichzeitig mit dem Fehler erster Art kontrollieren kann. Dazu definieren wir zunächst die Gütefunktion:

Die **Gütefunktion** ist definiert als die Ablehnwahrscheinlichkeit des Tests in Abhängigkeit des *wahren* Parameterwertes β_j .

Wieder unterscheiden wir zwischen zweiseitigen und einseitigen Tests:

Zweiseitige Tests

Zunächst ist klar, dass bei stetigen Parameterwerten (die in der Regressionsrechnung immer Voraussetzung sind!) eine Punkthypothese $\beta_j = \beta_{j0}$ eigentlich *nie* erfüllt sein kann, da die Menge von H_0 das "Maß null" (bzw. das H_0 entsprechende Intervall die Länge null) aufweist. Damit ist die Wahrscheinlichkeit eines Fehlers zweiter Art direkt durch eins minus der Ablehnwahrscheinlichkeit des Tests gegeben.

Zum Berechnen der Gütefunktion für $\beta_j \neq \beta_{j0}$ muss man sich klarmachen, dass dann die Test-Variable $(\hat{\beta}_j - \beta_{j0})/\hat{\sigma}_{\hat{\beta}_j}$ *nicht* mehr Student-t-verteilt ist, sondern vielmehr die Variable

$$T = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - \beta_{j0} + \beta_{j0} - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}} - \Delta T.$$

Im letzten Ausdruck ist der erste Summand die beim Test verwendete Testvariable und

$$\Delta T = \frac{\beta_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}}$$

der Abstand des wahren Wertes vom Grenzwert β_{j0} in Einheiten der geschätzten Standardabweichung. Gemäß Punkt 4 (Entscheidung) des Test-Schemas ist die Gütefunktion $G^{\text{eq}}(\Delta T)$, definiert als die *Ablehnwahrscheinlichkeit* der Gleichheitsbedingung (Superskript "eq" für "equality") bei einer formalen "Fehlerwahrscheinlichkeit" α in Abhängigkeit der Abweichung ΔT ("Gütefunktion") gegeben durch

$$\begin{aligned} G^{\text{eq}}(\Delta T) &= P\left(\left|\frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}}\right| > t_{1-\alpha/2}\right) \\ &= P(|T + \Delta T| > t_{1-\alpha/2}) \\ &= P(T + \Delta T > t_{1-\alpha/2}) + P(T + \Delta T < -t_{1-\alpha/2}) \\ &= 1 - P(T + \Delta T \leq t_{1-\alpha/2}) + P(T + \Delta T < -t_{1-\alpha/2}) \end{aligned}$$

2 Stetige ökonomische Modelle

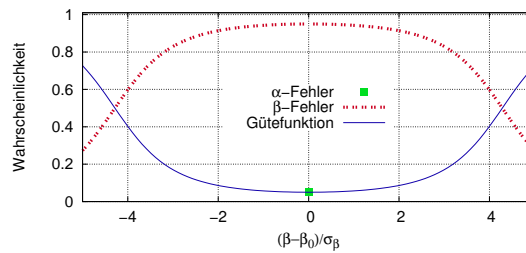


Abbildung 2.20: Gütefunktion und Fehler erster und zweiter Art bei zweiseitigen Tests (Punkthypothesen) in Abhängigkeit des skalierten Abstandes $\Delta t = (\beta_j - \beta_{0j}) / \hat{\sigma}_{\hat{\beta}_j}$ des wahren Parameterwertes von der Nullhypothese. Beispiel für $\alpha = 10\%$ und $n - J - 1 = 2$ Freiheitsgrade.

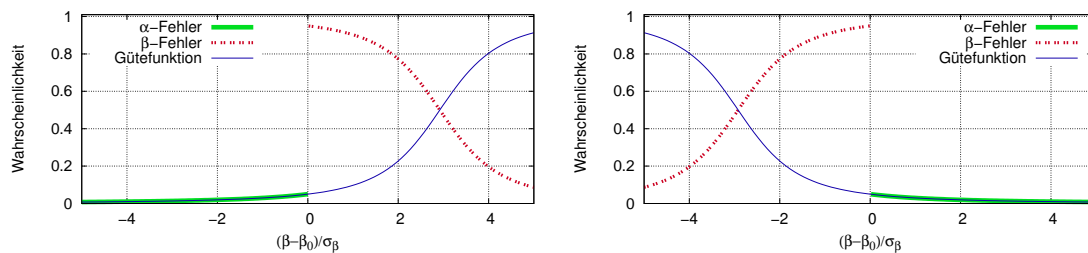


Abbildung 2.21: Gütefunktion und Fehler erster und zweiter Art bei einseitigen Tests auf $<, \leq$ (links) und $>, \geq$ (rechts) in Abhängigkeit des skalierten Abstandes $\Delta t = (\beta_j - \beta_{0j}) / \hat{\sigma}_{\hat{\beta}_j}$ des wahren Parameterwertes vom Grenzwert der Nullhypothese. Beispiele für $\alpha = 10\%$ und $n - J - 1 = 2$ Freiheitsgrade.

und mit der (kumulierten) Verteilungsfunktion $F_T(t)$ der Student-t-Verteilung unter Verwendung von $F_T(-t) = 1 - F_T(t)$

$$\begin{aligned} G^{\text{eq}}(\Delta T) &= 1 - F_T(t_{1-\alpha/2} - \Delta T) + F_T(-t_{1-\alpha/2} - \Delta T) \\ &= 2 - F_T(t_{1-\alpha/2} - \Delta T) - F_T(t_{1-\alpha/2} + \Delta T) \end{aligned}$$

Probe:

$$\begin{aligned} G^{\text{eq}}(0) &= 2 - F_T(t_{1-\alpha/2}) - F_T(t_{1-\alpha/2}) = 2 - 2 + 2\alpha/2 = \alpha \quad \checkmark \\ \lim_{\Delta T \rightarrow -\infty} G^{\text{eq}}(\Delta T) &= 2 - F_T(\infty) - F_T(-\infty) = 1 \quad \checkmark \\ \lim_{\Delta T \rightarrow +\infty} G^{\text{eq}}(\Delta T) &= 2 - F_T(-\infty) - F_T(\infty) = 1 \quad \checkmark \end{aligned}$$

Abbildung 2.20 zeigt das Ergebnis: Die Wahrscheinlichkeit eines Fehlers zweiter Art kann bis zu $1 - \alpha$ betragen, man weiß aber noch nicht mal den konkreten Wert dieser bedingten

2 Stetige ökonomische Modelle

Wahrscheinlichkeit, da die Bedingung (=der wahre Wert β_j) nicht bekannt ist. *in jedem Fall ist aber die Wahrscheinlichkeit eines Fehlers erster Art höchstens gleich α .*

Einseitige Tests

Die Berechnung der Gütefunktion für einseitige Tests geht analog:

Test auf “>” oder “ \geq ”:

$$\begin{aligned} G^{\geq}(\Delta T) &= P\left(\frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}} < t_{\alpha}\right) \\ &= P(T + \Delta T < t_{\alpha}) \\ &= P(T < -\Delta T + t_{\alpha}) \end{aligned}$$

und damit

$$G^{\geq}(\Delta T) = F_T(t_{\alpha} - \Delta T) \quad (2.78)$$

Probe:

$$\begin{aligned} G^{\geq}(0) &= F_T(t_{\alpha}) = \alpha \quad \checkmark \\ \lim_{\Delta T \rightarrow -\infty} G^{\geq}(\Delta T) &= F_T(\infty) = 1 \quad \checkmark \\ \lim_{\Delta T \rightarrow \infty} G^{\geq}(\Delta T) &= F_T(-\infty) = 0 \quad \checkmark. \end{aligned}$$

Test auf “<” oder “ \leq ”:

$$\begin{aligned} G^{\leq}(\Delta T) &= P\left(\frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}_{\hat{\beta}_j}} > t_{1-\alpha}\right) \\ &= P(T + \Delta T > t_{1-\alpha}) \\ &= P(T > -\Delta T + t_{1-\alpha}) \\ &= 1 - P(T < -\Delta T + t_{1-\alpha}) \end{aligned}$$

und damit

$$G^{\leq}(\Delta T) = F_T(\Delta T - t_{1-\alpha}) \quad (2.79)$$

Probe:

$$\begin{aligned} G^{\leq}(0) &= F_T(-t_{1-\alpha}) = 1 - F_T(t_{1-\alpha}) = 1 - (1 - \alpha) = \alpha \quad \checkmark, \\ \lim_{\Delta T \rightarrow -\infty} G^{\leq}(\Delta T) &= F_T(-\infty) = 0 \quad \checkmark, \\ \lim_{\Delta T \rightarrow \infty} G^{\leq}(\Delta T) &= F_T(\infty) = 1 \quad \checkmark \end{aligned}$$

Abbildung 2.21 zeigt, dass auch hier die Wahrscheinlichkeit eines Fehlers zweiter Art bis zu $1 - \alpha = 90\%$ betragen kann, während die Wahrscheinlichkeit eines Fehlers erster Art höchstens gleich α , u.U. aber auch wesentlich kleiner ist.

2 Stetige ökonomische Modelle

F-Test

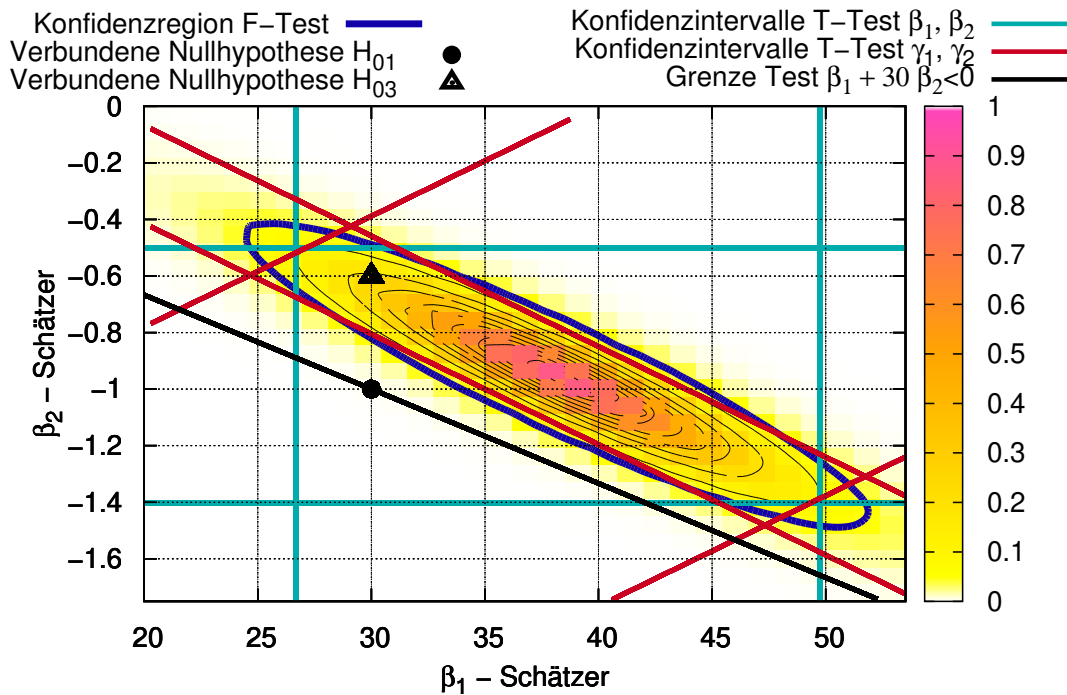


Abbildung 2.22: Zweidimensionale Konfidenzregionen des Hotelbeispiels für die beiden Anstiegsparameter β_1 und β_2 . Kontourplot: Auf den maximalen Wert 1 normierte Likelihood-Dichte der verbundenen Wahrscheinlichkeiten, dass β_1 und β_2 die jeweiligen Werte annehmen. Grünes Rechteck: Konfidenzregion als kartesisches Produkt zweier unverbundener Konfidenzintervalle (zu $\alpha = 5\%$). Punkt-Nullhypothesen, welche Punkten innerhalb des Rechtecks entsprechen, werden von keinem der beiden einfachen T -Tests abgelehnt. Blaue Ellipse: Verbundene Konfidenzregion, innerhalb der sich beide Anstiegsparameter simultan mit $1 - \alpha = 95\%$ Wahrscheinlichkeit aufhalten. Nullhypothesen außerhalb der blauen Ellipse aber innerhalb des Rechtecks werden vom verbundenen F -Test, nicht aber von den beiden T -Tests abgelehnt. Hellblaue parallele Linien: Konfidenzintervalle (=Abstand der jeweiligen Linien) der Linearkombinationen $\gamma_1 = \beta_2 + c\beta_1$ und $\gamma_2 = \beta_2 - c\beta_1$ mit $c = (V_{22}/V_{11})^{1/2}$. Auch das durch diese Linien aufgespannte Parallelogramm ist eine Konfidenzregion (mit einer kleineren Fläche als das Rechteck.) Die beiden schwarzen Symbole geben zwei verbundene Punkt-Nullhypothesen an, welche in Abschnitt 2.6.8 betrachtet werden: H_{01} (gefüllter Kreis) und H_{02} (Dreieck). Schließlich kennzeichnet die schwarze gerade die obere Grenze der kombinierten Intervall-Nullhypothese $\gamma = \beta_1 + 30\beta_2 < 0$.

2 Stetige ökonomische Modelle

2.6.7 Konfidenzregionen von Parameterkombinationen

Aufgrund der Korrelationen bei den Abweichungen der Parameterschätzer von den tatsächlichen Werten (vgl. Abb. 2.22) gibt es bei mehr als einem zu schätzenden Parameter gegenüber der naheliegenden seriellen Analyse jedes einzelnen Parameters nichttriviale weitere Schätz-Möglichkeiten, insbesondere Konfidenzintervalle von Parameterkombinationen und mehrdimensionale Konfidenzregionen.

Konfidenzintervalle von Parameterkombinationen

Das gewöhnliche, bisher betrachtete Konfidenzintervall eines Parameters gilt für beliebige Werte der anderen Parameter. Im Raum aller Parameter stellen die Begrenzungen der Intervalle daher Geraden (2 Parameter), Ebenen (3 Parameter) oder Hyperebenen (> 3 Parameter) dar. In Abb. 2.22 sind für das "Hotelbeispiel" die Begrenzungen der Konfidenzintervalle von β_1 und β_2 als grüne senkrechte bzw. waagerechte Geraden im von β_1 und β_2 aufgespannten Unterraum des Raums aller Parameter $(\beta_0, \beta_1, \beta_2)$ gezeichnet. Die Likelihood-Dichte (2.47) legt für dieses Beispiel nahe, dass *Linearkombinationen* der Ansteigparameter, z.B. durch die schrägen blauen Geraden charakterisiert, möglicherweise sehr viel kleinere Konfidenzintervalle aufweisen und damit schärfere Aussagen erlauben könnten. Eine allgemeine Linearkombination hat die Form

$$\gamma = \sum_j c_j \beta_j \quad (2.80)$$

mit dem Schätzer

$$\hat{\gamma} = \sum_j c_j \hat{\beta}_j. \quad (2.81)$$

Mit Hilfe elementarer Statistik-Regeln ist es ein Leichtes, Erwartungswert, Varianz und Verteilung von $\hat{\gamma}$ zu bestimmen:

- **Erwartungswert:**

$$E(\hat{\gamma}) = \sum_j c_j E(\hat{\beta}_j) = \sum_j c_j \beta_j.$$

- **Varianz:**

$$V(\hat{\gamma}) = \sum_j \sum_k c_j c_k \text{Cov}(\hat{\beta}_j, \hat{\beta}_k) = \sum_j \sum_k c_j c_k V_{jk} \quad (2.82)$$

- **Verteilung:** Eine Linearkombination von (korrelierten oder nicht korrelierten) gaußverteilten Zufallsvariablen ist wieder gaußverteilt. Also ist $\hat{\gamma}$ gaußverteilt.

Man bekommt also auch für $\hat{\gamma}$ die Student-t-verteilte Test-Statistik

$$T_\gamma = \frac{\hat{\gamma} - \gamma}{\sqrt{\hat{V}(\hat{\gamma})}} \sim T(n - J - 1), \quad (2.83)$$

2 Stetige ökonomische Modelle

wobei $\hat{V}(\hat{\gamma})$ durch (2.82) gegeben ist, wenn man dort V_{jk} durch den Varianz-Kovarianz-Schätzer \hat{V}_{jk} ersetzt. Damit ist das Konfidenzintervall von $\hat{\gamma}$ wie üblich gegeben durch

$$\gamma \in [\hat{\gamma} - \Delta\hat{\gamma}, \hat{\gamma} + \Delta\hat{\gamma}], \quad \Delta\hat{\gamma} = t_{(1-\alpha)/2}^{(n-J-1)} \sqrt{\hat{V}(\hat{\gamma})}. \quad (2.84)$$

In Abb. 2.22 sind in Türkis Konfidenzintervalle für die Linearkombinationen

$$\gamma_1 = \beta_2 + c\beta_1, \quad \gamma_2 = \beta_2 - c\beta_1, \quad c = \sqrt{\frac{V_{22}}{V_{11}}}$$

eingetragen. Der Koeffizient c wurde dabei so gewählt, dass die Geraden $\gamma_1 = \text{const.}$ parallel zur großen Halbachse der ellipsenförmigen Kurven konstanter Likelihood-Dichten (hellgelb=geringe, rot=hohe Dichte) liegen, so dass das Konfidenzintervall von γ_1 (die beiden fallenden türkisfarbenen Linien) besonders klein, dass von γ_2 (die steigenden Linien) hingegen besonders groß ist.

Verbundene Konfidenzregionen

Offensichtlich definieren sowohl das von den grünen "einfachen" Konfidenzintervallen von β_1 und β_2 eingeschlossene Rechteck (kartesisches Produkt), als auch das von den türkisfarbenen Intervallen der Parameterkombinationen γ_1 und γ_2 eingeschlossene Parallelogramm eine *Konfidenzregion*, in welcher beide Parameter β_1 und β_2 gleichzeitig mit einer Wahrscheinlichkeit von mindestens $(1 - \alpha)^2$ zu finden sind.²⁰ Offensichtlich sind darin aber durch die nichtberücksichtigte Korrelation jede Menge an Punkten enthalten, welche sehr unwahrscheinlichen Parameterkombinationen entsprechen (weiße Bereiche der Likelihood-Dichte in Abb. 2.22). Die Problemstellung ist daher, eine möglichst kleine Region zu finden, in welcher mit einer Wahrscheinlichkeit von $1 - \alpha$ alle betrachteten Parameter zu finden sind. Eine gute Näherung an eine solche Region wird mit Hilfe der Likelihood-Dichte (2.47) selbst konstruiert: Die *Konfidenzregion* ist die Menge aller Kombinationen von Parameterwerten, bei denen die Likelihood-Dichte (2.47) einen bestimmten Wert überschreitet. Die blau umrandete Region in Abb. 2.22 stellt eine solche Region dar. Im Zweidimensionalen (wie im Hotelbeispiel) ist die Konfidenzregion durch eine Ellipse eingeschlossen, im höherdimensionalen durch eine Ellipsoid-Hyperfläche. Wenn diese Region die "einfachen" Konfidenzintervalle gerade berühren würde, wäre die Wahrscheinlichkeit dafür, dass sie alle m betrachteten Parameterwerte enthält (im Beispiel ist $m = 2$), gleich $(1 - \alpha)^m$. Deshalb ist die wahre Region etwas größer und die blaue Ellipse geht etwas über die einfachen Konfidenzintervalle hinaus.²¹

Aufgrund von Korrelationen zwischen den Parameterschätzern ist die Konfidenzregion meist viel kleiner als das kartesische Produkt mehrerer einzelner Konfidenzintervalle (grünes Rechteck im Bild), die Aussagen also bei gleicher Fehlerwahrscheinlichkeit viel schärfer!

²⁰jeder Parameter liegt mit einer Wahrscheinlichkeit von $1 - \alpha$ außerhalb

²¹Streng genommen, müsste man wieder den Fall unbestimmter Varianz und damit eine mehrdimensionale Student-t Verteilung betrachten. Die dadurch hervorgerufenen Komplikationen (mehrdimensionale t -Verteilung) machen die Darstellung jedoch unübersichtlich, weshalb auf sie verzichtet wurde. Der weiter unten betrachtete F -Test (und auch die Grafik) berücksichtigen dies jedoch.

2 Stetige ökonomische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Hier kann man die Konsequenz sehr schön sehen: In Abb. 2.23 sind die Datenpunkte mit der Modellschätzung verglichen, und zwar für vier Sätze von Modellparametern $\hat{\beta}_1 \pm \Delta\hat{\beta}_1, \hat{\beta}_2 \pm \Delta\hat{\beta}_2$, welche der grünen Konfidenzregion entsprechen:

- Die Parametrisierungen $(\hat{\beta}_1 + \Delta\hat{\beta}_1, \hat{\beta}_2 - \Delta\hat{\beta}_2)$ und $(\hat{\beta}_1 - \Delta\hat{\beta}_1, \hat{\beta}_2 + \Delta\hat{\beta}_2)$ liegen innerhalb der verbundenen Konfidenzregion und beschreiben die Daten relativ gut (rechte Spalte der Abb. 2.23).
- Die Parametrisierungen $(\hat{\beta}_1 + \Delta\hat{\beta}_1, \hat{\beta}_2 + \Delta\hat{\beta}_2)$ und $(\hat{\beta}_1 - \Delta\hat{\beta}_1, \hat{\beta}_2 - \Delta\hat{\beta}_2)$ liegen weit außerhalb der verbundenen Konfidenzregion und führen zu extrem schlechten Übereinstimmungen mit den Daten (linke Spalte), so dass diese Nullhypothesen eigentlich hochsignifikant abzulehnen sind.

Offensichtlich liefert die verbundene Konfidenzregion, nicht aber das kartesische Produkt der einfachen Konfidenzintervalle, eine verlässliche Aussage darüber, ob eine gewisse Parameterkombination auszuschließen (bzw. die entsprechende Nullhypothese zu verwerfen) ist oder nicht.

Schlussfolgerung: Wegen der Korrelationen (im Beispiel ist $r_{12} > 0.9$) sind verbundene Konfidenzregionen und verbundene Tests viel zuverlässiger und trennschärfer im Vergleich zur sequentiellen Anwendung von eindimensionalen Konfidenzintervallen und Tests einzelner Parameter.

2.6.8 Simultane Tests verbundener Nullhypothesen: F -Test

Das Kapitel über Konfidenzregionen legt nahe, dass ein *simultaner* Test mehrerer Punkt-Nullhypothesen, also ein Test *verbundener* Nullhypothesen wie beispielsweise

$$H_0 : (\beta_1 = \beta_{10}) \cap (\beta_2 = \beta_{20})$$

andere Ergebnisse liefert, als wenn man die Nullhypothesen nacheinander einzeln testet. Meistens ist das Ergebnis trennschärfer (es gibt große Gebiete im Hypothesenraum, in welchem die verbundene Nullhypothese abgelehnt werden kann, aber keine einzige der Einzelhypothesen), es kann aber im Einzelfall auch anders herum sein wie im "Hotelbeispiel", Abb. 2.22, in der Nähe der Punkte $(\hat{\beta}_1 + \Delta\hat{\beta}_1, \hat{\beta}_2 - \Delta\hat{\beta}_2)$ bzw. $(\hat{\beta}_1 - \Delta\hat{\beta}_1, \hat{\beta}_2 + \Delta\hat{\beta}_2)$: Dort gibt es Punkte (β_{10}, β_{20}) , bei denen man die verbundene Hypothese nicht ablehnen kann, wohl aber eine oder sogar beide der Einzelhypothesen.

2 Stetige ökonomische Modelle

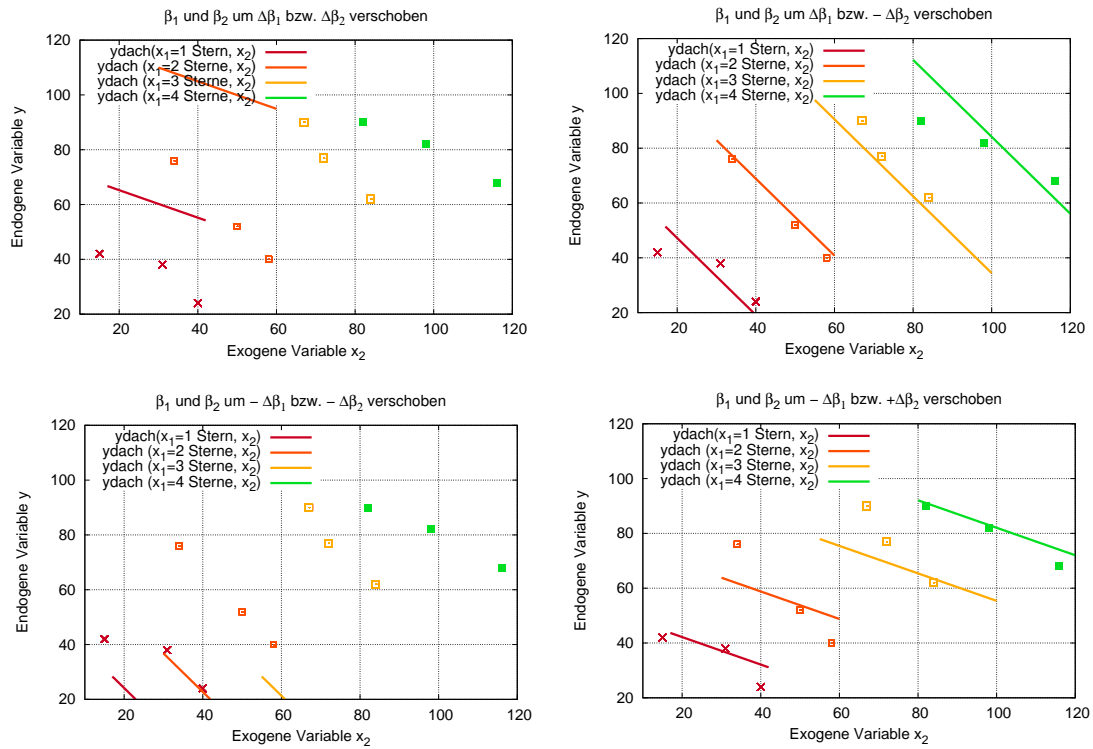


Abbildung 2.23: Übereinstimmung zwischen Modell und Daten für vier verschiedene Parametrisierungen. Die verschiedenfarbigen Datenpunkte entsprechen Hotels mit einem bis vier Sterne. Die dazugehörigen gleichfarbigen Geraden entsprechen der modellierten Auslastung als Funktion des Preises x_2 für gegebene Sternenzahl.

2 Stetige ökonomische Modelle

Der verbundene Test wird als F -Test durchgeführt.²² Das “Hotelbeispiel” mit der verbundenen Nullhypothese

$$H_{01} : (\beta_1 = \beta_{10} = 30) \cap (\beta_2 = \beta_{20} = -1)$$

erläutert das Vorgehen, welches wieder in den bekanntern vier Schritten abläuft.

1. **Nullhypothese:** H_{01} ist eine Punkt-Hypothese und entspricht dem ausgefüllten kleinem Kreis in Abb. 2.22.
2. **Bestimmen der Test-Statistik:** Die Testvariable

$$F = \frac{\left(S_{\min}^{(0)} - S_{\min} \right) / (J - J^{(0)})}{S_{\min} / (n - J - 1)} \sim F(J - J^{(0)}, n - J - 1) \quad (2.85)$$

bzw. hier

$$F = \frac{\left(S_{\min}^{(0)} - S_{\min} \right) / 2}{S_{\min} / 9} \sim F(2, 9) \quad (2.86)$$

ist (was hier nicht gezeigt wird) bei Zutreffen der Nullhypothese Fisher-F-verteilt mit $f_Z = J - J^{(0)} = 2$ Zähler-Freiheitsgraden und $f_N = n - J - 1 = n - 3$ Nenner-Freiheitsgraden.²³

- Im Zähler dieser Gleichung steht dabei die Reduktion der Fehlerquadratsumme des vollen Modells gegenüber dem Nullmodell (*restringiertes Modell*) pro zusätzlichen Parameter. Hier hat das volle Modell ($J = 2$) drei und das restringierte Modell ($J^{(0)} = 0$) einen Parameter, das volle Modell hat also zwei zusätzliche Parameter.
- Im Nenner steht die Fehlerquadratsumme des vollen Modells pro Freiheitsgrad. Die Zahl der Nenner-Freiheitsgrade $f_N = n - 1 - J$ ist gleich der Zahl n der Datensätze abzüglich der Zahl $(J + 1)$ der zu schätzenden Parameter des Vollen Modells.

Offensichtlich ist das Nullmodell wesentlich “schlechter” (und damit die verbundene Nullhypothese zu verwerfen), wenn man beim Übergang vom restringierten zum vollen Modell pro zusätzlich aufgewendeten Freiheitsgrad (hier zwei) wesentlich mehr Fehlerquadratsumme spart als im vollen Modell pro Freiheitsgrad noch übrig bleiben: Je höher der Wert f von F aus der Stichprobe, desto wahrscheinlicher trifft H_0 nicht zu.

²²Man kann auch einfach testen, ob die H_0 entsprechende Parameterkombination innerhalb der verbundenen Konfidenzregion liegt, denn die Äquivalenz zwischen Punkttests und Konfidenzregionen gilt auch im Mehrdimensionalen. Allerdings ist die genaue Festlegung der Größe der Konfidenzregion, so dass sie $1 - \alpha$ der Wahrscheinlichkeitsmasse (und nicht etwa $(1 - \alpha)^2$) enthält, schwierig und der Übergang von einer Gauß zu einer (mehrdimensionalen) t -Verteilung bei unbekannter Varianz ebenfalls schwierig. All dies ist im F -Test berücksichtigt.

²³die zweiparametrische Fisher'sche Verteilung $F(f_Z, f_N)$ mit f_Z bzw. f_N Freiheitsgraden im Zähler bzw. Nenner wird realisiert durch den Quotient zweier unabhängiger, auf den Erwartungswert 1 normierter χ^2 verteilter Zufallsgrößen mit der entsprechenden Zahl an Freiheitsgraden.

2 Stetige ökonomische Modelle

3. Bestimmen der Realisierung aus der Stichprobe

Die minimale Fehlerquadratsumme

$$S_{\min} = n\hat{\sigma}_\epsilon^2 = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}(\mathbf{x}_i))^2$$

des vollständigen geschätzten Modells hat man typischerweise schon bei der OLS-Parameterschätzung berechnet, da die Schätzung ja gerade durch die Minimierung der Fehlerquadratsumme mit dem Ergebnis S_{\min} definiert ist. Im Hotelbeispiel gilt $n = 12$ und $S_{\min} = 498$.

Die Fehlerquadratsumme $S_{\min}^{(0)}$ des *neu geschätzten, auf die Nullhypothese restringierten Modells*

$$\hat{y}^{(0)}(x_1, x_2; \beta_0^{(0)}) = \beta_0^{(0)} + \beta_{10}x_1 + \beta_{20}x_2 \quad (2.87)$$

muss neu berechnet werden. Zunächst muss der verbleibende Parameter $\beta_0^{(0)}$ des restringierten Modells geschätzt werden:²⁴ Dazu nutzt man aus, dass das restringierte Modell als ein gewöhnliches “Trivialmodell”

$$\hat{z} = \beta_0^{(0)}$$

mit der neuen endogenen Variablen

$$z = y - \beta_{10}x_1 - \beta_{20}x_2$$

geschrieben werden, wobei die “Messwerte” dieser neuen endogenen Variablen durch

$$z_i = y_i - \beta_{10}x_{i1} - \beta_{20}x_{i2}$$

gegeben sind. Dadurch erhält man direkt den Schätzer für den einzigen Parameter des Nullmodells,

$$\hat{\beta}_0^{(0)} = \bar{z} = \bar{y} - \beta_{10}\bar{x}_1 - \beta_{20}\bar{x}_2 = 49$$

und die dazugehörige Fehlerquadratsumme

$$S_{\min}^{(0)} = \sum_{i=1}^n (z_i - \beta_0^{(0)})^2 = 1808.$$

Setzt man all das in (2.85) ein, ergibt sich die Realisierung der Fisher-(2, $n - 3$) verteilten Test-Statistik F :

$$f = \frac{(1808 - 498)/(2 - 0)}{498/(12 - 2 - 1)} = 11.8.$$

²⁴Die OLS-Schätzung des Achsenabschnittsparameters $\beta_0^{(0)}$ des restringierten Modells ist im Allgemeinen *nicht* gleich der Schätzung $\hat{\beta}_0$ für das volle Modell!

2 Stetige ökonomische Modelle

4. **Auswertung und Entscheidung:** Pro “aufgewendeten” Freiheitsgrad konnte man durch das “Upgrade” vom Nullmodell auf das volle Modell also das 11.8 fache an Fehlerquadratsumme “entsorgen”, als im vollen Modell pro Freiheitsgrad noch übrig bleibt: Mit zwei Freiheitsgraden hat man $S_{\min}^{(0)} - S_{\min} = 1310$ an Fehlerquadratsumme, also 655 pro Freiheitsgrad, “entsorgt”. Es verbleibt eine Fehlerquadratsumme von $S_{\min} = 498$, aufgeteilt auf $n - 3 = 9$ Freiheitsgrade, also 55 pro Freiheitsgrad.

Damit ist klar, dass die Verbesserung durch das volle Modell signifikant ist, also H_{01} abgelehnt werden kann, wenn die Realisierung f von F hinreichend groß ist (größer als das entsprechende Quantil der Fisher-F-Verteilung), genau wie beim Test auf “Kleiner-Gleich”:

$$H_0 \text{ bei Fehlerwahrsch. } \alpha \text{ ablehnbar} \Leftrightarrow f > f_{1-\alpha}^{(J-J^{(0)}, n-J-1)} = f_{1-\alpha}^{(2,9)} \quad (2.88)$$

Der dazugehörige p -Wert ergibt sich analog aus der Verteilungsfunktion $F_F^{(2, n-3)}(x)$ der Fisher’schen F-Statistik zu

$$p = 1 - F_F^{(J-J^{(0)}, n-J-1)}(f) = 1 - F_F^{(2,9)}(11.8) = 0.3\% \quad (2.89)$$

Bei Zutreffen der Nullhypothese würde sich also nur in 0.3% der Fälle ein Wert der F -Statistik von $\geq f$ und damit ein Fehler erster Art ergeben (Abb. 2.24).

Das Ablehnen der Nullhypothese in diesem Beispiel ist auch konsistent mit der Konfidenzregion und der Likelihood-Dichte der Abbildung 2.22: Dort ist die Nullhypothese als schwarzer gefüllter Kreis eingezeichnet. Ebenfalls konsistent mit den anderen Analysemethoden ist der F -Test der verbundenen Nullhypothese

$$H_{02} : (\beta_{10} = 34) \cap (\beta_{20} = -1)$$

(schwarzes Dreieck in Abb. 2.22), welcher bei $\alpha = 0.05$ noch nicht abgelehnt werden kann (p -Wert 6%, vgl. Abb. 2.24).

2.6.9 Test auf Relevanz von Einflussfaktoren

Eine sehr schöne Anwendung des F -Tests verbundener Nullhypothesen ist der Vergleich zweier verschachtelter (engl. *genesteter*) Modelle:

- Das einfachere Modell M1 hat J_1 Einflussfaktoren (exogenen Variablen),
- Das komplexere Modell M2 hat $J_2 > J_1$ exogene Variablen, darunter auch alle des einfacheren Modells.

Dieser F -Test hat eine ähnliche Anwendung wie der später bei den diskreten Modellen vorgestellte *Likelihood-Ration-Test*. Überprüfen, ob man überflüssige (irrelevante) Einflussfaktoren bzw. Modellparameter eliminieren bzw. neue relevante Einflussfaktoren aufnehmen muss, um ein korrekt spezifiziertes Modell zu erhalten.

2 Stetige ökonomische Modelle

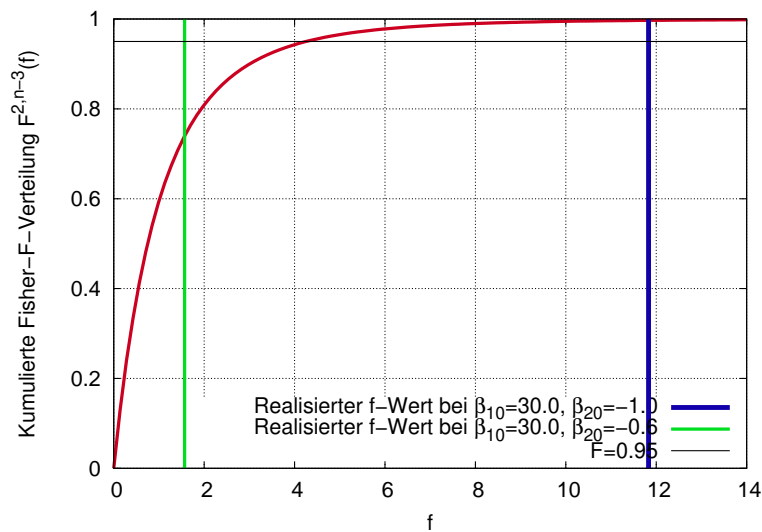


Abbildung 2.24: F-Test zweier verbundenen Nullhypothesen im Hotelbeispiel. Die Nullhypothese $H_{01} : \beta_{10} = 30$ und $\beta_{20} = -1$ wird signifikant abgelehnt ($f = 11.8, p = 0.3\%$), während $H_{02} : \beta_{10} = 34$ und $\beta_{20} = -1$ bei 5% Fehlerwahrscheinlichkeit nicht abgelehnt werden kann ($f = 3.9, p = 6.1\%$).

Bezeichnet man mit x_1, \dots, x_{J_1} die exogenen Variablen von M1 und mit $x_{J_1+1}, \dots, x_{J_2}$ die zusätzlichen Variablen von M2, so ist offensichtlich Modell M2 nicht signifikant besser als M1, wenn die Nullhypothese

$$H_0 : (\beta_{J_1+1} = 0) \cap \dots \cap (\beta_{J_2} = 0) \quad (2.90)$$

nicht verworfen werden kann. Kalibriert man jedes Modell separat und berechnet die daraus resultierenden minimalen Fehlerquadratsummen S_{\min}^{M1} und S_{\min}^{M2} , so sind *alle* zusätzlichen Einflussfaktoren des Modells M2 irrelevant (bzw. nicht als relevant nachzuweisen), falls die Realisierung f der Fisher'sche Test-Statistik

$$F = \frac{\frac{S_{\min}^{M1} - S_{\min}^{M2}}{J_2 - J_1}}{\frac{S_{\min}^{M2}}{n - J_2}} \sim F(J_2 - J_1, n - J_2) \quad (2.91)$$

zu p -Werten

$$p = 1 - F_F^{(J_2 - J_1, n - J_2)}(f) > \alpha = 5\% \quad (2.92)$$

führt.

2 Stetige ökonomische Modelle

Beispiel zur Begleitung: Hotel-Auslastung (vgl. S. 35)

Aus Abb. 2.22 sieht man sofort, dass keinesfalls *beide* Einflussfaktoren (Sternezahl x_1 und Preis x_2) für die Belegung irrelevant sind. Ist aber jeweils eine dieser Einflussfaktoren irrelevant?

- **Test, ob der Preis irrelevant ist:** Es ergeben sich für das Modell

$$M_1 : \hat{y} = \beta_0^{M1} + \beta_1^{M1} x_1$$

im Vergleich mit dem vollen Modell

$$M : \hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

und seiner Fehlerquadratsumme $S_{\min}^{M1} = S_{\min} = 498$ die Größen

$$S_{\min}^{M1} = 1762, \quad f = 22.8, \quad p = 0.1 \%$$

Der p -Wert ist deutlich unter 5%. Der Preis ist also ein signifikanter Einflussfaktor und seine Berücksichtigung eine notwendige Bedingung, um ein korrekt spezifiziertes Modell dieses Sachverhaltes zu erhalten.

- **Test, ob die Sternezahl irrelevant ist:** Es ergeben sich für das Modell

$$M_2 : \hat{y} = \beta_0^{M2} + \beta_2^{M2} x_2$$

im Vergleich mit dem vollen Modell die Größen

$$S_{\min}^{M2} = 3605, \quad f = 56, \quad p = 0.004 \%$$

Der p -Wert ist unter 0.1%. Die Sternezahl ist also ein hochsignifikanter Einflussfaktor.

Schließlich sei noch darauf hingewiesen, dass bei nur einem Einflussfaktor Unterschied (wie hier) der F -Test zum T -Test des entsprechenden Parameters des komplexeren Modells auf den Wert null exakt äquivalent ist: Man kann die F -Statistik in diesem Fall in eine T -Statistik umrechnen. Konsequenterweise sind die p -Werte des F -Tests und die der entsprechenden T -Tests (vgl. das Ende des Abschnitt 2.6.5) identisch. Dies gilt aber nicht mehr bei mehr als einem Parameter Unterschied.

2.7 Nichtmetrische exogene Variable

Es gibt, von Feinheiten abgesehen, drei Skalierungsarten von Merkmalen bzw. Variablen:

- **Metrische** bzw. **kardinalskalierte** Variable haben Ausprägungen (Werte), wel-

2 Stetige ökonomische Modelle

che sich addieren und multiplizieren lassen. Dies ist der übliche Variablentyp bei Regressionsmodellen. Die Operatoren $+$, $*$, $>$ und $'='$ (sowie Varianten wie $-$, $/$, \leq und \neq) sind definiert.

- **Ordinale** oder **ordinalskalierten** Variablen haben Werte, welche sich in eine Rangfolge anordnen lassen (z.B. Schulnoten): Die Operatoren $>$ und $=$ sind definiert
- Bei **Qualitativen** bzw. **nominalskalierten** bzw. **kategorial skalierten** Variablen hingegen kann man nur sagen, ob sie denselben oder unterschiedliche Wert haben wie eine andere qualitative Variable: Nur der Operator $'='$ ist definiert. Aus der Definition folgt, dass nichtmetrische Variable immer diskretwertig sind.

Während bei den Regressionsmodellen die endogene Variable immer metrisch (und kontinuierlich bzw. quasikontinuierlich) sein muss, können die exogenen Variablen auch einem anderen Wertetyp angehören. Da in diesen Modellen jedoch auch exogene Variable multipliziert und addiert werden, müssen solche Variable durch *Platzhalter-* bzw. **Dummyvariablen** ersetzt werden.

Damit der Charakter der Werte solcher Variablen (nur $'='$ bzw. \neq ist definiert) nicht verfälscht wird, kommen ausschließlich binärwertige (*dichotome*) Dummyvariable zum Einsatz, die typischerweise die Werte 1 (“trifft zu”) und 0 (“trifft nicht zu”) annehmen können. Ferner sollte das Merkmal nicht **häufbar** sein, also immer genau einen Wert annehmen: Das Merkmal “Schulabschluss” ist beispielsweise häufbar (und ordinalskaliert), während das Merkmal “höchster Schulabschluss” nichthäufbar ist.

Um einen nichtmetrischen (ordinalen oder nominalen) nichthäufbaren Einflussfaktor mit m möglichen Ausprägungen in ein Regressionsmodell zu bringen, wird dieser Einflussfaktor durch $m - 1$ dichotome exogene Dummyvariable modelliert.

Hat man ausschließlich nichtmetrische exogene Variable, bezeichnet man die Regressionsanalyse auch als **Varianzanalyse**. Die exogenen Variablen werden dann auch als *Faktoren* und deren Ausprägungen als *Faktorstufen* bezeichnet.

2 Stetige ökonometrische Modelle

Beispiel: Kraftstoffverbrauch

Wichtige Einflussfaktoren für den auf 100 km bezogene Verbrauch y sind das Fahrzeuggewicht x_1 , die Motorleistung x_2 , das Alter x_3 des Fahrzeugs, die Kraftstoffsorte (Benzin oder Diesel) und der Streckentyp (Stadt-Landstraße-Autobahn). Während x_1 bis x_3 metrisch sind, sind die letzten beiden Einflussfaktoren qualitative Variablen mit 2 bzw. 3 Ausprägungen. Man bringt sie in das Regressionsmodell durch die dichotomen Variablen x_4 (=1 bei Benzin, 0 bei Diesel), x_5 (=1 bei Stadtstraßen, 0 sonst) und x_6 (=1 bei Landstraßen, 0 sonst.). Wegen der Nichthäufbarkeit ist keine Variable x_7 (=1 bei Autobahnen, 0 sonst) notwendig oder erlaubt.

Im Falle einer einzigen dichotomen qualitativen Variable und keiner weiteren exogenen Variablen ist der Test auf Relevanz dieser Variable identisch zum nichtverbundenen Differenztest.

Beispiel: Verbrauchen Dieselfahrzeuge weniger als Benzinler?

Gegeben sind n Kfz, davon ist die Hälfte mit Benzin und die Hälfte mit Diesel betrieben. Hängt der Verbrauch signifikant von der Kraftstoffsorte ab? Untersuchung mit dem Regressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \epsilon, \quad x_1 = \begin{cases} 1 & \text{Benzin} \\ 0 & \text{Diesel.} \end{cases}$$

Schätzung von β_1 :

$$\hat{\beta}_1 = \frac{s_{1y}}{s_x^2} = \bar{y}^B - \bar{y}^D$$

wobei \bar{y}^B bzw. \bar{y}^D die Verbrauchsmittelwerte der Benzin- und Dieselfahrzeuge darstellen, wenn man sie getrennt betrachtet. Schätzung der Unschärfe von β_1 :

$$\hat{V}(\hat{\beta}_1) = \hat{V}_{11} = \frac{\hat{\sigma}_\epsilon^2}{n s_{11}} = \frac{2}{n-2} (s_{y,B}^2 + s_{y,D}^2)$$

wobei $s_{y,B}^2 = 2/n \sum_{i=1}^{n/2} (y_i - \bar{y}^B)^2$ und analog $s_{y,D}^2$ die deskriptiven Varianzen der Verbrauchsschwankungen der Benzin- und Dieselfahrzeuge sind, wenn man diese getrennt betrachtet.

Das Ergebnis ist für $n \gg 1$ nichts anderes als der in diversen Statistikbüchern (und der Statistik-2 Vorlesung) betrachtete Differenztest unverbundener Stichproben (für $n_1 = n_2 = n/2$). Die Formulierung als Regressionsmodell ist jedoch auch gültig, wenn die Bedingungen $n_1 = n_2 \gg 1$ nicht zutreffen.

2 Stetige ökonometrische Modelle

2.8 Logistische Regression

Bekanntlich hat das allgemeine Regressionsmodell die Form (2.27), also

$$Y(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x} + \epsilon, \quad \epsilon \sim \text{i.i.d.}N(0, 1).$$

Daraus folgt, aufgrund der reellwertigen Modellparameter und der gaußverteilten Zufallsvariable, dass die endogene Variable Y stetig sein muss. Bisweilen wendet man die Regressionsmodelle aber auch auf Sachverhalte mit einer *diskreten* und/oder *qualitativen*, insbesondere zweiwertigen (*binären*) endogenen Variablen an, z.B. wenn eine Entscheidung ansteht, ob ich einen bestimmten Weg mit dem ÖPNV ($Y = 1$) oder auf andere Art ($Y = 0$) zurücklegen will. In der logistischen Regression wird dabei die normale Regression auf eine nicht-beobachtbare reellwertige Zwischenvariable $Y^*(\mathbf{x})$ angewandt und – im Sinne eines verketteten Modells – diese Zwischenvariable auf den binären Output abgebildet:

$$Y(\mathbf{x}) = \begin{cases} 1 & Y^*(\mathbf{x}) > 0 \\ 0 & \text{sonst,} \end{cases}, \quad Y^*(\mathbf{x}) = \hat{y}^*(\mathbf{x}) + \epsilon = \boldsymbol{\beta}'\mathbf{x} + \epsilon, \quad \epsilon \sim \text{i.i.d. Logistic.} \quad (2.93)$$

Außerdem wird nun der Zufallsanteil der Zwischenvariable, die sogenannte **latente Variable**, als logistisch verteilt statt standardnormalverteilt angenommen. Die logistische Verteilung ist dabei durch die Verteilungsfunktion

$$F(x) = \frac{e^x}{e^x + 1}. \quad (2.94)$$

definiert.

Wahrscheinlichkeit des Ausgangs '1'

Diese ergibt sich direkt durch die Definition der Verteilungsfunktion: Der Wert $F(x) = P(X \leq x)$ gibt die Wahrscheinlichkeit dafür an, dass die Zufallsvariable X den Wert $X \leq x$ hat: Die Wahrscheinlichkeit dafür, dass die binärwertige endogene Variable den Wert $Y = 1$ hat, ergibt sich damit wie folgt:

$$\begin{aligned} P(Y = 1) &= P(Y^*(\mathbf{x}) > 0) \\ &= P(\boldsymbol{\beta}'\mathbf{x} + \epsilon > 0) \\ &= 1 - P(\boldsymbol{\beta}'\mathbf{x} + \epsilon \leq 0) \\ &= 1 - P(\epsilon \leq -\boldsymbol{\beta}'\mathbf{x}) \\ &= 1 - F_\epsilon(-\boldsymbol{\beta}'\mathbf{x}). \end{aligned}$$

Nutzt man aus, dass die Dichtefunktion der Verteilung symmetrisch um 0 ist, erhält man das Ergebnis

$$P(Y = 1) = F_\epsilon(\boldsymbol{\beta}'\mathbf{x}). \quad (2.95)$$

Dies gilt für beliebige symmetrische Verteilungen, insbesondere für die logistische und die Normalverteilung.

2 Stetige ökonomische Modelle

Odds-Ratio und Log-Odds-Ratio

Für die logistische Verteilung ergibt (2.95) die Wahrscheinlichkeiten des Logitmodells,

$$P_1 = P(Y = 1) = \frac{e^{\beta' \mathbf{x}}}{e^{\beta' \mathbf{x}} + 1}. \quad (2.96)$$

Dies kann man auch nach der Zwischengröße $\hat{y}^*(\mathbf{x}) = \beta' \mathbf{x}$ auflösen und erhält damit eine Interpretation dieser Zwischengröße:

$$\hat{y}^*(\mathbf{x}) = \beta' \mathbf{x} = \ln \left(\frac{P_1}{1 - P_1} \right) = \ln \left(\frac{P_1}{P_0} \right). \quad (2.97)$$

Also:

In der logistischen Regression gibt der deterministische Anteil $\hat{y}^* = \beta' \mathbf{x}$ der un beobachtbaren Zwischengröße das logarithmierte Wahrscheinlichkeitsverhältnis $\ln(P_1/(1 - P_1)) = \ln(P(Y = 1)/P(Y = 0))$, oder auf englisch die **Log-Odds-Ratio**, an.

Ordnet man alle Stichprobenelemente mit denselben Werten der exogenen Variablen (Klasse k) einer Klasse k zu, oder hat man a priori makroskopische Daten, z.B. den ÖV-Anteil verschiedener Städte k , kann man das *Odds-Ratio* dieser Klasse durch die beobachteten relativen Häufigkeiten f_{1k} für den Wert $Y = 1$ bestimmen und damit "Datenwerte" für Y^* bekommen:

$$y_{k,\text{data}}^* = \ln \left(\frac{f_{1k}}{1 - f_{1k}} \right). \quad (2.98)$$

Damit könnte man eine normale OLS-Parameterschätzung durchführen, indem man die Fehlerquadratsumme $\sum_k [y_{k,\text{data}}^* - \hat{y}^*(\mathbf{x}_k)]^2$ durchführen. Dies ist jedoch methodisch unsauber und insbesondere nicht anwendbar, wenn eine beobachtete relative Häufigkeit $f_{1k} = 0$ oder $f_{1k} = 1$ ist, da dann $y_k \rightarrow -\infty$ bzw. $y_k \rightarrow +\infty$ geht. Die korrekte und auch in den Statistikprogrammen angewandte Schätzung wird mit der **Maximum-Likelihood-Methode** (Abschnitt 4.6) durchgeführt. Die Unterschiede der beiden Schätzmethoden und die Unzulänglichkeit der "naiven" OLS-Methode werden in Gegenbeispiel auf Seite 88 gezeigt.

Zusammenhang mit Modellen der diskreten Wahltheorie

Im Kontext des *Homo Oeconomicus* ergibt sich für die in (2.93) definierte Zwischenvariable $Y^*(\mathbf{x})$ noch eine Bedeutung: Da der *Homo Oeconomicus* genau dann $Y = 1$ wählt, wenn $Y^* > 0$, kann man $Y = 0$ und $Y = 1$ als zwei wählbare Alternativen sowie Y^* als *Nutzendifferenz* der beiden $Y = 1$ und $Y = 0$ zugeordneten Alternativen auffassen. Mit (2.97) kann man den Zusammenhang wie folgt charakterisieren:

2 Stetige ökonometrische Modelle

- Die Zwischenvariable $Y^* = \beta' \mathbf{x} + \epsilon$ der logistischen Regression mit einer “logistisch” verteilten latenten Variable (Zufallsvariable) entspricht der Nutzen-differenz des binomialen Logit-Modells der diskreten Wahltheorie, wo jeder Nutzen eine gumbelverteilte latente Variable hat,
- Der deterministische Teil $\hat{y}^* = \beta' \mathbf{x}$ dieser Variable ist einerseits gleich dem logarithmierten Odds-Ratio $P_1/(1 - P_1)$ und andererseits gleich der deterministischen Nutzendifferenz.

In Kapitel 4 zeigen wir sogar Äquivalenzen zu Modellen der diskreten Wahltheorie:

Die mit der Maximum-Likelihood-Methode geschätzte logistische Regression ist äquivalent zum **binomialen Logitmodell**, denn die Differenz der unabhängigen gumbelverteilten Zufallsnutzen dieses Modells ist logistisch verteilt. Ist der latente Anteil ϵ der Zwischenvariablen wie in der linearen Regression normalverteilt, ist die logistische Regression identisch zum **binomialen Probitmodell**, denn Differenzen gaußverteilter Zufallsvariablen sind gaußverteilt.

Mehr als zwei Alternativen und Kalibrierung

Unter den Namen **multinomiale logistische Regression** wird das logistische Regressionsmodell auf mehr als zwei Ausprägungen der nominalskalierten bzw. kategorialen exogenen endogenen Variablen Y verallgemeinert. Bei I Ausprägungen bekommt man $I - 1$ Zwischenvariable $Y_i^* = y_i^* + \epsilon_i$. Nimmt man die latenten Variablen ϵ_i wieder i.i.d. logistisch an, erhält man dieselben Odds-Ratios und Wahrscheinlichkeitsverhältnisse wie bei der binomialen logistischen Regression. Kalibriert man mit der Maximum-Likelihood-Methode, ist die multinomiale logistische Regression identisch zum Multinomial-Logit-Modell (Kap. 4.4.2), da die entsprechende zu maximierende Log-Likelihood nur von den absoluten Häufigkeiten in den Daten und den Wahrscheinlichkeiten P_i für Ausprägung i abhängen.

Da aber die in Kap. 4 betrachteten Modelle der diskreten Wahltheorie anschaulicher formulierbar und vor allem besser verallgemeinerbar sind, wird die logistische Regression, abgesehen vom folgenden “Gegenbeispiel”, nicht weiter betrachtet.

Gegenbeispiel: Naive OLS-Parameterschätzung des logistischen Modells



Die naive OLS-Schätzung der logistischen Regression ist anwendbar, wenn es in der Stichprobe mehrere Klassen k mit jeweils Umfang n_k gibt, die jeweils die gleichen Werte

2 Stetige ökonomische Modelle

\mathbf{x}_k aller exogene Variablen besitzen. Ferner ist der beobachtete Mittelwert der endogenen Variablen in jeder Klasse gleich der relative Häufigkeit f_k , wobei $f_k > 0$ und $f_k < 1$ sein muss (bei der Maximum-Likelihood-Schätzung sind auch $f_k = 0$ und $f_k = 1$ erlaubt). Bezeichnet man (anders als im Haupttext!) mit f_k und P_k die relativen Häufigkeiten bzw. modellierten Wahrscheinlichkeiten dafür, dass in Klasse k die endogene Variable $Y = 1$ ist, ergibt sich mit (2.97) und (2.98) die zu minimierende Fehlerquadratsumme wie folgt:

$$S(\boldsymbol{\beta}) = \sum_k n_k \left[\ln \left(\frac{f_k}{1 - f_k} \right) - \ln \left(\frac{P_k}{1 - P_k} \right) \right]^2 \quad (2.99)$$

Die Maximum-Likelihood-Schätzung des binomiales Logitmodells (siehe Abschnitt 4.7.5 weiter hinten im Skript) maximiert hingegen die Log-Likelihood

$$\tilde{L} = \sum_k n_k [f_k \ln P_k - (1 - f_k) \ln(1 - P_k)].$$

Die Bedingung für die Minimierung dieser Log-Likelihood bezüglich $\boldsymbol{\beta}$ lautet

$$\frac{\partial \tilde{L}}{\partial \boldsymbol{\beta}} = \sum_k n_k \frac{f_k - P_k}{P_k(1 - P_k)} \frac{\partial P_k}{\partial \boldsymbol{\beta}} \stackrel{!}{=} 0. \quad (2.100)$$

Leitet man hingegen die Fehlerquadratsumme (2.99) bezüglich $\boldsymbol{\beta}$ ab, erhält man²⁵

$$\begin{aligned} -\frac{1}{2} \frac{\partial S}{\partial \boldsymbol{\beta}} &= \sum_k n_k \left[\ln \left(\frac{f_k}{1 - f_k} \right) - \ln \left(\frac{P_k}{1 - P_k} \right) \right] \frac{\partial}{\partial \boldsymbol{\beta}} \left(\ln \left(\frac{P_k}{1 - P_k} \right) \right) \\ &= \sum_k n_k \ln \left(\frac{f_k(1 - P_k)}{P_k(1 - f_k)} \right) \frac{1}{P_k(1 - P_k)} \frac{\partial P_k}{\partial \boldsymbol{\beta}} \stackrel{!}{=} 0 \end{aligned}$$

Liegt nun die modellierte Wahrscheinlichkeit P_k in der Nähe von f_k , ist das Argument des Logarithmus nahe 1 und es ergibt Sinn, den Logarithmus in eine Reihe zu entwickeln, $\ln(1 + z) = z - z^2/2 + \dots$. Damit erhält man

$$\sum_k \left(\frac{1 + \mathcal{O}(f_k - P_k)}{P_k(1 - f_k)} \right) n_k \frac{f_k - P_k}{P_k(1 - P_k)} \frac{\partial P_k}{\partial \boldsymbol{\beta}} \stackrel{!}{=} 0. \quad (2.101)$$

Hierbei ist $\mathcal{O}(f_k - P_k)$ eine allgemeine Funktion, die linear gegen null geht, wenn P_k gegen f_k geht. *Der Unterschied zu der Log-Likelihood-Bedingung (2.100) liegt also im ersten geklammerten Faktor.* Unterschiede gibt es vor allem bei relativen Häufigkeiten oder Wahrscheinlichkeiten nahe bei 0 oder 1. Für $f_k = 1$ ist die Schätzvorschrift (2.101) sogar undefiniert, im Gegensatz zur Schätzvorschrift (2.100) des "echten" binomialen Logitmodells. Dies ist klar, da dann der entsprechende y Wert wegen des Nenners $(1 - f_k)$ nicht definiert ist. Für nicht zu extreme relative Häufigkeiten sind die Unterschiede hingegen gering.

²⁵Natürlich verwendet man zur expliziten Schätzung die normale OLS-Formel (2.37) mit den $y_{k,\text{data}}^*$ -Werten aus (2.98), aber hier geht es ja darum, die Unterschiede zu der Maximum-Likelihood-Schätzung zu zeigen, welche durch Auswahlwahrscheinlichkeiten P_k formuliert ist. Deshalb formulieren wir auch die OLS-Schätzung durch die Wahrscheinlichkeiten P_k .

2.9 Bayes'sche Inferenz: Was sagt der p-Wert wirklich aus?

2.9.1 Problemstellung

Das allgemeine Vorgehen bei Signifikanztests ist bekanntlich folgendes:

1. Formulierung der Nullhypothese, z.B. $\beta \leq \beta_0$, wobei β eine beliebige zu schätzende Größe sein kann, z.B. ein Regressionsparameter, eine Korrelation oder der Erwartungswert der endogenen Variablen selbst (dann wäre β der Achsenabschnitt β_0 des Modells $y = \beta_0 + \epsilon$, also $\hat{y} = \beta_0$),
2. Formulierung eines geeigneten, aus den Daten zu bestimmenden Schätzers $\hat{\beta}$ (z.B. mit der OLS-Methode) und einer zur Testentscheidung dienenden Testfunktion,
3. Ausrechnen der Realisierung b von $\hat{\beta}$ aus den Daten,
4. Testentscheidung bzw. Berechnung des p -Wertes.

Zur Rechenerleichterung und aus Gründen einer einfacheren Darstellung nehmen wir im Folgenden bekannte Standardabweichungen σ_b der Schätzer an, was zu Testfunktionen $Z = (\hat{\beta} - \beta_0)/\sigma_b \sim N(0, 1)$ anstelle von $T = (\hat{\beta} - \beta_0)/\hat{\sigma}_b \sim T(n - J - 1)$ führt.²⁶ Um mögliche Probleme dieses klassischen "frequentistischen" Vorgehensweise aufzuzeigen, müssen wir ganz genau anschauen, was

- (i) einerseits der p -Wert genau bedeutet,
- (ii) Wie hoch andererseits die Wahrscheinlichkeit für H_0 ist, falls der Wert b gemessen wurde und wie sich dies zum p -Wert verhält.

Der p -Wert

In Worten kann man den p -Wert wie folgt definieren:

Der p -Wert zu einer Beobachtung b gibt die Wahrscheinlichkeit dafür an, die Beobachtung oder einen "extremere" Wert zu erhalten, falls die Nullhypothese *grenzwertig zutrifft*.

Die Frage, was ein "extremere" Wert ist, hängt von H_0 ab. Generell ist ein Wert umso extremer, je weiter er vom Rand von H_0 entfernt ist. Zur formalen Definition führen wir die zur jeweiligen Nullhypothese H_0 und Beobachtung b gehörige "Extremmenge" E_b ein:

$$E_b = \begin{cases} [b, \infty] & \text{Tests auf "}\leq\text{"}, \chi^2\text{- und } F\text{-Tests} \\ [-\infty, b] & \text{Tests auf "}\geq\text{"} \\ \mathbb{R} \setminus [\beta_0 - |b - \beta_0|, \beta_0 + |b - \beta_0|] & \text{Tests auf "}" \end{cases} \quad (2.102)$$

²⁶Dies ist jedoch keine prinzipielle Einschränkung und ist bei sinnvollen Stichprobengrößen auch quantitativ kaum von Bedeutung.

2 Stetige ökonomische Modelle

Bezeichnet man den Rand der Nullhypothese mit H_0^* , dann gilt allgemein

$$p = P(E_b|H_0^*). \quad (2.103)$$

Hier bezeichnet $P(A)$ die Wahrscheinlichkeit für Ereignis A . Für das obige Beispiel einer “Kleiner-Gleich” Nullhypothese gilt

$$p_b = P(E_b|H_0^*) = P(E_b|\beta = \beta_0). \quad (2.104)$$

Die normale bzw. **frequentistische** induktive Statistik macht also aus einer bedingten Wahrscheinlichkeit der Art

$$\text{Prob}(\text{Messergebnis}|H_0^*)$$

Aussagen der Art

$$\text{Prob}(H_0|\text{Messergebnis}),$$

es werden also bei den Schlussfolgerungen *Ereignisse und Bedingungen vertauscht*. Aussagen über die Wahrscheinlichkeiten bei derartiger Vertauschung macht der Satz von Bayes, den es also eigentlich anzuwenden gilt. Dies ist die Grundlage der **Bayes’schen Statistik**.

Das grundlegende Vorgehen der Bayes’schen Inferenz

Warum wendet man dann den Satz von Bayes nicht immer in der induktiven Statistik an? Der Knackpunkt seiner Anwendung ist, dass man zur Berechnung der eigentlich gesuchten bedingten **A-Posteriori-Wahrscheinlichkeit** für H_0 , also die Wahrscheinlichkeit nach Berücksichtigung der Information durch die Messung,²⁷ die **A-Priori-Wahrscheinlichkeit** $P(H_0)$, also die unbedingte Wahrscheinlichkeit vor der Messung,²⁸ kennen muss, was im Allgemeinen zusätzliche Annahmen impliziert und die ganze Vorgehensweise verkompliziert. A-Priori-Informationen kann man mit Hilfe plausibler Annahmen für den konkreten Sachverhalt oder aus vergangenen Messungen erhalten.²⁹ Bedeutet β beispielsweise einen Zeitwert für Fahrten mit dem ÖPNV (Euro pro Stunde) und ist $H_0 : \beta > 0$, so ist die A-Priori-Wahrscheinlichkeit ziemlich hoch: Auch ohne Messung ist es naheliegend, dass die ÖV-Nutzer zumindest kleine Geldbeträge für eine Zeitersparnis ausgeben würden. Will man die morgige Höchsttemperatur abschätzen, dann repräsentiert die A-Priori-Informationen den Bereich typischer Temperaturen für den gegebenen Ort zur jeweiligen Jahreszeit, ggf. unter Berücksichtigung des heutigen Wetters und des vergangenen Wetterberichts, während die “Messung” das Lesen einer aktuellen Wettervorhersage beinhaltet.

Zur mathematischen Behandlung definieren wir folgendes *Beobachtungsereignis* B als kleines Intervall um die tatsächliche Beobachtung b :

$$B = [b - \delta/2, b + \delta/2]. \quad (2.105)$$

²⁷ *A-posteriori* (lat.) bedeutet “im Nachhinein”.

²⁸ *A-priori* (lat.) bedeutet “von vorneherein” oder “davor”.

²⁹ Wenn von vorneherein gar nichts bekannt ist, behilft man sich ggf. mit einer Gleichverteilung sehr großer Spannweite, welche alle sinnvollen Werte abdeckt.

2 Stetige ökonomische Modelle

Die sehr kleine, aber endliche Konstante (halbe Intervallbreite) δ ist eine Hilfsgröße, um Ereignisse mit endlichen Wahrscheinlichkeiten zu erhalten³⁰ und damit den Satz von Bayes anwenden zu können: Bei der konkreten Rechnung wird sich δ zum Schluss wegekürzen. Die gesuchte A-Posteriori-Wahrscheinlichkeit $P(H_0|B) \approx P(H_0|b)$ (für $\delta \rightarrow 0$ wird dies exakt) hängt nach dem Satz von Bayes oder, direkt ausgehend von der Definition von Wahrscheinlichkeiten von Schnittmengen, $P(B \cap H_0) = P(H_0)P(B|H_0)$, von den A-Priori-Wahrscheinlichkeiten $P(H_0)$ und $P(B)$ wie folgt ab:

$$P(H_0|B) = \frac{P(B \cap H_0)}{P(B)} = \frac{P(H_0)P(B|H_0)}{P(B)}. \quad (2.106)$$

Dies ist die grundlegende Gleichung der **Bayes'schen Inferenz**, also der auf den Satz von Bayes beruhenden induktiven Statistik. Es sei daran erinnert, dass (2.106) nur ausgerechnet werden kann, wenn die A-Priori-Verteilung der zu schätzenden Größe bekannt ist. In diesem Fall kann man nicht nur die unbedingte Wahrscheinlichkeit $P(H_0)$, sondern auch die unbedingte Verteilungsfunktion der Beobachtung $\hat{\beta}$ und damit die unbedingte (nicht von irgendwelchen H_0 -Annahmen abhängige) Wahrscheinlichkeit $P(B)$ einer Beobachtung in B und damit (2.106) für beliebige Nullhypothesen ausrechnen.

Das Ergebnis ist manchmal überraschend. Beispielsweise ergibt (2.106) *exakt* $P(H_0|B) = 0$, wann immer β stetig und H_0 eine Punkt-Nullhypothese ist: Einfach dadurch, dass die A-Priori-Wahrscheinlichkeit $P(H_0)$ gleich null $=0$ ist, während die Wahrscheinlichkeiten $P(B|H_0)$ sowie $P(B)$ endlich sind.

Im weiten Netz gibt es von "Bayes-Revolutionären" oft sehr drastisch geschriebene Artikel, welche den Untergang der klassischen **frequentistischen Statistik**, d.h. den Inhalt aller bisherigen Abschnitte dieses Kapitels und insbesondere des p -Wertes proklamieren. Typischerweise werden hier sehr spezielle A-Priori-Verteilungen angenommen, welche in Einzelfällen aber durchaus gültig sind:

- Beim [Klassiker](#) ist β die Wirkung eines neuen Medikaments und es wird eine diskrete oder gemischte A-Priori-Verteilung angenommen, z.B. $\beta = H_0 = 0$ in 90% der Fälle (Medikament wirkt nicht) und $\beta = c > 0$ bzw. gaußverteilt β in den restlichen 10% (Medikament wirkt).
- Im Verkehrskontext wäre *Map-Matching* (siehe Aufgabe "Map-Matching" auf S. 101) ein Beispiel: Hier ist der wahre Wert β die diskrete bzw. bimodal oder multimodal verteilte Transversalkoordinate der tatsächlich genutzte Straße bzw. des genutzten Fahrstreifens, beispielsweise die Autobahn oder die parallele Nebenstraße. Dann ist
 - H_0 eine bestimmte probeweise Zuordnung, z.B. das Ereignis "das Auto befindet sich auf der Autobahn" (β ist im entsprechenden Intervall),
 - die Beobachtungen b bzw. B werden durch GPS-Messungen realisiert,

³⁰Bekanntlich ist die Wahrscheinlichkeit, einen bestimmten Wert einer kontinuierlichen Zufallsvariablen zu bekommen, exakt gleich Null.

2 Stetige ökonomische Modelle

- die A-Priori-Wahrscheinlichkeit $P(H_0)$ ist gleich dem aus historischen Daten bekannten Verkehrsflussanteil auf der Autobahn,
- die gesuchte Größe ist $P(H_0|B)$: Falls $P(H_0|B) \geq 0.5$, ordnet das Navi – formal wie bei der logistischen Regression – das Fahrzeug der Autobahn zu.

In diesen Fällen gibt es durchaus Situationen bzw. Beobachtungen, welche nach (2.104) einem p -Wert von 5 % entsprechen und bei denen dennoch $P(H_0|B) = 95\%$ statt der eigentlich erwarteten 5 % gilt (siehe Aufgabe “Map-Matching” auf S. 101): Die mit der frequentistischen Statistik *berechnete* Grenz-Fehlerwahrscheinlichkeit einer Ablehnung ist also $\alpha_{\min} = p = 5\%$, während die *tatsächliche* 95 % beträgt!

In Regressionsmodellen sind die Schätzgrößen (die Modellparameter) nach Konstruktion immer stetig. Nimmt man neben den Gauß-Markow-Annahmen zusätzlich unimodal verteilte A-Priori-Werte an (aufgrund des Zentralen Grenzwertsatzes ist eine Gaußverteilung häufig eine gute Annahme), ist der p -Wert, im Gegensatz zu den obigen Beispielen, eine gute Abschätzung der tatsächlichen A-Posteriori-Wahrscheinlichkeiten, falls

- (1) Die Nullhypothese eine Intervallhypothese ist,
- (2) Die Schätzfehler deutlich kleiner sind als die Streubreite der A-Priori-Werte.

Diese *Resurrection of the Frequentist Statistics* wird im Folgenden gezeigt.

2.9.2 Gaußverteilte a-priori-Werte

Wir wollen die allgemeinen Bayes’schen Inferenzformel (2.106) für die gesuchte A-Posteriori Wahrscheinlichkeit des Zutreffens von H_0 bei Vorliegen der Messung b ,

$$P(H_0|B_b) = \frac{P(B_b \cap H_0)}{P(B_b)} = \frac{P(H_0)P(B_b|H_0)}{P(B_b)}, \quad (2.107)$$

für eine gaußverteilte A-Priori-Verteilung des wahren Wertes³¹ ausrechnen. Ohne Einschränkung der Allgemeinheit nehmen wir um Null verteilte wahre Werte,

$$\beta \sim N(0, \sigma_\beta^2)$$

und die Intervall-Nullhypothese

$$H_0 : \beta \leq \beta_0$$

an.³² Damit lautet die A-priori-Wahrscheinlichkeit dafür, dass H_0 wahr ist,

$$P(H_0) = \Phi\left(\frac{\beta_0}{\sigma_\beta}\right), \quad (2.108)$$

³¹Der *eigentliche* wahre Wert ist natürlich fest, uns aber (weder vor noch nach der Messung) bekannt, deshalb muss auch für den wahren Wert eine Verteilungsannahme getroffen werden.

³²Das dies keine Einschränkung ist, wird am Ergebnis deutlich: dort wird die A-Posteriori-Wahrscheinlichkeit von H_0 als Funktion des p -Wertes und der A-Priori-Wahrscheinlichkeit $P(H_0)$ durch (2.113) mit (2.115), (2.117) und (2.118) ausgedrückt. Dieser Ausdruck hängt weder vom Erwartungswert der Nullhypothesenverteilung noch von β_0 noch von der Art der Intervall-Nullhypothese (“ \leq ” oder “ \geq ”) ab; allgemeine Werte von β_0 führen wir aber zum einfacheren Verständnis mit.

2 Stetige ökonomische Modelle

wobei $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Die Messung $\hat{\beta}$ von β aus einer Stichprobe sei unverzerrt mit gaußverteilt, vom wahren Wert β unabhängigen Messfehlern $\Delta\hat{\beta} = \hat{\beta} - \beta$, deren Varianz $\sigma_{\hat{\beta}}^2$ als bekannt und von β unabhängig angenommen wird.³³ Da hier sowohl der eigentliche wahre Wert β als auch der Messfehler $\Delta\hat{\beta}$ als Zufallsgrößen aufzufassen sind, ist das Messergebnis $\hat{\beta}$ die *Summe* zweier unabhängiger Zufallsvariablen,³⁴

$$\hat{\beta} = \beta + \Delta\hat{\beta}, \quad \beta \sim N(0, \sigma_{\beta}^2), \quad \Delta\hat{\beta} \sim N(0, \sigma_{\hat{\beta}}^2), \quad (2.109)$$

und somit gaußverteilt mit Erwartungswert 0 und Varianz gleich der Summe der Einzelvarianzen. Um nun die A-Priori-Wahrscheinlichkeit $P(B)$ für das Beobachtungsintervall $B = [b - \delta/2, b + \delta/2]$ um die Messung sowie $P(B \cap H_0)$ zu berechnen, rastern wir alle Möglichkeiten, die zum Messergebnis b bzw. dem Intervall B führen können, über die möglichen wahren Werte β durch. Dabei sind bei der unbedingten Wahrscheinlichkeit $P(B)$ alle möglichen wahren Werte erlaubt, bei der kombinierten Wahrscheinlichkeit $P(B \cap H_0)$ nur Werte $\beta \in H_0$. Setzt man für β die Dichtefunktion $g(\beta)$ und für $\Delta\hat{\beta}$ die Dichtefunktion $h(\beta)$ an, führt dies zu folgenden unbedingten bzw. kombinierten Wahrscheinlichkeitsdichten für ein Messergebnis bei b :

$$f_{\hat{\beta}}(b) = \int_{-\infty}^{\infty} g(\beta)h(b - \beta) d\beta, \quad (2.110)$$

$$f_{\hat{\beta}}(b \cap H_0) = \int_{-\infty}^{\beta_0} g(\beta)h(b - \beta) d\beta. \quad (2.111)$$

Man beachte, dass (2.110) nichts anderes als den Faltungssatz für die Dichte der Summe unabhängiger Zufallsvariablen angibt und sich die kombinierte Dichtefunktion (2.111) davon nur durch das Integrationsintervall unterscheidet.³⁵ Die endlichen, aber sehr kleinen Wahrscheinlichkeiten $P(B)$ und $Pf(B|H_0)$ ergeben sich durch Integration von (2.110) bzw. (2.111) über das Intervall B . Für sehr kleine halbe Intervallbreiten δ kann man im Integranden $\hat{\beta} = b = \text{const.}$ setzen mit dem Ergebnis

$$P(B) = P(\hat{\beta} \in [b - \delta/2, b + \delta/2]) \approx 2\delta f_{\hat{\beta}}(b), \quad (2.112)$$

$$P(B \cap H_0) = P(\hat{\beta} \in [b - \delta/2, b + \delta/2] \cap H_0) \approx \delta f_{\hat{\beta}}(b \cap H_0). \quad (2.113)$$

Setzt man diese beiden Ausdrücke in die Bayes'sche Inferenzformel (2.106) ein, kürzt sich δ heraus und man erhält das gesuchte $P(H_0|B) \rightarrow P(H_0|b)$ als Funktion zweier Integrale. Für normalverteilte Dichten kann man diese ausrechnen und erhält die

³³Diese vereinfachenden Annahmen sind harmlos.

³⁴Achtung! $\hat{\beta}$ hat nun neben der üblichen Varianzkomponente $\sigma_{\hat{\beta}}^2$ bei festen β auch noch die Varianzkomponente der A-Priori Verteilung von β !

³⁵Streng genommen ist (2.111) nur eine Pseudo-Dichtefunktion, da das Integral über b nicht 1, sondern $P(H_0)$ ergibt.

2 Stetige ökonometrische Modelle

A-Posteriori Wahrscheinlichkeit für H_0 in Abhängigkeit des realisierten Beobachtungswertes b (bzw. B), der Grenze β_0 der Nullhypothese und der Varianzen σ_b^2 und $\sigma_{\hat{\beta}}^2$ der A-Priori-Verteilung bzw. der Messfehler,

$$P(H_0|B) = P(H_0|\hat{\beta} = b) = \Phi\left(\frac{\beta_0 - \mu}{\sigma}\right) \quad (2.114)$$

mit den formalen Hilfsgrößen

$$\mu = b \frac{\sigma_{\hat{\beta}}^2}{\sigma_b^2 + \sigma_{\hat{\beta}}^2}, \quad \sigma = \frac{\sigma_b \sigma_{\hat{\beta}}}{\sqrt{\sigma_b^2 + \sigma_{\hat{\beta}}^2}}. \quad (2.115)$$

Schließlich kann man $P(H_0|B)$ auch noch als Funktion der A-Priori-Wahrscheinlichkeit $P(H_0)$ und des p -Wertes ausdrücken, wenn man Gl. (2.108) für die A-Priori-Wahrscheinlichkeit anwendet und den p -Wert aus der Definition berechnet:

$$\begin{aligned} p &= P(E_b|\beta = \beta_0) \\ &= P(\hat{\beta} > b|\beta = \beta_0) \\ &= P(\hat{\beta} > b - \beta_0|\beta = 0) \\ &= 1 - P(\hat{\beta} \leq b - \beta_0|\beta = 0) \\ &= 1 - \Phi\left(\frac{b - \beta_0}{\sigma_{\hat{\beta}}}\right) = \Phi\left(\frac{\beta_0 - b}{\sigma_{\hat{\beta}}}\right). \end{aligned} \quad (2.116)$$

Drückt man nun umgekehrt b durch den p -Wert und mit (2.108) β_0 durch die A-Priori-Wahrscheinlichkeit aus,³⁶

$$\beta_0 = \sigma_{\hat{\beta}} \Phi^{-1}(P(H_0)), \quad (2.117)$$

$$b = \beta_0 + \sigma_{\hat{\beta}} \Phi^{-1}(1 - p) \quad (2.118)$$

lässt sich die A-Posteriori-Wahrscheinlichkeit (2.114) mit (2.115)-(2.118) als Funktion des p -Wertes und der A-Priori-Wahrscheinlichkeit schreiben. Hat man also erst einmal klassisch den p -Wert bestimmt und kann $P(H_0)$ abschätzen, ergibt sich aus dieser Formel das Ergebnis der Bayes'schen Inferenz. Die Abbildungen 2.25 bis 2.28 zeigen die A-Posteriori-Wahrscheinlichkeit als Funktion von p für verschiedene Varianzverhältnisse und für die H_0 -Grenzen $\beta_0 = 0$ (entspricht $P(H_0) = 0.5$), $\beta_0 = \sigma_{\hat{\beta}}$ (entspricht $P(H_0) = 0.84$), $\beta_0 = 3\sigma_{\hat{\beta}}$ (entspricht $P(H_0) = 0.9987$) und $\beta_0 = -0.5\sigma_{\hat{\beta}}$ (entspricht $P(H_0) = 0.31$). Aus den Formeln und Abbildungen ergeben sich folgende wichtige Erkenntnisse:

- Aus der Herleitung folgt, dass die A-Posteriori-Wahrscheinlichkeit (2.114) mit (2.115)-(2.118) als Funktion des p -Wertes und der A-Priori-Wahrscheinlichkeit unverändert auch für Intervallhypothesen der Form $H_0 : \beta \geq \beta_0$ sowie für beliebige Erwartungswerte der A-Priori-Verteilung von β gelten. Außerdem sind sie unabhängig

³⁶Die Umkehrfunktion $\Phi^{-1}(q)$ gibt die q -Quantile der Standardnormalverteilung an.

2 Stetige ökonometrische Modelle

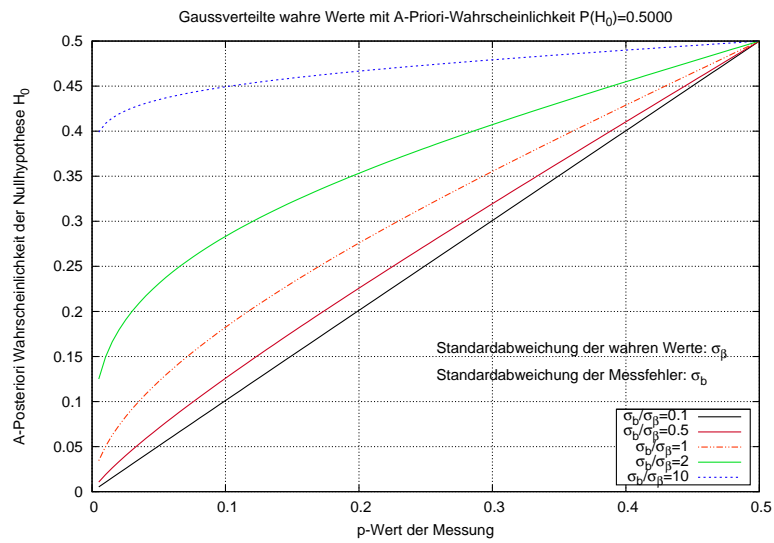


Abbildung 2.25: A-Posteriori Wahrscheinlichkeiten für H_0 als Funktion des p -Wertes

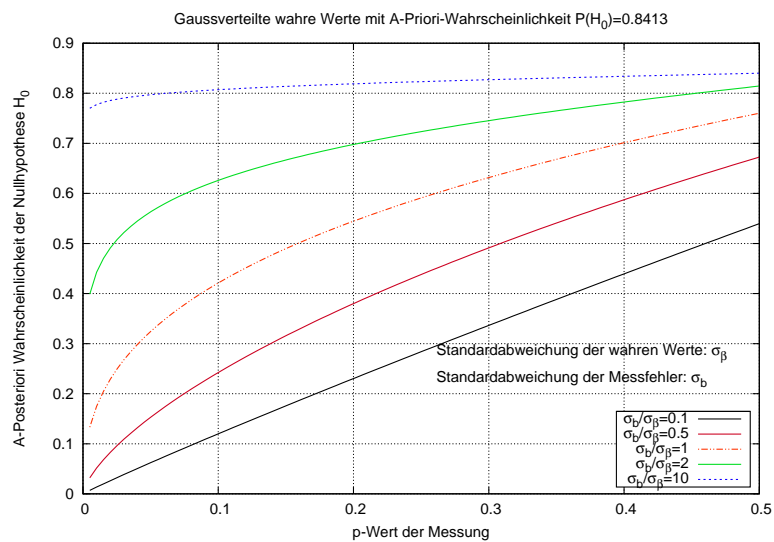


Abbildung 2.26: A-Posteriori Wahrscheinlichkeiten für H_0 als Funktion des p -Wertes

2 Stetige ökonometrische Modelle

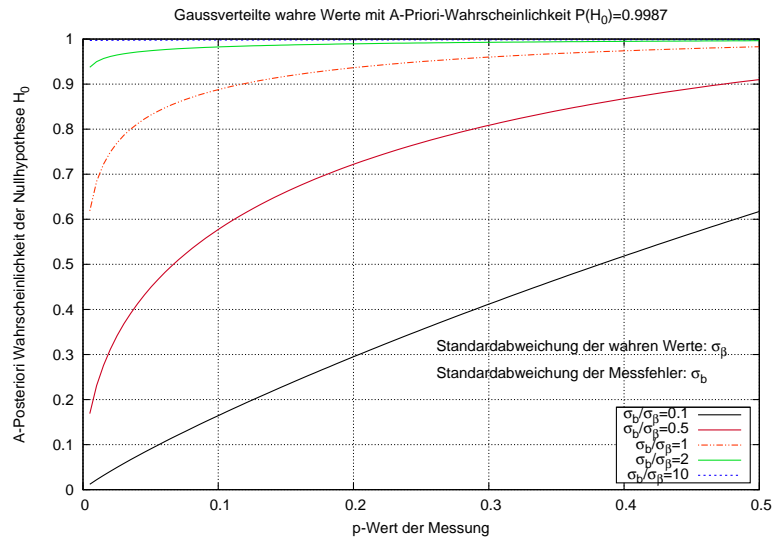


Abbildung 2.27: A-Posteriori Wahrscheinlichkeiten für H_0 als Funktion des p -Wertes

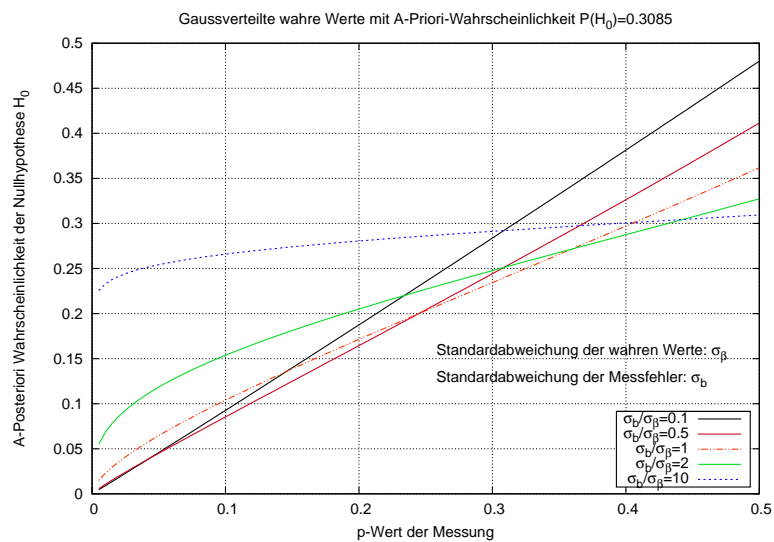


Abbildung 2.28: A-Posteriori Wahrscheinlichkeiten für H_0 als Funktion des p -Wertes

2 Stetige ökonomische Modelle

vom Grenzwert β_0 des H_0 -Intervalls: Weder b noch β_0 tauchen nach Einsetzen von (2.115) - (2.118) in (2.114) auf. Nur bei der Ermittlung von $P(H_0)$ und p selbst muss man die jeweils zutreffenden Nullhypothesen und A-Priori-Verteilungen berücksichtigen!

- Ist der Messfehler sehr viel kleiner als die A-Priori-Schwankungsbreite der Nullhypothese, gibt der p -Wert in guter Näherung die gesuchte A-Posteriori-Wahrscheinlichkeit für H_0 an und zwar weitgehend unabhängig von der A-Priori-Wahrscheinlichkeit: Die Schlussfolgerungen der klassischen “frequentistischen” Statistik, also Konfidenzintervalle, Tests, p -Werte & Co, sind dann also trotz Vertauschung von Messereignis und Bedingung gültig. In der Tat kann man mit Gl. (2.114) zeigen (siehe Übungsaufgabe “Grenzfälle” auf S. 99), dass $P(H_0|B) \rightarrow p$, wenn $\sigma_{\hat{\beta}}/\sigma_{\beta} \rightarrow 0$.
- Ist der Messfehler sehr viel größer als der A-Priori-Schwankungsbereich der gesuchten Variablen, ändert die Messung wenig an der A-Priori-Wahrscheinlichkeit $P(H_0)$. Man benötigt dann schon extreme p -Werte (sehr nahe bei Null oder Eins und sehr weit weg von $P(H_0)$), damit die Messung überhaupt zu signifikanten Änderungen führt. In der Tat kann man mit Gl. (2.114) zeigen (siehe wieder Übungsaufgabe “Grenzfälle” auf S. 99), dass $P(H_0|B) \rightarrow P(H_0)$, wenn $\sigma_{\beta}/\sigma_{\hat{\beta}} \rightarrow 0$. Dies ist unmittelbar anschaulich: Wenn eine Messung ungenauer ist als die A-Priori-Schärfe, kann man nicht viel von ihr erwarten!
- Die Aussagen gelten angenähert auch bei unbekannter Varianz der Messung oder anderer unimodaler A-Priori-Verteilungen, nicht jedoch bei Punkthypothesen und/oder diskreten oder multimodalen A-Priori-Verteilungen, siehe Aufgabe “Map-Matching” auf s. 101 und die Abbildungen 2.29 und 2.30.

Die zentrale Erkenntnis kann man wie folgt zusammenfassen:

Ist die zu schätzende Größe stetig und hat eine unimodale A-Priori-Verteilung, so ist der p -Wert eines Intervalltests eine gute Annäherung an die eigentlich gesuchte A-Posteriori-Wahrscheinlichkeit $P(H_0|B)$. Ist auch nur eine dieser Annahmen nicht erfüllt (die zu schätzende Größe ist diskret oder hat eine multimodale Prior-Verteilung, die Nullhypothese ist eine Punkthypothese, die Schätzfehler sind nicht sehr viel kleiner als die A-Priori-Ungenauigkeit), kann der p -Wert zu drastischen Fehlschlüssen führen. Dann ist eine Bayes'sche Inferenz zwingend, die allerdings immer zusätzliche Modelle bzw. Annahmen impliziert.

2 Stetige ökonometrische Modelle

Aufgabe: Grenzfälle

Werten Sie die A-Posteriori-Wahrscheinlichkeit (2.114) mit (2.118) und (2.117) für gaußverteilte A-Priori-Werte für die Spezialfälle (i) $\sigma_{\hat{\beta}}/\sigma_{\beta} \ll 1$ (genaue Messung) und (ii) $\sigma_{\hat{\beta}}/\sigma_{\beta} \gg 1$ (Messung viel unschärfer als die A-Priori-Information) aus.

Lösung:

- (i) Hier gilt (2.114) mit $\mu \rightarrow b$ und $\sigma \rightarrow \sigma_{\hat{\beta}}$, also

$$P(H_0|\hat{\beta} = b) = \Phi\left(\frac{\beta_0 - b}{\sigma_{\hat{\beta}}}\right) = p.$$

Das zweite Gleichheitszeichen folgt aus (2.116). In diesem Fall gibt der p -Wert also die korrekte A-Posteriori-Wahrscheinlichkeit für H_0 an!

- (ii) Hier gilt (2.114) mit $\mu \rightarrow 0$ und $\sigma \rightarrow \sigma_{\beta}$, also mit (2.108)

$$P(H_0|\hat{\beta} = b) = \Phi\left(\frac{\beta_0}{\sigma_{\beta}}\right) = P(H_0),$$

In diesem Fall ist – unabhängig vom p -Wert – die A-Posteriori-Wahrscheinlichkeit gleich der A-Priori-Wahrscheinlichkeit. Der p -Wert sagt nun also gar nichts aus!

2 Stetige ökonomische Modelle

Aufgabe: MIV-Anteil

Aus vergangenen Umfragen ergab sich in einer gewissen Stadt ein MIV-Anteil von $55\% \pm 3\%$ (Erwartungswert \pm Standardabweichung der als gaußverteilt angenommenen bisherigen Unschärfen). In der aktuellen Untersuchung ist das Ergebnis hingegen $49\% \pm 3\%$. Die grünen Stadträte wollen dieses Ergebnis nun als verstärktes Umweltbewusstsein ihrer Bürger "verkaufen" und fragen die Statistiker, ob sie die Nullhypothese H_0 : "der aktuelle Anteil ist 55% oder höher" widerlegen können. (i) Wie hoch ist der p -Wert? (ii) Wie hoch ist hingegen die tatsächliche A-Posteriori-Wahrscheinlichkeit, wenn die vergangenen Untersuchungen als A-Priori-Information dienen?

Lösung: Sei β der tatsächliche MIV-Anteil der Grundgesamtheit in Prozent. Dann ist $H_0: \beta \geq \beta_0 = 55$, $b = 49$, $\sigma_\beta = 3$ und $\sigma_{\hat{\beta}} = 3$.

- (i) Der p -Wert für " \geq " Nullhypothesen bei bekannter Varianz ergibt sich analog zur Rechnung für " \leq " im Haupttext zu

$$p = P(\hat{\beta} < b | \beta = \beta_0) = \Phi\left(\frac{b - \beta_0}{\sigma_{\hat{\beta}}}\right) = \Phi(-2) = 0.0227.$$

Der p -Wert bzw. der dazugehörige Test legt also eine Ablehnung von H_0 (und damit in der Tat ein "grüneres" Verhalten) nahe.

- (ii) Zunächst gilt offensichtlich $P(H_0) = 1/2$, da der Rand der Nullhypothese im Zentrum der A-Priori-Verteilung liegt. Wie erwähnt, gilt Formel (2.114) auch für " \geq "-Tests, wenn man sie mit den (unveränderten!) Beziehungen (2.117) und (2.118) als Funktion von p und $P(H_0) = 1/2$ ausdrückt:^a

$$\begin{aligned}\beta_0 &= \sigma_\beta \Phi^{-1}(P(H_0)) = 3\Phi^{-1}(1/2) = 0, \\ b &= \beta_0 + \sigma_{\hat{\beta}} \Phi^{-1}(1 - p) = 3\Phi^{-1}(\Phi(2)) = 6, \\ \sigma &= 2.12, \\ \mu &= 3.0\end{aligned}$$

und damit

$$P(H_0|B) = \Phi\left(\frac{\beta_0 - \mu}{\sigma}\right) = \Phi\left(\frac{-3}{2.12}\right) = 0.079.$$

In Wirklichkeit ist hier die H_0 -Wahrscheinlichkeit also etwa das dreifache des p -Wertes!

^aLetztere Formeln transformieren den Sachverhalt hier automatisch auf einen " \leq "-Test und einen Prior-Erwartungswert von Null um, indem implizit β durch $\beta' = 55 - \beta$ ersetzt wird.

2 Stetige ökonomische Modelle

Aufgabe: Map Matching

Die satellitengestützte Lokalisierung von Navigationsgeräten weist üblicherweise Fehler auf, sodass bei parallelen Straßen die Zuordnung (*Map Matching*) nicht immer eindeutig ist. Es seien eine Autobahn (Transversalkoordinate y der rechten Spur $y = 0$) und rechts von ihr eine parallele Straße (Transversalkoordinate $y = 50$ m im selben Koordinatensystem) gegeben, wobei 80 % des Verkehrs auf der Autobahn stattfindet. Zur Vereinfachung nehmen wir an, dass nur die rechten Spuren benutzt werden, also (in Metern) $y = 0$ oder $y = 50$ (der allgemeine Fall mit zwei Intervallen für y entsprechend der fahrbahnen Breiten der Richtungsfahrbahnen ist qualitativ nicht anders, aber verkompliziert die Rechnung nur unnötigerweise). Das Modell sei $\hat{y} = \beta + \epsilon$ mit der bekannten GPS-Standardabweichung $\sigma = \sqrt{V(\epsilon)} = 10$ m.

- (a) Eine GPS-Messung ergab $\hat{\beta} = b = 20$ m. Testen Sie die Nullhypothese “das Fahrzeug befindet sich auf der Autobahn” bei einer Fehlerwahrscheinlichkeit von 5 % und geben Sie den p -Wert an. Wie groß ist hingegen die tatsächliche Wahrscheinlichkeit (i) vor der Messung unter Berücksichtigung der Verkehrsflussanteile, (ii) nach der Messung?
- (b) Mit welchen zusätzlichen Informationen kann man die A-Posteriori-Wahrscheinlichkeit noch weiter in Richtung Sicherheit (0 oder 100 %) bringen?

Lösung:

- (a) Zunächst der p -Wert des Tests von $H_0 : \beta = 0$ mit der Extremmenge $E = \{\hat{\beta} : |\hat{\beta}| \geq 20\}$ und dem Rand $H_0^* = H_0 : \beta = 0$.

$$p = P(\hat{\beta} \in E | H_0^*) = 2(1 - \Phi((20 - 0)/10)) = 2(1 - \Phi(2)) = 0.0454.$$

Damit wäre die Nullhypothese “das Fahrzeug befindet sich auf der Autobahn” bei $\alpha = 5\%$ abzulehnen.

Die tatsächliche A-Posteriori-Wahrscheinlichkeit ergibt sich wieder durch

$$P(H_0|B) = \frac{P(B \cap H_0)}{P(B)} = \frac{P(H_0)P(B|H_0)}{P(B)},$$

nur werden diesmal die einzelnen Wahrscheinlichkeiten anders berechnet. Mit der Messfehler-Dichtefunktion $h(u) = f_{\Phi}(u/10)/10$ (mit der Dichte $f_{\Phi}(u)$ der Dichte der Standardnormalverteilung) und der Annahme, dass die Autobahn-Fahrer immer auf der rechten Spur ($\beta = 0$) fahren, erhalten wir

$$\begin{aligned} P(H_0) &= 0.8 \text{ (vor der Messung, da 80 \% des Verkehrs auf der Autobahn)} \\ P(B|H_0) &= \frac{\delta}{10} f_{\Phi}\left(\frac{20-0}{10}\right) = 0.00540 \delta, \\ P(B|\overline{H_0}) &= \frac{\delta}{10} f_{\Phi}\left(\frac{20-50}{10}\right) = 0.000443 \delta, \\ P(B) &= P(H_0)P(B|H_0) + P(\overline{H_0})P(B|\overline{H_0}) = 0.00441 \delta, \\ P(H_0|B) &= 0.980. \end{aligned}$$

Bei dieser Messung ist die A-Posteriori-Wahrscheinlichkeit für die Autobahn also 98 % (siehe schwarze Kurve von Abb. 2.29) und nicht unterhalb 5 %, wie der Signifikanztest nahelegt! Allerdings ist die Wahrscheinlichkeit, dass man so irreführende oder noch irreführendere Daten bekommt, sehr gering: *Diese* Wahrscheinlichkeit ist hier nämlich durch den p -Wert gegeben!

- (b) Man könnte z.B. den vergangenen Verlauf berücksichtigen und mit Hilfe der Kontinuität (das Auto springt nicht zwischen den Straßen hin und her) weitere Indizien gewinnen.

2 Stetige ökonometrische Modelle

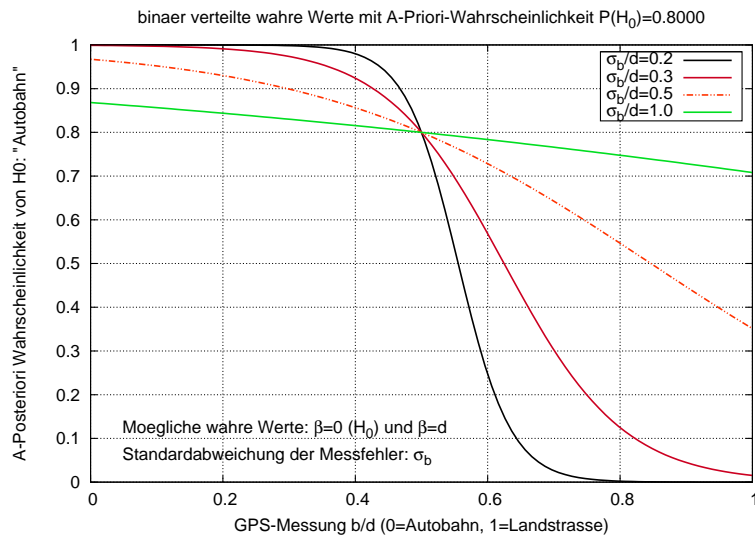


Abbildung 2.29: A-Posteriori Wahrscheinlichkeiten für $H_0 : \beta = 0$, falls die wahren Werte nur $\beta = 0$ und $\beta = d = 50$ betragen können (vgl. Aufgabe)

2.10 Einige Herleitungen

Dieser Abschnitt kann beim ersten Lesen übersprungen werden. Jedoch eignet er sich ideal, um die Zufallseigenschaft der Parameterschätzer zu vertiefen (Abschnitt 2.10.1) oder das Rechnen mit der Matrix-Vektor-Formulierung zu üben (Abschnitte 2.10.2) und 2.10.3).

2.10.1 Statistischen Eigenschaften der Einfachregression "zu Fuß"



Wir betrachten hier die Herleitung der statistischen Eigenschaften des geschätzten linearen Modells bei einer exogenen Variable $x_1 = x$ in Summen-Schreibweise, also ohne Vektoren und Matrizen.

Aus den allgemeinen Gauß-Markow-Annahmen, vor allem der in Kap. 2.3.2 beschriebenen statistischen Spezifikation folgt, dass in den Systemgleichungen

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim i.i.d.N(0, \sigma_\epsilon^2) \quad (2.119)$$

die Messwerte x_i und y_i deterministisch sind und die einzigen Zufallskomponenten in den Größen ϵ_i liegen.

2 Stetige ökonomische Modelle

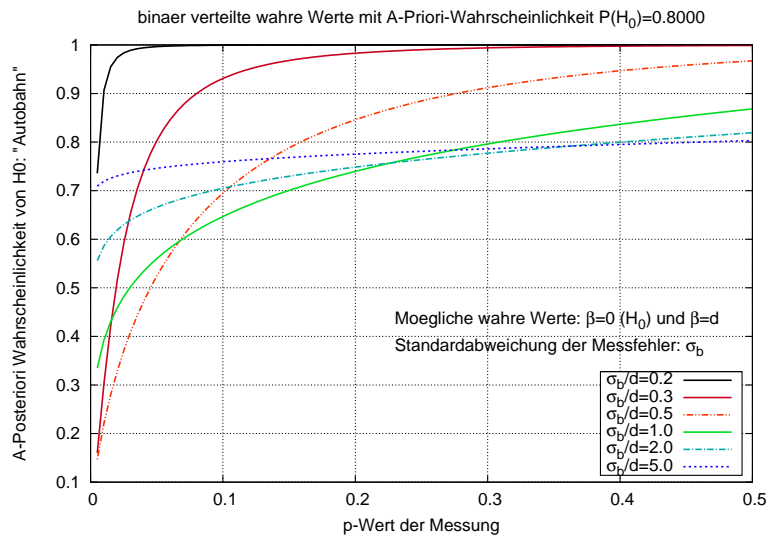


Abbildung 2.30: A-Posteriori Wahrscheinlichkeiten wie in Abb. 2.29, aber nun allgemeiner in Abhängigkeit des p -Wertes des dazugehörigen Tests.

Schätzer des linearen Anstiegsparameters β_1

Zunächst zeigen wir, dass der Schätzer $\hat{\beta}_1$ linear von den Zufallsanteilen ϵ_i abhängt. Aus (2.26) ergibt sich

$$\begin{aligned}\hat{\beta}_1 &= \frac{s_{1y}}{s_{11}} = \frac{1}{ns_{xx}} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{ns_{xx}} \sum_{i=1}^n (x_i - \bar{x}) [\beta_1 x_i + \epsilon_i - (\beta_1 \bar{x} + \bar{\epsilon})] \\ &= \frac{1}{ns_{xx}} \left[b \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x}) \epsilon_i - \bar{\epsilon} \sum_{i=1}^n (x_i - \bar{x}) \right]\end{aligned}$$

und mit $\sum_i (x_i - \bar{x}) = 0$ sowie $\sum_i (x_i - \bar{x})^2 = ns_{xx}$ schließlich

$$\hat{\beta}_1 = \beta_1 + \frac{1}{ns_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \epsilon_i. \quad (2.120)$$

Erwartungswert: Wegen $E(\epsilon_i) = 0$ (Man beachte, dass die x_i nur *Zahlenwerte* sind!) gilt

$$E(\hat{\beta}_1) = \beta_1, \quad (2.121)$$

der Schätzer für den linearen Anstieg ist also erwartungstreu.

2 Stetige ökonomische Modelle

Varianz: Mit der Definition der Varianz und (2.120) ergibt sich

$$\begin{aligned} V(\hat{\beta}_1) &= E\left((\hat{\beta}_1 - \beta_1)^2\right) \\ &= E\left(\frac{1}{n^2 s_{xx}^2} \sum_{i=1}^n \sum_j (x_i - \bar{x})(x_j - \bar{x}) \epsilon_i \epsilon_j\right) \end{aligned}$$

und mit $E(\epsilon_i \epsilon_j) = 0$ für $i \neq j$ und $E(\epsilon_i^2) = \sigma_\epsilon^2$ (vgl. Gl. (2.119)) schließlich

$$V(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n s_{xx}}. \quad (2.122)$$

Schätzer des konstanten Regressionsparameters β_0

Das Vorgehen ist analog zu dem für den linearen Anstiegparameter:

$$\begin{aligned} \hat{\beta}_0 &= \hat{y} - \hat{\beta}_1 \bar{x} \\ &= \beta_0 + \beta_1 \bar{x} + \bar{\epsilon} - \hat{\beta}_1 \bar{x} \\ &= \beta_0 + \bar{\epsilon} + (\beta_1 - \hat{\beta}_1) \bar{x} \end{aligned}$$

und mit (2.43)

$$\hat{\beta}_0 = \beta_0 + \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}}\right) \epsilon_i. \quad (2.123)$$

Erwartungswert: Wegen $E(\epsilon_i) = 0$ gilt

$$E(\hat{\beta}_0) = \beta_0. \quad (2.124)$$

Der Schätzer für den konstanten Parameter ist also ebenfalls erwartungstreu.

Varianz: Die Rechnung geht analog wie bei der Herleitung der Varianz des kalibrierten Anstiegparameters: Mit (2.119) und (2.123) erhält man

$$\begin{aligned} V(\hat{\beta}_0) &= E((\hat{\beta}_0 - \beta_0)^2) \\ &= \frac{\sigma_\epsilon^2}{n^2} \sum_{i=1}^n \left(1 - \frac{(x_i - \bar{x})\bar{x}}{s_{xx}}\right)^2 \\ &= \frac{\sigma_\epsilon^2}{n^2} \left(n + \frac{n\bar{x}^2}{s_{xx}}\right), \end{aligned}$$

(In der letzten Umformung wurde $\sum_i (x_i - \bar{x})\bar{x} = 0$ verwendet) und damit

$$V(\hat{\beta}_0) = \frac{\sigma_\epsilon^2}{n} \left(1 + \frac{\bar{x}^2}{s_{xx}}\right),$$

also (2.56)

2 Stetige ökonomische Modelle

2.10.2 Additionsregel der Varianzen und Bestimmtheitsmaß



Als Übung wird nun die Additionsrelation (2.21) und die Formel (2.22) der erklärten Varianz hergeleitet.

Die erklärte Varianz $s_{\hat{y}}^2$ lässt sich in Matrixnotation folgendermaßen umformen:

$$\begin{aligned} ns_{\hat{y}}^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 \\ &= (\mathbf{X}\hat{\boldsymbol{\beta}})' \cdot (\mathbf{X}\hat{\boldsymbol{\beta}}) - n\bar{y}^2 \\ &= \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} - n\bar{y}^2. \end{aligned}$$

Ersetzt man nun den zweiten Faktor $\boldsymbol{\beta}$ durch $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, erhält man

$$\begin{aligned} ns_{\hat{y}}^2 &= \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2 \\ &= \mathbf{y}' \mathbf{X} \hat{\boldsymbol{\beta}} - n\bar{y}^2. \end{aligned} \tag{2.125}$$

Weiteres Umformen ergibt

$$\begin{aligned} &= \sum_{i=1}^n y_i (\mathbf{X}\hat{\boldsymbol{\beta}})_i - n\bar{y}^2 \\ &= \sum_{i=1}^n y_i \left(\hat{\beta}_0 + \sum_{j=1}^J x_{ij} \hat{\beta}_j \right) - n\bar{y}^2 \\ &= \sum_{i=1}^n y_i \left(\bar{y} - \sum_{j=1}^J \hat{\beta}_j \bar{x}_j + \sum_{j=1}^J x_{ij} \hat{\beta}_j \right) - n\bar{y}^2 \\ &= n\bar{y}^2 + \sum_{j=1}^J \hat{\beta}_j \left(\sum_i x_{ij} y_i - \bar{x}_j \bar{y} \right) - n\bar{y}^2 \\ &= n\bar{y}^2 + n \sum_{j=1}^J \hat{\beta}_j s_{jy} - n\bar{y}^2 \end{aligned}$$

und damit

$$s_{\hat{y}}^2 = \sum_{j=1}^J \hat{\beta}_j s_{jy}$$

also Formel (2.22).

2 Stetige ökonometrische Modelle

Zur Herleitung der Additionsregel schaut man sich die Fehlerquadratsumme näher an:

$$\begin{aligned}
 ns_{\epsilon}^2 &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
 &= \mathbf{y}'\mathbf{y} - (\mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{y} - \mathbf{y}' (\mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{X}\hat{\boldsymbol{\beta}}) \\
 &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\
 &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}
 \end{aligned}$$

mit (2.125) ergibt dies

$$ns_{\epsilon}^2 = \mathbf{y}'\mathbf{y} - n\bar{y}^2 - ns_{\hat{y}}^2 = n(s_y^2 - s_{\hat{y}}^2),$$

also die Additionsregel (2.21).

2.10.3 Fehlerquadratsumme und Schätzer der Residualvarianz



Multipliziert man (2.62) aus, setzt also in $S(\boldsymbol{\beta})$, Gl. (2.33), die Parameterschätzer $\hat{\boldsymbol{\beta}}$ ein, ergibt sich zunächst

$$S_{\min} = ns_{\epsilon}^2 = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Nun wird S_{\min} allein als Funktion der deterministischen Datenmatrix \mathbf{X} der exogenen Variablen und der Residualterme $\boldsymbol{\epsilon}$ ausgedrückt. Sukzessive Ersetzung der abhängigen stochastischen Größen \mathbf{y} und $\hat{\boldsymbol{\beta}}$ durch (2.37) und die Definition des Modells,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

liefert

$$S_{\min} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})' (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) - (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$$

Da nach den Gauß-Markow-Annahmen $E(\boldsymbol{\epsilon}) = 0$ und $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma_{\epsilon}^2 \mathbf{1}$, gilt für den Erwartungswert

$$E(S_{\min}) = E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) - E[\boldsymbol{\epsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon}]$$

Der erste Summand ist

$$E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) = n\sigma_{\epsilon}^2$$

Beim zweiten Summanden kann man leider keine Matrix-Rechenregeln direkt zur Vereinfachung nehmen (insbesondere nicht die Regel $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ anwenden, da weder $\mathbf{A} = \mathbf{X}'$ noch $\mathbf{B} = \mathbf{X}$ invertierbar sind!). Vielmehr muss die Faktoren, welche $\boldsymbol{\epsilon}$ enthalten, in Summen zerlegen, so dass man die Gauß-Markow-Annahmen anwenden

2 Stetige ökonomische Modelle

kann:

$$\begin{aligned}
 E [\boldsymbol{\epsilon}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\epsilon}] &= E \left[\sum_{j=0}^J \sum_{k=0}^J (\boldsymbol{\epsilon}' \mathbf{X})'_j (\mathbf{X}' \mathbf{X})_{jk}^{-1} (\mathbf{X}' \boldsymbol{\epsilon})_k \right] \\
 &= \sum_j \sum_k (\mathbf{X}' \mathbf{X})_{jk}^{-1} E [(\mathbf{X}' \boldsymbol{\epsilon})_j (\mathbf{X}' \boldsymbol{\epsilon})_k] \\
 &= \sum_j \sum_k (\mathbf{X}' \mathbf{X})_{jk}^{-1} E \left(\sum_{i=1}^n x_{ij} \epsilon_i \sum_{l=1}^n x_{kl} \epsilon_l \right) \\
 &= \sum_j \sum_k (\mathbf{X}' \mathbf{X})_{jk}^{-1} \sigma_\epsilon^2 \sum_i x_{ij} x_{ki} \\
 &= \sigma_\epsilon^2 \sum_j \sum_k (\mathbf{X}' \mathbf{X})_{jk}^{-1} (\mathbf{X}' \mathbf{X})_{jk}
 \end{aligned}$$

Da $\mathbf{X}' \mathbf{X}$ symmetrisch ist, kann man das auch schreiben als

$$\begin{aligned}
 E [\boldsymbol{\epsilon}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\epsilon}] &= \sigma_\epsilon^2 \sum_j \sum_k (\mathbf{X}' \mathbf{X})_{jk}^{-1} (\mathbf{X}' \mathbf{X})_{kj} \\
 &= \sigma_\epsilon^2 \sum_j [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}]_{jj} \\
 &= (J + 1) \sigma_\epsilon^2
 \end{aligned}$$

Dies ergibt für die Fehlerquadratsumme den Erwartungswert

$$E(S_{\min}) = (n - J - 1) \sigma_\epsilon^2$$

und damit Gl. (2.63).

2.11 Rechnen mit Vektoren und Matrizen

2.11.1 Wichtige Definitionen

- Spaltenvektor \mathbf{a} mit n Komponenten:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

- Zeilenvektor = transponierter Spaltenvektor:

$$\mathbf{a}' = (a_1, \dots, a_n)$$

- $n \times m$ -Matrix: Rechteckiges Zahlenschema mit n Zeilen und m Spalten:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}.$$

- Transposition: Tausch von Zeilen und Spalten bei Matrizen und Vektoren:

$$(\mathbf{A})_{ij} = a_{ij} \Leftrightarrow (\mathbf{A}')_{ij} = a_{ji}$$

- Die Einheitsmatrix \mathbf{E} bzw. $\mathbf{1}$ ist das neutrale Element bezüglich der Multiplikation quadratischer Matrizen:

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{Es gilt} \quad \mathbf{A} \cdot \mathbf{E} = \mathbf{E} \cdot \mathbf{A} = \mathbf{A}$$

- Die Matrixinverse \mathbf{A}^{-1} einer **regulären**, also quadratischen und invertierbaren Matrix \mathbf{A} ist definiert durch die Matrixmultiplikation:

$$\mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{E}$$

(Man beachte, dass im Allgemeinen bei der Matrixmultiplikation keine Kommutativität gilt!)

Für 2×2 und 3×3 -Matrizen gilt

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - cb} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad (2.126)$$

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}^{-1} = \frac{1}{aei + bfg + cdh - afh - bdi - ceg} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}$$

2 Stetige ökonomische Modelle

- Die Determinante $\text{Det } \mathbf{A}$, manchmal auch als $|\mathbf{A}|$ geschrieben, ist eine multilineare alternierende normierte Abbildung einer quadratischen Matrix auf ein Skalar. Diese existiert und ist eindeutig. Wer dies nicht versteht, merkt sich einfach die Determinanten von 2×2 und 3×3 -Matrizen (dies sind übrigens die Nenner obiger Matrixinversionsformeln):

$$\text{Det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - cb,$$

$$\text{Det} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \text{Det} \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \text{Det} \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \text{Det} \begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

- Der Rang einer rechteckigen Matrix ist die maximale Zahl an Zeilen oder Spalten, innerhalb denen keine Multikolarität (vgl. Gl. (2.7)) herrscht. Quadratische $n \times n$ -Matrizen vom Rang n haben eine Determinante ungleich null und solche vom Rang $r < n$ eine Determinante gleich null.

2 Stetige ökonomische Modelle

2.11.2 Additionen und Multiplikationen

(Die Punkte für Skalar- und Matrixprodukte werden oft weggelassen.)

Operation	Definition	Bedingung	Ergebnis
Vektoraddition	$(\mathbf{a} + \mathbf{b})_i = a_i + b_i$	$n_a = n_b$	Vektor mit n_a Komponenten
Matrixaddition	$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$	$n_A = n_B, m_A = m_B$	$n_A \times m_A$ -Matrix
Zahlenmultiplikation	$(c\mathbf{a})_i = ca_i, (\mathbf{cA})_{ij} = ca_{ij}$	keine	Vektor bzw. Matrix
Skalarprodukt	$\mathbf{a}' \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i$	$n_a = n_b$	Zahl ("Skalar")
Vektorprodukt	$\mathbf{a} \times \mathbf{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$	$n_a = n_b = 3$	3-Vektor
Tensorprodukt	$\mathbf{a} \cdot \mathbf{b}' = \begin{pmatrix} a_1 b_1 & \dots & a_1 b_{n_b} \\ \vdots & & \vdots \\ a_{n_a} b_1 & \dots & a_{n_a} b_{n_b} \end{pmatrix}$	keine	$n_a \times n_b$ - Matrix
Matrixmal Vektor	$(\mathbf{A} \cdot \mathbf{b})_i = \sum_{j=1}^m a_{ij} b_j$	$\mathbf{A} = n \times m$ -Matrix, $\mathbf{b} = m$ -Vektor	n -Vektor
Matrixmultiplikation	$(\mathbf{A} \cdot \mathbf{B})_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$	$\mathbf{A} = n \times m$ -Matrix, $\mathbf{B} = m \times l$ - Matrix	$n \times l$ -Matrix

(2.128)

2.11.3 Wichtige Matrix-Rechenregeln

- Kommutativität bei Skalarprodukten, $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$, aber im Allgemeinen keine Kommutativität bei Matrixprodukten: $\mathbf{AB} \neq \mathbf{BA}$
- Assoziativität bei Matrixprodukten: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- Distributivität bei diversen Produkten: $\mathbf{A}(\mathbf{b} + \mathbf{c}) = \mathbf{Ab} + \mathbf{Ac}$
sowie $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

2 Stetige ökonomische Modelle

- “Kippschalter-Eigenschaft” der Transposition: $(\mathbf{A}')' = \mathbf{A}$
- Transpositionsregeln $(\mathbf{A}\mathbf{b})' = \mathbf{b}'\mathbf{A}'$ sowie $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$
- Für beliebige $n \times m$ -Matrizen \mathbf{X} ist $\mathbf{X}'\mathbf{X}$ eine symmetrische $m \times m$ -Matrix, also

$$(\mathbf{X}'\mathbf{X})_{ij} = (\mathbf{X}'\mathbf{X})_{ji}$$

- Inverse und Transposition sind bei regulären bzw. invertierbaren Matrizen \mathbf{A} vertauschbar:

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

- Für das Inverse eines Matrixproduktes zweier invertierbarer (also insbesondere quadratischer) Matrizen \mathbf{A} und \mathbf{B} sowie bei doppelter Inversion gelten *rein formal* dieselben Regeln wie bei der Transposition:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}, \quad (\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

- Determinante der Matrixinversen: $\text{Det}(\mathbf{A}^{-1}) = 1/(\text{Det } \mathbf{A})$
- Definition der Ableitung von Skalarfunktionen $f(\boldsymbol{\beta})$ nach dem Vektor $\boldsymbol{\beta}$:

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial f}{\partial \beta_0}, \frac{\partial f}{\partial \beta_1}, \dots, \frac{\partial f}{\partial \beta_J} \right)'$$

- Ableitung einer linearen Funktion von $\boldsymbol{\beta}$:

$$\frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta}'\mathbf{a}) = \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{a}'\boldsymbol{\beta}) = \mathbf{a},$$

- Ableitung einer quadratischen Funktion von $\boldsymbol{\beta}$:

$$\frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}) = (\mathbf{A} + \mathbf{A}')\boldsymbol{\beta}.$$

2.11.4 Multivariate Normalverteilung

Dichtefunktion der n -dimensionalen Gaußverteilung mit Erwartungswert $E(\mathbf{z}) = \boldsymbol{\mu}$ und Kovarianzmatrix $E((\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})') = \boldsymbol{\Sigma}$:

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^n \text{Det } \boldsymbol{\Sigma}}} \exp \left[-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z} - \boldsymbol{\mu}) \right]$$

2 Stetige ökonomische Modelle

2.12 Verwendete Symbole

Y	Abhängige bzw. erklärte Variable. Ohne Einschränkung der Allgemeinheit wird nur eine Variable betrachtet.
$x_j,$ $j = 0, \dots, J$	Unabhängige bzw. exogene Variable bei parameterlinearen Modellen
\mathbf{x}	Vektor der exogenen Variablen
β_0	Achsenabschnitt (engl. <i>intercept</i>)
$\beta_j,$ $j = 1, \dots, J$	Lineare Anstiegsparameter
$\boldsymbol{\beta}$	Parametervektor mit den Komponenten β_0, \dots, β_J
ϵ	Ein additiver i.i.d. Zufallsterm (“unbestimmter Anteil”) mit Erwartungswert 0 und der Residualvarianz $V(\epsilon) = \sigma_\epsilon^2$
n	Zahl der Stichprobenelemente bzw. Mess-Sätze aller Variablen
$y_i, x_{ij},$ $i = 1, \dots, n$	Wert der abhängigen bzw. der m -ten unabhängigen Variablen bei der i -ten Messung bzw. dem i -ten Element der Stichprobe.
\mathbf{X}	Systemmatrix der exogenen Variablen mit n Zeilen und $J+1$ Spalten
$\mathbf{y}, \boldsymbol{\epsilon}$	Vektor der gemessenen Werte der endogenen Variablen y_i und des Restterms ϵ_i
$\hat{\boldsymbol{\beta}}, \hat{\sigma}_\epsilon^2, \hat{y}_i$ usw.	Schätzer der entsprechenden Größen
s_{jm}, r_{jm}	Elemente der deskriptiven Kovarianzmatrix der exogenen Variablen bzw. daraus abgeleitete Korrelationen
s_{jy}	Deskriptive Kovarianz zwischen den exogenen und der endogenen Variable
$s_y^2, s_{\hat{y}}^2$	Deskriptive Varianz der endogenen Variable und ihres OLS-Schätzers
s_ϵ^2	Deskriptive Varianz des unbestimmten Anteils
$S(\boldsymbol{\beta})$	Fehlerquadratsumme
$S_{\min} = S(\hat{\boldsymbol{\beta}})$	Minimale Fehlerquadratsumme (OLS)
B, \tilde{B}	Bestimmtheitsmaß, korrigiertes Bestimmtheitsmaß

2 Stetige ökonomische Modelle

$E(\cdot)$	Erwartungswert der statistischen Größe im Argument
$V(\cdot), \text{Cov}(\cdot, \cdot)$	Varianz und Kovarianz der statistischen Größen im Argument
$\mathbf{V} = \mathbf{V}_{\hat{\beta}}$	Varianz-Kovarianz-Matrix der Parameterschätzer
$\hat{\mathbf{V}}$	Schätzer der Varianz-Kovarianz-Matrix
$V(\hat{\beta}_j) = \hat{\sigma}_{\hat{\beta}_j}^2$	Varianz der Parameterschätzer
$f_{\hat{\beta}}(\Delta\hat{\beta})$	Dichtefunktion der Parameterschätzer in Abhängigkeit der Abweichungen $\Delta\hat{\beta} = \hat{\beta} - \beta$ vom wahren Wert
\mathbf{H}	Hessematrix der zweiten Ableitungen von $S(\beta)$
α	Fehlerwahrscheinlichkeit bei Konfidenzintervallen und statistischen Tests
p	p -Wert, minimale Fehlerwahrscheinlichkeit, bei der ein Test gerade noch abgelehnt werden kann
$T_{\hat{\beta}_j}$	Student-t verteilte Test-Variable für den Schätzer $\hat{\beta}_j$
β_{j0}	(Grenzwert der) Nullhypothese H_0 des entsprechenden Tests.
$T(n)$	Student-t-Verteilung mit n "Freiheitsgraden"
$t_q^{(n)}$	q -Quantil der Student-t Verteilung mit n "Freiheitsgraden"
$F(n, m)$	F -Verteilung mit n Freiheitsgraden im Zähler und m im Nenner
$G(\Delta T)$	Gütefunktion in Abhängigkeit der normierten Abweichung ΔT des wahren Parameterwertes von der Nullhypothese

3 Datengewinnung: Verkehrserhebungen

Die Erhebung der Daten zum jeweiligen Sachverhalt ist ein wesentlicher Bestandteil der Vorgehensweise in der Ökonometrie. Insbesondere muss die Art der Erhebung und die erhobenen Daten zum vorgesehenen ökonometrischen Modell passen. In diesem Kapitel werden neben der Erhebungsmethodik (Abschnitt 3.2) die für die Verkehrsökonomie wichtigen Datenkategorien kurz vorgestellt: Verkehrsflussdaten (Abschnitt 3.3), Daten zum Mobilitätsverhalten (Abschnitt 3.4) und die direkte Nutzermessung (Abschnitt 3.5). Schließlich wird in Kap. 3.6 die wahlbasierte Conjoint-Analyse als wichtigste Methode der direkten Nutzermessung im Verkehrskontext vorgestellt. Das Flussdiagramm 3.1 zeigt als Übersicht die Einbindung der Datenerhebung in den allgemeinen Ablauf einer ökonometrischen Untersuchung.

3.1 Ablauf einer Erhebung

Eine Datenerhebung (vgl. Abb. 3.2) beginnt mit der Definition der statistischen Einheiten (Merkmalsträger), der räumlich, zeitlich und sachlich abzugrenzenden Grundgesamtheit und der nicht notwendigerweise identischen Ziehungsgrundlage (Abschnitt 3.2).

Die zu erhebenden Merkmale sind durch den Sachverhalt bzw. die Modellspezifikation vorgegeben und müssen sowohl die exogenen als auch die endogenen Variablen dieser Modelle enthalten.

Anschließend wird die Erhebung selbst spezifiziert: Meist wird man als Ziehungs-methode eine Stichprobe wählen, deren Umfang man in Abhängigkeit der angestrebten Genauigkeit festlegt (vgl. Statistik-Vorlesungen). Die Modalität der Stichprobe (Zufallsstichprobe, geschichtete Stichprobe, Klumpenstichprobe) richtet sich einerseits nach den zur Verfügung stehenden Mitteln (Klumpenstichproben sind bei gleichem Umfang weniger aufwändig zu erheben) und nach der Priorität, welcher auf Repräsentativität und die Vermeidung systematischer Verzerrungen gelegt wird (dann ist die Zufallsstichprobe mit Entzerrung optimal).

Je nach Umfang und Komplexität der Erhebung, der angestrebten Rücklaufquote und der verfügbaren Mittel wählt man die Erhebungsform aus (vgl. Abb. 3.2). Eine Erhebung in Form eines persönlichen Interviews ergibt die höchste Rücklaufquote und die höchste Datenqualität, da Missverständnisse sofort geklärt werden können.

Nach Festlegung aller Erhebungsdetails wird zunächst an wenigen Probanden in einem **Pre-Test** geprüft, ob die Erhebung so funktioniert wie vorgesehen, insbesondere, ob es Schwierigkeiten bei der Fragenbeantwortung gibt und sich die Verweigererquote in den anvisierten Grenzen hält. Nach Optimierung des Erhebungsprotokolls wird dann die eigentliche Erhebung durchgeführt.

3 Datengewinnung: Verkehrserhebungen

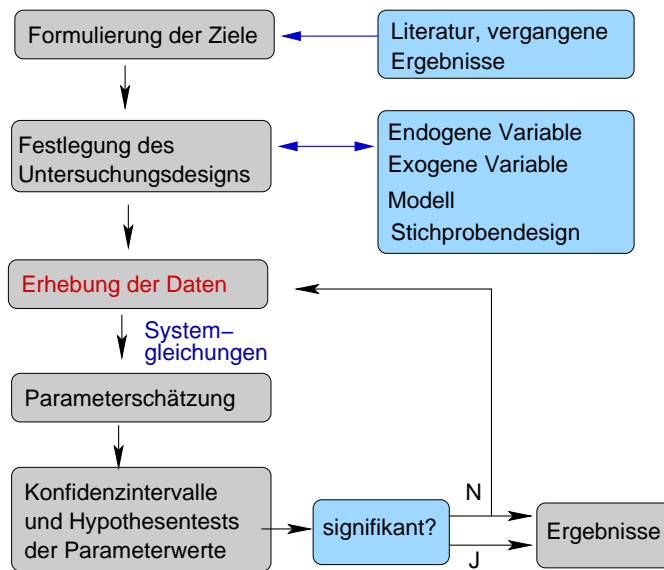


Abbildung 3.1: Eingliederung der Datenerhebung in den allgemeinen Ablauf einer ökonomischen Untersuchung.

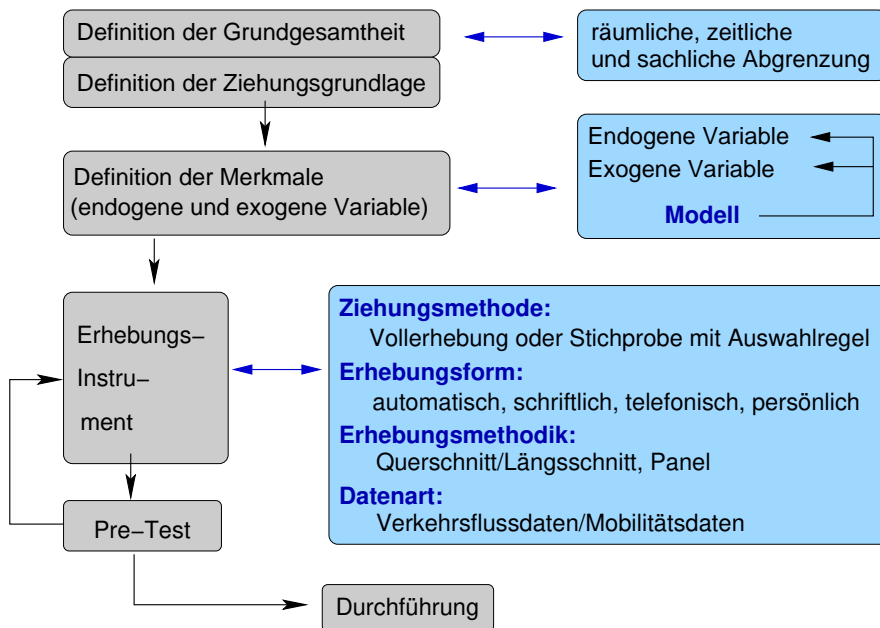


Abbildung 3.2: Ablaufdiagramm einer verkehrsökonomischen Datenerhebung.

3 Datengewinnung: Verkehrserhebungen

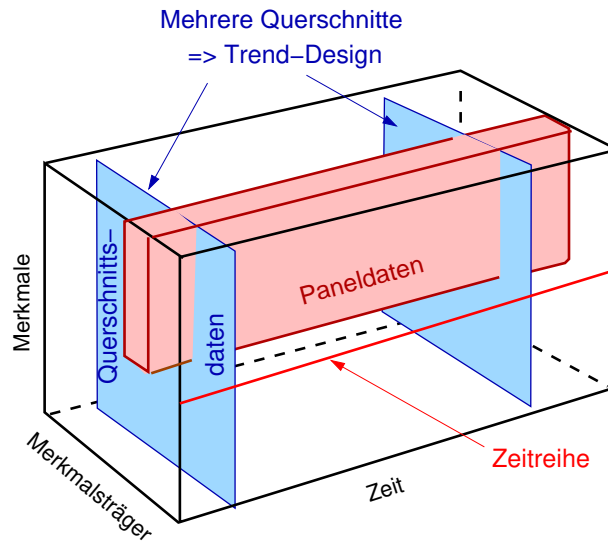


Abbildung 3.3: Verschiedene Datenarten bzw. Erhebungsdesigns, aufgegliedert bezüglich der Dimensionen Zeit, Stichprobenumfang (Achse "Merkmalsträger") und Zahl der Merkmale bzw. typischer Umfang des Fragebogens (Achse "Merkmale").

Die Auswertung der Erhebung beinhaltet schließlich das Aufstellen der Systemgleichungen anhand des vorgesehenen ökonomischen Modells. Diese Systemgleichungen dienen dann als Ausgangspunkt für die Parameterschätzung und den Test der gewonnenen Aussagen.

3.2 Erhebungsdesign

Je nach Art der erhofften Aussagen bzw. der zu prüfenden Hypothesen wird die Erhebungsmethodik bezüglich folgender Kategorien festgelegt:

3.2.1 Aggregierungsebene

Diese reicht von einer globalen Untersuchung der relevanten Merkmale (**makroskopische Ebene**) bis hin zur Berücksichtigung einzelner Merkmalsträger bzw. Individuen (**mikroskopische Ebene**). Den Spezialfall einer Disaggregation makroskopischer Daten nach Altersgruppen bezeichnet man auch als **Kohortendesign**.

3.2.2 Abgrenzung bzw. Zeit- und Merkmalsträgerdimension

Die Unterscheidungen werden in Abb. 3.3 veranschaulicht:

3 Datengewinnung: Verkehrserhebungen

- **Querschnittsdaten** werden zu einem festen Zeitpunkt oder Zeitintervall erhoben und stellen quasi einen “Schnappschuss” der Situation dar.¹
- Bei **Längsschnittdaten** hingegen wird auch die Veränderung in der Zeit berücksichtigt. Dies kann in Form mehrerer Querschnitterhebungen (**Trend-Design**) erfolgen, welche in Zeitabständen von i.d.R. mehreren Jahren mit jeweils neuen Stichproben wiederholt werden. Diese Methodik wird z.B. bei den großen Mobilitätshebungen SrV (System relevanter Verkehrserhebungen) und MiD (Mobilität in Deutschland) angewandt. Untersucht man hingegen ein und dieselbe Stichprobe bzw. Personengruppe fortlaufend über mehrere Jahre, spricht man von **Panel-daten**.² Um einen übermäßigen Ausfall der im Panel enthaltenen Personen im Laufe der Zeit bzw. Fluktuationen zu vermeiden, muss man bei solchen Befragungen jedoch die Zahl der erhobenen Merkmale beschränken, was in Abb. 3.3 durch eine gegenüber der Querschnittserhebung reduzierten “Höhe” symbolisiert ist. Ferner ist bei einem Panel-Design der Umfang der Stichprobe (in Abb. 3.3 symbolisiert durch die “Tiefe” des entsprechenden Quaders) in der Regel geringer als bei Querschnitts-Befragungen.
- Zeichnet man hingegen nur einzelne Merkmale kontinuierlich über die Zeit auf, spricht man von einer **Zeitreihe**.

3.2.3 Ausmaß der Kontrolle über den Untersuchungsgegenstand

- **Objektive Messung**. Die Daten werden hier ohne jede Beeinflussung der zu untersuchenden Subjekte und Objekte erhoben.³ Dies führt zu maximaler Objektivität, da durch Überprüfung der Messgeräte der Wahrheitsgehalt der Messung im Prinzip immer getestet werden kann. Beispiele hierfür sind Verkehrszählungen sowie Messung anderer objektiver Verkehrseigenschaften wie Geschwindigkeit, Verkehrsdichte etc. (Abschnitt 3.3).
- **Erhebung tatsächlicher Ereignisse durch Befragung** (*Revealed Choice*). Hier werden tatsächlich realisierte Wahlentscheidungen wie Wegekettwahl, Verkehrsmittelwahl und Routenwahl und die relevanten erklärenden Variablen wie Alter, Geschlecht, Zeit- und Geldaufwände für die verschiedenen Verkehrsmittel durch Befragung, in der Regel außerhalb des Verkehrsraums, erhoben (Abschnitt 3.4).

Man erhält eine Vielzahl von Informationen, welche man durch eine objektive Messung nie gewinnen kann: Man sieht einem vorbeifahrenden Kfz oder einem Buspassagier ja nicht an, wo der Ausgangspunkt und das Ziel des Weges ist und welchem Zweck er dient. Andererseits ist der Wahrheitsgehalt viel schwerer zu überprüfen. Fehler ergeben sich nicht nur durch absichtliche Falschaussagen, son-

¹Dies hat nichts mit den “Querschnittsdaten” der Verkehrsflussmessung durch stationäre Detektoren zu tun, vgl. Abschnitt 3.3.

²Die daran beteiligten Personen bezeichnet man auch als Panel.

³Für Physiker: Abgesehen vom quantenmechanische Messproblem.

3 Datengewinnung: Verkehrserhebungen

dem auch dadurch, dass sich die befragten Personen an viele Details nicht oder nur unvollständig erinnern.

- **Experientelles Design: Direkte Nutzermessung** (*Stated Choice*). Hier wird nicht nach realisierten Entscheidungen gefragt, sondern Gedankenexperimente durchgeführt nach dem Motto “Wie würden Sie sich entscheiden, wenn Sie ein Kfz zur Verfügung hätten und einen Weg zurücklegen müssten, der zu Fuß 30 min, mit dem Rad 10 min, mit ÖPNV 20 min und mit dem Kfz 8 min beträgt? Im Gegensatz zur Erhebung realisierter Entscheidungen hat hier der Experimentator viele der exogenen Variablen unter Kontrolle und kann sie genau in den interessanten Bereich bringen. Hingegen sinkt naturgemäß die Objektivität. Verzerrungen und systematische Fehler lassen sich kaum kontrollieren. Näheres in den Abschnitten ab 3.5).
- **Kombiniertes Design** Hier werden Revealed-Choice Daten als Angelpunkt oder **Pivot** einer Stated-Choice Befragung derselben Personen genommen, um den Stated-Choice Fragebogen zu optimieren. Typischerweise wird als Referenz der durch die Revealed Choice Befragung aufgedeckte wahre Sachverhalt verwendet und dieser dann schrittweise verändert.

3.2.4 Ziehungsmethode

Bei jeder ökonomischen Untersuchung muss man zunächst die Merkmalsträger und deren **statistische Gesamtheit** bzw. **Grundgesamtheit** definieren. Dies geschieht (vgl. Statistik-Grundvorlesung) durch

- **räumliche Eingrenzung** (z.B. Untersuchungsgebiet Deutschland oder Dresden),
- **zeitliche Eingrenzung** (Zeitpunkt oder Zeitraum, z.B. Sommer 2008)
- **sachliche Eingrenzung** (z.B. nur erwachsene Personen, einschließlich Ausländer, die in Deutschland leben aber ausschließlich Deutsche, die im Ausland leben).

Die Ziehungsmethode wird dann charakterisiert durch

- Die **Ziehungsgrundlage**. Diese ist bei objektiven Messungen eine Teilmenge der räumlichen und zeitlichen Abgrenzung (z.B. die mit Detektoren ausgestattete Streckenabschnitte) und bei Befragungen meist ein Register, z.B. das Einwohnermelderegister oder das Kfz-Zulassungsregister. Die Ziehungsgrundlage muss nicht identisch zur Grundgesamtheit sein. Vielmehr kann es sich um eine Obermenge handeln (z.B. Einwohnerregister bei einer aus Motorradfahrern bestehenden Grundgesamtheit), um einer Untermenge (viele Gewohnheiten deutscher Einwohner werden anhand einer repräsentativen Durchschnittsstadt erhoben), oder um eine abgeleitete Menge (z.B. Einwohnerregister bei einer aus Haushalten bestehenden Grundgesamtheit).⁴

⁴Dieser Sachverhalt wurde in der Bachelor-Vorlesung in Verkehrsökonomie ausführlich behandelt.

3 Datengewinnung: Verkehrserhebungen

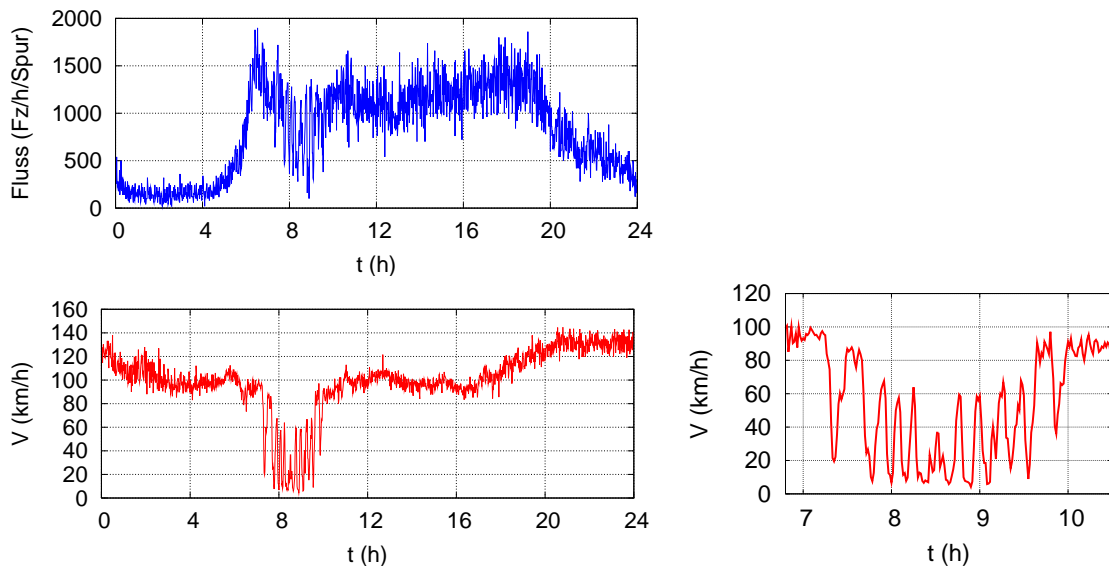


Abbildung 3.4: Verkehrsbelastungs- und Geschwindigkeitsganglinien auf der A9-Süd bei München. Der Ausschnitt zeigt, dass in der morgendlichen Rush-Hour Stop-and-Go-Verkehr herrscht.

- **Vollerhebung** oder **Stichprobe**. Vollerhebungen sind extrem aufwändig und kommen (abgesehen von der Volkszählung) nur für kleine statistische Gesamtheiten in Frage. Bei der Stichprobe wiederum unterscheidet man im Wesentlichen die **Zufallsauswahl**, die **Proportionalauswahl** (geschichtete Stichprobe) und die **Klumpenauswahl**

3.2.5 Modalität der Erhebung

Bei Messungen objektiver Merkmale wird die Erhebung in der Regel automatisch durchgeführt, sie kann aber auch händisch z.B. durch Zählgeräte, Notebookeingabe oder Strichlisten erfolgen. Wichtige Erhebungsmodi bei der Befragung sind

- Persönliche Befragung durch eine Betreuungsperson,
- schriftlich,
- telefonisch,
- und per Internet.

Kombinationen (z.B. telefonisches Nachfragen bei Ausbleiben einer schriftlichen Antwort und Hotlines) sind ebenfalls üblich.

3.3 Beispiel einer objektiven Messung: Verkehrsflussdaten

Verkehrsflussdaten werden direkt im Verkehrsgeschehen an festen Orten erhoben. In der Regel sind sich die Verkehrsteilnehmer (Fußgänger, Radfahrer, Autofahrer) der Erhebung nicht bewusst.

3.3.1 Mesung

Die Erfassung erfolgt in der Regel auf einer der folgenden Arten:

- Personen mit Zählgeräten oder Strichlisten (Stadtverkehr, Untersuchungen für spezielle Zwecke),
- Videoaufzeichnung mit anschließender manueller Auswertung,
- automatisch Messung an **Messquerschnitten**⁵ Die Messung kann dabei erfolgen durch
 - **pneumatische Zähler**: Vor allem Stadtverkehr als temporäre Anwendung,
 - **Induktionsschleifen** oder Infrarot-Messung: In der Regel permanente Anwendung, im Stadtverkehr z.B. als Sensor für verkehrsabhängig gesteuerte Lichtsignalanlagen, auf der Autobahn zur Verkehrslageschätzung und als Sensor für Verkehrsbeeinflussungsanlagen
 - Videocamera mit automatischer Auswertung (Tracking). Damit können neben Kfz-Ströme auch z.B. Fußgängerströme erfasst werden.

Typische erhobenen Merkmale sind die Folgenden:

- Verkehrsfluss im engeren Sinne, d.h. Zahl der Fußgänger oder Kfz, welche einen bestimmten Querschnitt pro Zeitintervall überqueren,
- Geschwindigkeit, entweder über Zeitintervalle aggregiert oder als Einzelmerkmal (nur bei Doppellinduktionsschleifen oder aus Videodaten),
- evtl. Art der Fahrzeuge oder zumindest die Fahrzeuglänge
- evtl. Trajektorien in der Nähe des Querschnitts.

Abbildung 3.4 zeigt als Beispiel den Verkehrsfluss und die in Minutenintervallen aggregierte Geschwindigkeit von Messquerschnitten auf der Autobahn A9-Süd nördlich von München. Man sieht insbesondere, dass es in der morgendlichen Rush-Hour (Maximum im Fluss) zu einem Stau (Zusammenbruch der Geschwindigkeit) kommt. In der Detailansicht der Geschwindigkeits-Zeitreihe (rechts) sieht man, dass es sich um die Stauart "Stop-and-Go-Verkehr" handelt.

⁵Manchmal werden auf diese Weise erhobene Verkehrsflussdaten als "Querschnittsdaten" bezeichnet, da sie sich auf den Querschnitt einer Straße beziehen. In Abschnitt 3.2 wurde dieser Begriff jedoch definiert als Daten, welche zu einem Zeitpunkt bzw. einem Zeitraum erhoben wurden, während die hier gemeinten Daten der Kategorie "Zeitreihe" zuzuordnen sind. Um Verwechslungen zu vermeiden, wird im Folgenden der Begriff "Querschnittsdaten" nur in der in Abschnitt 3.2 eingeführten Bedeutung als "Schnappschuss zu einem Zeitpunkt bzw. Zeitraum" verwendet.

3 Datengewinnung: Verkehrserhebungen

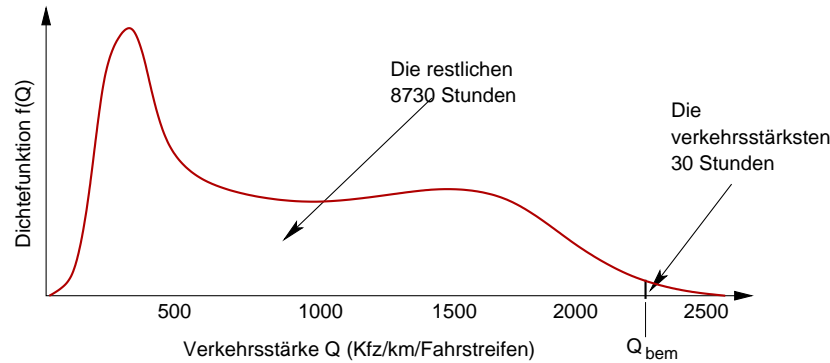


Abbildung 3.5: Zur Definition der Bemessungsverkehrsstärke Q_{bem}

3.3.2 Tagesganglinien und Bemessungsverkehrsstärke

Die über einen Tag aufgenommene Zeitreihe des Flusses bezeichnet man auch als **Tagesganglinie**. Natürlich hängt die Tagesganglinie auch vom Wochentag und (in geringerem Umfang) von der Saison (Sommer, Winter etc) sowie von Sondersituationen (z.B. Ferienanfang, Fußballspiel) ab. Man definiert deshalb verschiedene Fälle, z.B. “Normalwerktag”=Dienstag bis Donnerstag, Freitag, Samstag usw., jeweils in “normalen” Wochen und Wochen mit Sondersituationen.

In ihrer Gesamtheit geben die Tagesganglinien Aufschluss über die Stärke und zeitliche Verteilung der Verkehrsbelastung. Insbesondere lässt sich damit die sogenannte **Bemessungsverkehrsstärke** als Quantil der Verkehrsnachfrageverteilung definieren (vgl. Abb. 3.5):

Die **Bemessungsverstärke** einer Richtungsfahrbahn ist die Verkehrsnachfrage (in Kfz/h oder Kfz/h/Fahstreifen), welcher im statistischen Mittel nur an 30 Stunden im Jahr überschritten wird.

Hintergrund ist, dass man aus ökonomischen Gründen die Kapazität der Straßen nicht nach der maximal zu erwarteten Nachfrage dimensioniert, sondern man vielmehr einen gewissen Zeitanteil mit Streckenüberlastung und daraus resultierenden Staus “in Kauf” nimmt.⁶

Verständnisfrage:

Welchem Quantil der Flussverteilung entspricht obige Definition der Bemessungsverkehrsstärke?

⁶Dasselbe Prinzip gilt bei der Bemessung von Hochwasserverbauungen: Die Schutzmaßnahmen schließen regelmäßig Ereignisse vom Typ “HW200” oder “HW1000” aus, also Hochwasser, welche im Mittel nur alle 200 bzw. 1000 Jahre vorkommen.

3 Datengewinnung: Verkehrserhebungen

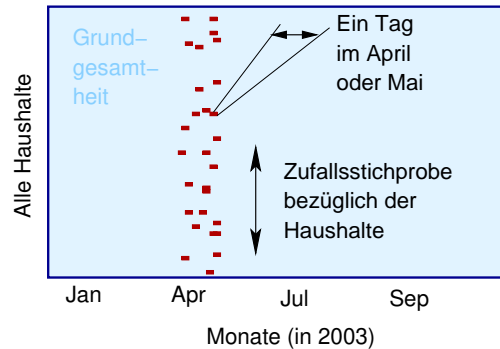


Abbildung 3.6: Visualisierung der Grundgesamtheit (hellblau) und der Stichprobe (dunkelrot) bei der Mobilitätserhebung SrV. Grundgesamtheit und Stichprobe sind dabei als zweidimensionale Mengen in den Dimensionen “statistischen Einheiten” (Haushalte 1 bis N mit N der Zahl der Haushalte in der Grundgesamtheit) und der Zeit aufgetragen.

3.4 Beispiel zu einer Revealed-Choice Erhebung: SrV

Mobilitätsdaten beinhalten alle exogenen und endogenen Variablen des jeweils verwendeten ökonomischen Modells, also im Gegensatz zu den Verkehrsflussdaten nicht nur das Ergebnis in Form von Verkehrsbewegungen, sondern auch das Wozu, das Woher-Wohin, sowie die das Subjekt (den Verkehrsteilnehmer) und seinen/ihren Hintergrund näher beschreibenden sozioökonomischen Variablen. Damit ist klar, dass eine Erhebung nicht automatisch, sondern nur durch *Befragung* durchgeführt werden kann.

Die Erhebung von Mobilitätsdaten soll am Beispiel der in Zusammenarbeit mit der TU Dresden (Prof. Ahrens) durchgeführten Erhebung **System relevanter Verkehrsbefragungen** (SrV) verdeutlicht werden. Die **statistische Gesamtheit** wird charakterisiert durch

- Statistische Einheit: (i) Haushalte (Abb. 3.7), (ii) Personen (Abb. 3.8)
- Räumliche Abgrenzung: 34 Ausgewählte Städte und Gemeinden in Deutschland, u.a. Dresden
- Zeitliche Abgrenzung: Alle “mittleren Werktage” (Dienstag-Donnerstag) im Jahr 2003
- Sachliche Abgrenzung: Alle in den jeweiligen Städten wohnende Personen, also einschließlich Kinder und Ausländer.

Die **Stichprobe** wird aus den *Einwohnermelderegistern* durch Auswahl von *Haushalten* per *Zufallsauswahl* gezogen. Aus den gewählten Haushalten werden *alle Personen* einbezogen. Der *Stichprobenumfang* beträgt etwa 6000 Personen (etwa 3000) Haushalte im Großraum Dresden und etwa ein Sechstel davon in den meisten übrigen Städten/Gemeinden.

3 Datengewinnung: Verkehrserhebungen

Wichtige Hinweise zum Ausfüllen der Fragebögen

- Ihre Teilnahme an der Befragung ist freiwillig. Der Erfolg der Untersuchung hängt jedoch entscheidend von Ihrer Mitarbeit ab.
- Alle Ihre Angaben werden **vertraulich** behandelt und ohne Bezug auf Ihren Namen ausgewertet.
- Bitte füllen Sie den Fragebogen für den Stichtag aus, der auf dem Begleitschreiben genannt ist.
- Für jede Person in Ihrem Haushalt gibt es einen Wegefragebogen. Die Person unter laufender Nummer, z.B. 1, im Personenfragebogen muß mit der Person unter laufender Nummer im Wegefragebogen übereinstimmen.
- Jede Person sollte die Fragen möglichst selbst beantworten.
- Ein Weg ist immer mit einem bestimmten Zweck (z.B. Einkauf, Freizeitweg) verbunden.
- Auf einem Weg können mehrere Verkehrsmittel genutzt werden.
- Bitte tragen Sie alle Wege des ganzen Tages ein und vergessen Sie keinen Weg. Auch Heimwege und kurze Wege z.B. zum Postkasten sind für die Befragung wichtig!
- Bitte füllen Sie die Fragebögen auch dann aus, wenn Sie am Stichtag nicht unterwegs gewesen sind!
- Wenn in Ihrem Haushalt mehr als 5 Personen leben, dann notieren Sie bitte die Personenangaben ab der 6. Person formlos auf einem gesonderten Blatt.
- Wenn eine Person mehr als 7 Wege am Stichtag unternommen hat, dann ist ein zweiter Wegefragebogen auszufüllen und die laufende Nummer der Person zu übernehmen
- Vielen Dank dafür, dass Sie mit der Beantwortung zum Gelingen der Untersuchung beitragen!

Sollten Sie Fragen haben, wenden Sie sich bitte an:

kostenloses Infotelefon
0800 / 60 60 604
 omniphon GmbH

VERKEHRSBEFragung

'Mobilität in Städten - SrV 2003'

durchgeführt von der
 Technischen Universität Dresden
 Lehrstuhl Verkehrs- und Infrastrukturplanung



Bevor Sie die Fragebögen ausfüllen: Lesen Sie bitte die Hinweise im oberen Teil dieser Seite!

Wir empfehlen Ihnen, die Fragebögen in dieser Reihenfolge auszufüllen:

1. Haushaltsfragebogen
2. Personenfragebogen auf der Rückseite dieses Blattes.
3. Wegefragebögen

HAUSHALTSFRAGEBOGEN

Zum Haushalt gehören alle Personen (Sie selbst eingeschlossen), die ständig mit Ihnen zusammenleben. Ein Haushalt kann auch aus einer Person bestehen (Einpersonenhaushalt).

Über welche Telefonnummer(n) ist Ihr Haushalt am besten zu erreichen (z.B. für eventuelle Nachfragen)?
 1.) / Vorwahl Anschluss
 2.) / Vorwahl Anschluss

1 Wieviele Personen leben ständig in Ihrem Haushalt, Sie selbst mit eingeschlossen? Anzahl der Personen insgesamt: <input type="text"/>	2 Welche Fahrzeuge befinden sich im Besitz Ihres Haushaltes? Bitte geben Sie jeweils die Anzahl an! Privat - Pkw <input type="checkbox"/> Dienst - Pkw <input type="checkbox"/> Moped/Motorrad/Motoroller <input type="checkbox"/> Fahrrad <input type="checkbox"/> sonstige Fahrzeuge <input type="checkbox"/>	3 Welchen Pkw-Stellplatz nutzen sie überwiegend? eigener Stellplatz/ Garage/ Carport <input type="checkbox"/> im öffentlichen Raum <input type="checkbox"/>	4 In welcher Zeit ist die von Ihrer Wohnung aus meistbenutzte Haltestelle für Bus und Bahn zu Fuß erreichbar? Fußwegezeit Straßenbahn <input type="text"/> min Fußwegezeit Bus <input type="text"/> min Fußwegezeit S-/bzw.U-Bahn bzw. Nahverkehrszug <input type="text"/> min	5 Wird in Ihrem Haushalt eine übertragbare Fahrkarte für Bus oder Bahn (z.B. übertragbare Monatskarte) genutzt? Ja <input type="checkbox"/> Nein <input type="checkbox"/>	6 Über welche technische Ausstattung verfügt Ihr Haushalt? Bitte geben Sie jeweils die Anzahl der Nutzer an! Festnetztelefon <input type="text"/> Mobiltelefon <input type="text"/> Computer <input type="text"/> Internet <input type="text"/> <small>(Diese Frage dient der Analyse der verkehrlichen Wirkung neuer Medien.)</small>
--	--	---	---	--	---

Bitte beantworten Sie jetzt für alle Haushaltsmitglieder die Fragen auf der Innenseite dieses Fragebogens!

Abbildung 3.7: Haushaltsfragebogen der Mobilitätserhebung SrV 2003.

3 Datengewinnung: Verkehrserhebungen

PERSONENFRAGEBOGEN						
Bitte beantworten Sie jetzt die nachfolgenden Fragen für alle Haushaltsmitglieder!						
Personen im Haushalt		Älteste Person	Zweitälteste Person	Drittälteste Person	Viertälteste Person	Fünftälteste Person
Laufende Nummer der Person		1	2	3	4	5
Geburtsjahr		<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Geschlecht		m <input type="checkbox"/> w <input type="checkbox"/>	m <input type="checkbox"/> w <input type="checkbox"/>	m <input type="checkbox"/> w <input type="checkbox"/>	m <input type="checkbox"/> w <input type="checkbox"/>	m <input type="checkbox"/> w <input type="checkbox"/>
BERUFSTÄTIGKEIT	nicht berufstätig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Vorschulkind	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Hausfrau /-mann	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Rentner /Vormuhestand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Schüler	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	In Ausbildung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Student	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Auszubildender/ Umschüler	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Berufstätig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Volzeit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Teilzeit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ruhendes Beschäftigungsverhältnis (z.B.: Wehrdienst, Kur, Mutterschaftsurlaub)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
z.Zt. arbeitslos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
STELLUNG IM BERUF	Arbeiter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Angestellter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Beamter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Mithelfender Familienangehöriger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Selbständiger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
SCHULABSCHLUSS	ohne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	mit (ohne Abitur)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Abitur	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
BERUFSAUSBILDUNG	ohne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Facharbeiter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Meister	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Hoch- oder Fachschule	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FUHRERSCHEIN-BESITZ	PKW	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
	Moped/Motorrad	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
KFZ-VERFÜGBARKEIT	ständig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Wie oft konnten Sie am Stichtag über ein Kraftfahrzeug Ihres Haushalts verfügen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	teilweise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
gar nicht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
FAHRKARTENART	Einzelfahrschein, Tageskarte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Wenn sie mit öffentlichen Verkehrsmitteln fahren, welche Fahrkartenart nutzen Sie dabei normalerweise hauptsächlich?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Mehrfachkarte, Streifenkarte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Wochenkarte, Monatskarte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jobticket, Semesterticket, etc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
NUTZUNG NEUER MEDIEN	Telearbeit	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Falls Sie zu Hause einen Internetzugang haben: Welches sind Ihre wichtigsten Nutzungszwecke?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Internetshopping	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Internetbanking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	e-mail	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
andere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Abbildung 3.8: Personenfragebogen der Mobilitätserhebung SrV 2003.

3 Datengewinnung: Verkehrserhebungen

Die Stichprobe ist aber nicht nur in der Dimension der statistischen Einheiten ein Teil der Grundgesamtheit, sondern auch in der zeitlichen Dimension: Erhebungen wurden nur im April und Mai 2003 durchgeführt und es wurden pro Haushalt bzw. Person nur die Mobilitätsdaten für einen einzigen *Bezugstag* erhoben (vgl. Abb. 3.6)

Das **Erhebungsdesign** sieht eine Erhebung *tatsächlicher Ereignisse* auf *individueller* Ebene im *Trend-Design* vor, d.h. eine Querschnitterhebung in gewissen (mehrjährigen) Zeitabständen (ab 1972; die letzten beiden Erhebungen in den Jahren 2003 und 2008). Die **Modalität der Erhebung** ist entweder ein *Telefoninterview* oder eine *schriftliche Befragung per Fragebogen*.

Die erhobenen **Merkmale** der Personen können nach mehreren Kategorien unterschieden werden:

- **Allgemeine soziodemografische und sozioökonomische Merkmale:** Wohnort und -Lage, Zugehörigkeit zu Haushalt mit n Personen, Alter, Geschlecht, Beruf, Ausbildungsniveau usw. Diese Merkmale werden im Haushalts- und Personenfragebogen erhoben (Abb. 3.7 bzw. 3.8).
- **Mobilitätsbezogene sozioökonomische Merkmale:** Führerscheinbesitz, Kfz- und Fahrradverfügbarkeit, Besitz von ÖPNV-Dauerkarten, Entfernung zur nächsten ÖPNV-Haltestelle. Diese Merkmale werden ebenfalls im Haushalts- und Personenfragebogen erhoben.
- **Aktivitätenbezogene Merkmale:** Zahl der Wege (definiert als zielgerichtete Ortsveränderung im öffentlichen Verkehrsraum) sowie der Wegezweck (die Zielaktivität). Merkmale dieser und der folgende Datenkategorie werden in den Wegefragebögen (Abb. 3.9) erhoben.
- **Wege- bzw. alternativenbezogene Merkmale:** Start und Zielpunkt eines Weges; Gewähltes Verkehrsmittel sowie Reisezeit, Reiseweite und die Kosten des gewählten Verkehrsmittels für diesen Weg. Aus diesen und den sozioökonomischen Merkmalen sowie den Merkmalen der Infrastruktur lassen sich auch Schätzungen für Reisezeit, Reiseweite und die Kosten der verfügbaren Verkehrsmittelalternativen ableiten.

Je nach Modell handelt es sich dabei um exogene oder endogene Variable. Für ein Verkehrsmittelwahlmodell ist beispielsweise nur das gewählte Verkehrsmittel eine endogene Variable, während alle anderen Merkmale die Rolle von exogenen Variablen spielen. Bei Modellen der Verkehrserzeugung sind die beiden ersten Kategorien exogen, die aktivitätsbezogenen Merkmale endogen und die Wege- bzw. alternativenbezogene Merkmale nicht relevant.

Für die Modellierung ist es relevant, dass die Merkmale verschiedene Skalierungen aufweisen, z.B. dichotom bzw. zweiwertig wie das Geschlecht, allgemein nominalskaliert (z.B. Beruf, gewähltes Verkehrsmittel), ordinalskaliert (z.B. Ausbildungsgrad, Kfz-Verfügbarkeit) oder kardinalskaliert (z.B. Alter, Reisezeiten und -weiten). Ferner ist für die Modellierung wesentlich, dass einige Variable personengebunden sind und sich von Tag

3 Datengewinnung: Verkehrserhebungen

WEGE-FRAGEBOGEN		Wochentag (Stichtag)		Ausgangspunkt des ersten Weges:	
Laufende Nummer der Person		An diesem Tag nicht weggegangen, weil: (Bitte Grund angeben!)		Wohnung <input type="checkbox"/>	
				Anderer und zwar: <input type="checkbox"/>	
				(Straße, Nummer)	
				(PLZ) (Ort)	

	ERSTER WEG	ZWEITER WEG	DRITTER WEG
Um wieviel Uhr haben Sie diesen Weg begonnen?	BEGINN (Uhrzeit) <input type="text"/> : <input type="text"/>	BEGINN (Uhrzeit) <input type="text"/> : <input type="text"/>	BEGINN (Uhrzeit) <input type="text"/> : <input type="text"/>
Zu welchem ZWECK haben Sie diesen Weg unternommen?	ZWECK zur Arbeit <input type="checkbox"/> Kindereinrichtung <input type="checkbox"/> Schule/Ausbildung <input type="checkbox"/> dienstlich/geschäftl. <input type="checkbox"/> Einkauf <input type="checkbox"/> private Erledigung (Bank, Arzt, Dienstleistungen ...) <input type="checkbox"/> Freizeitaktivität <input type="checkbox"/> Holen/Bringen v. Personen nach Hause <input type="checkbox"/> sonstiges, und zwar: <input type="text"/>	ZWECK zur Arbeit <input type="checkbox"/> Kindereinrichtung <input type="checkbox"/> Schule/Ausbildung <input type="checkbox"/> dienstlich/geschäftl. <input type="checkbox"/> Einkauf <input type="checkbox"/> private Erledigung (Bank, Arzt, Dienstleistungen ...) <input type="checkbox"/> Freizeitaktivität <input type="checkbox"/> Holen/Bringen v. Personen nach Hause <input type="checkbox"/> sonstiges, und zwar: <input type="text"/>	ZWECK zur Arbeit <input type="checkbox"/> Kindereinrichtung <input type="checkbox"/> Schule/Ausbildung <input type="checkbox"/> dienstlich/geschäftl. <input type="checkbox"/> Einkauf <input type="checkbox"/> private Erledigung (Bank, Arzt, Dienstleistungen ...) <input type="checkbox"/> Freizeitaktivität <input type="checkbox"/> Holen/Bringen v. Personen nach Hause <input type="checkbox"/> sonstiges, und zwar: <input type="text"/>
Welche VERKEHRSMITTEL haben Sie auf diesem Weg genutzt? (Bitte alle genutzten Verkehrsmittel angeben)	VERKEHRSMITTEL 1. zu Fuß <input type="checkbox"/> 2. Fahrrad <input type="checkbox"/> 3. Moped, Motorrad <input type="checkbox"/> PKW als Fahrer 4. im Haushalts- PKW <input type="checkbox"/> 5. im anderen PKW <input type="checkbox"/> PKW als Mitfahrer 6. im Haushalts- PKW <input type="checkbox"/> 7. anderen PKW <input type="checkbox"/> 8. Bus <input type="checkbox"/> 9. Straßenbahn <input type="checkbox"/> 10. U-Bahn <input type="checkbox"/> 11. S-Bahn, Nahverkehrszug <input type="checkbox"/> 12. Fernzug <input type="checkbox"/> 13. Anderes und zwar <input type="text"/>	VERKEHRSMITTEL 1. zu Fuß <input type="checkbox"/> 2. Fahrrad <input type="checkbox"/> 3. Moped, Motorrad <input type="checkbox"/> PKW als Fahrer 4. im Haushalts- PKW <input type="checkbox"/> 5. im anderen PKW <input type="checkbox"/> PKW als Mitfahrer 6. im Haushalts- PKW <input type="checkbox"/> 7. anderen PKW <input type="checkbox"/> 8. Bus <input type="checkbox"/> 9. Straßenbahn <input type="checkbox"/> 10. U-Bahn <input type="checkbox"/> 11. S-Bahn, Nahverkehrszug <input type="checkbox"/> 12. Fernzug <input type="checkbox"/> 13. Anderes und zwar <input type="text"/>	VERKEHRSMITTEL zu Fuß <input type="checkbox"/> Fahrrad <input type="checkbox"/> Moped, Motorrad <input type="checkbox"/> PKW als Fahrer 4. im Haushalts- PKW <input type="checkbox"/> 5. im anderen PKW <input type="checkbox"/> PKW als Mitfahrer 6. im Haushalts- PKW <input type="checkbox"/> 7. anderen PKW <input type="checkbox"/> 8. Bus <input type="checkbox"/> Straßenbahn <input type="checkbox"/> U-Bahn <input type="checkbox"/> S-Bahn, Nahverkehrszug <input type="checkbox"/> Fernzug <input type="checkbox"/> Anderes und zwar <input type="text"/>
Mit welchem dieser Verkehrsmittel haben Sie den längsten Teilweg zurückgelegt? (bitte Nummer des Verkehrsmittels angeben)	Nummer des Verkehrsmittels <input type="text"/>	Nummer des Verkehrsmittels <input type="text"/>	Nummer des Verkehrsmittels <input type="text"/>
Wo lag das ZIEL dieses Weges? <i>Geben Sie bitte die Adresse an, wenn das Ziel nicht die eigene Wohnung ist!</i>	ZIELADRESSE <input type="text"/> (Straße, Nummer) <input type="text"/> (Ort, Ortsteil) Wohnort <input type="checkbox"/> anderer Ort <input type="checkbox"/>	ZIELADRESSE <input type="text"/> (Straße, Nummer) <input type="text"/> (Ort, Ortsteil) Wohnort <input type="checkbox"/> anderer Ort <input type="checkbox"/>	ZIELADRESSE <input type="text"/> (Straße, Nummer) <input type="text"/> (Ort, Ortsteil) Wohnort <input type="checkbox"/> anderer Ort <input type="checkbox"/>
Um wieviel Uhr sind Sie dort angekommen?	Ankunft (Uhrzeit) <input type="text"/> : <input type="text"/>	Ankunft (Uhrzeit) <input type="text"/> : <input type="text"/>	Ankunft (Uhrzeit) <input type="text"/> : <input type="text"/>
Schätzen Sie bitte die Länge des Weges möglichst genau!	Länge des Weges ca. <input type="text"/> , <input type="text"/> km	Länge des Weges ca. <input type="text"/> , <input type="text"/> km	Länge des Weges ca. <input type="text"/> , <input type="text"/> km
	nächster Weg: nächste Spalte	nächster Weg: nächste Spalte	nächster Weg: nächste Spalte

Bitte wenden! Die weiteren Wege dieses Tages bitte auf der Rückseite eintragen!

Abbildung 3.9: Wegefragebogen der Mobilitätserhebung SrV 2003.

3 Datengewinnung: Verkehrserhebungen

zu Tag nicht ändern⁷ (sozioökonomische Merkmale), andere sind personengebunden und ändern sich täglich (aktivitätenbezogene Merkmale). Schließlich hängen die wege- und alternativenbezogenen Merkmale nicht nur von der Person und dem Tag, sondern auch vom Weg und dem Verkehrsmittel ab.

3.5 Direkte Nutzermessung: Experimentelles Design

In den im vorherigen Abschnitt besprochenen Mobilitätserhebungen werden tatsächlich realisierte Entscheidungen abgefragt und daraus dann mit Hilfe von Modellen der diskreten Wahltheorie Schlussfolgerungen über die Bewertungen der verschiedenen Alternativen gezogen. Deshalb nennt man diese Erhebungen auch **revealed preference**, also “aufgedeckte Präferenzen” (bei der Bewertung des Nutzens). Man kann aber die Präferenzen, also die Gewichtung der Anteile der Nutzenfunktion, direkt messen, indem man die befragten Personen mit *hypothetischen Situationen* konfrontiert, also *Gedankenexperimente* durchführt. Es gibt drei Varianten dieser auch als **experimentelles Design** bekannten Versuchsanordnung:

- (a) **Stated preference.** Verschiedene Alternativen oder auch Aspekte bzw. Merkmale eines Angebots werden mit einer Rangskala bewertet. Beispiel: *Bewerten Sie allgemein die Attraktivität des ÖV-Angebots in Ihrer Stadt mit Schulnoten von 1 bis 6. Likert-Skala 1-5 mit Bild einfüegen!*
- (b) **Stated ranking.** Hier werden Alternativen nach der Rangfolge ihrer Attraktivität (beste Alternative, zweitbeste, etc) geordnet: *Was ist Ihnen im öffentliche Nahverkehr wichtig: Ordnen Sie die Kriterien Fahrpreis, Fahrzeit, Zuverlässigkeit und Sicherheit nach Ihrer persönlichen Wichtigkeit.*
- (c) **Stated choice.** Hier wird nur nach der besten aus zwei oder mehr Alternativen gefragt. Für zwei Alternativen ist dies äquivalent zum Stated Ranking. Beispiel: *Für den Tageseinkauf in einem 3km entfernten Supermarkt stehen Ihnen ein Kfz (Reisezeit 10 min), das Rad (15 min) sowie der Bus (15 min und Kosten 2 €) zur Verfügung. Welches Verkehrsmittel wählen Sie?*

Die *Vorteile* dieser Erhebungsmethode sind

- Einzelne Einflussfaktoren (Zeit, Geld, Flexibilität) sind leichter separierbar und ihre relative Gewichtung (z.B. Zeit/Geld Faktor) direkt ermittelbar,
- da man die Versuchsanordnung und die exogenen Variablen unter Kontrolle hat, kann man genau das erfragen, was für die Untersuchung relevant ist.
- man kann auch das Wahlverhalten nach Realisierung von geplanten Maßnahmen erfragen, für die es naturgemäß keine *Revealed Choice* Erhebung geben kann.

⁷Zumindest innerhalb des persönlichen Planungszeithorizonts für die alltäglichen Ziel- und Verkehrsmittelsentscheidungen. Auf längeren Zeitskalen finden natürlich durch Umzüge, Anschaffungen, berufliche Veränderungen oder auch schlicht durch das Älterwerden Veränderungen statt.

3 Datengewinnung: Verkehrserhebungen

Es gibt aber auch *Nachteile*:

- Subjektive Bewertung einer oft hypothetischen Situation. Die Probanden entscheiden sich bisweilen anders, als sie es bei einer “echten” Entscheidung tun würden,
- Leichter manipulierbar.

3.6 Konkretes Anwendungsbeispiel: Wahlbasierte Conjoint-Analyse

3.6.1 Allgemeines

Die *Stated-Choice* Befragung liefert die empirische Grundlage der wahlbasierten Conjoint-Analyse. Allgemein bezeichnet die **Conjoint-Analyse** (Akronym aus dem Englischen: *considered jointly*), ganzheitlich betrachtet) ein in der Psychologie entwickeltes und nun hauptsächlich im Produktmarketing angewandtes multivariates Verfahren zur gemeinsamen Bewertung der Stärke verschiedener Einflussfaktoren auf den Gesamtnutzen eines bestimmten realen oder fiktiven Produkts.

Für die Verkehrsökonomie, insbesondere zur Analyse der Einflussfaktoren verkehrsbezogener Wahlentscheidungen, ist eine Weiterentwicklung der traditionellen Conjoint-Analyse, die **wahlbasierte Conjoint-Analyse**, relevant. Dabei werden den befragten Personen mehrere Alternativenmengen vorgelegt, woraus im *stated-choice* Verfahren jeweils die (hypothetisch) gewählte Alternative abgefragt wird. In jeder Alternativenmenge (engl. *choice set*) werden alle oder einige Alternativen mit jeweils genau definierten und von *choice set* zu *choice set* im allgemeinen unterschiedlichen Attributwerten zur (hypothetischen) Wahl gestellt:

- Bei der Wahlentscheidung zum Autokauf können die Attribute (Merkmale, Einflussfaktoren, exogene Variable) beispielsweise den Kaufpreis, die Marke, den Typ (Fließheck, Kombi, SUV, Van, Zweisitzer ...), den Verbrauch, die Motorleistung, den Wiederverkaufswert nach 5 Jahren, die Zuladungskapazität und vieles mehr enthalten.
- Bei der Verkehrsmittelwahl sind wichtige Einflussfaktoren sicherlich der Modus als solches, die (Grenz-)kosten, die Reisezeit und die Zuverlässigkeit.

Dieser Fragebogen wird dann mit den Mitteln der in Kapitel 4 vorgestellten **diskreten Wahltheorie** analysiert.

3.6.2 Beispiel: Studentenbefragung

Als einfaches Beispiel zur wahlbasierten Conjoint-Analyse kann als Input eine *Stated-Choice* Befragung zur Verkehrsmittelwahl unter 75 Studenten einer vergangenen Vorlesung dienen. Die Ergebnisse sind in der Tabelle 3.1 wiedergegeben. Dabei wird ein bestimmter Wegezweck angenommen, welcher von den Probanden gut vorstellbar ist

3 Datengewinnung: Verkehrserhebungen

SS 2009

Hypoth. Situation	Altern. 1: Fuß/Rad	Altern. 2: ÖPNV	Wahl h_1	Wahl h_2	Logit: $(h_1)_{\text{pred}}$	Logit: $(h_2)_{\text{pred}}$
1	30 min	30 min + 0€	5	70	3.9	71.1
2	30 min	40 min + 0€	21	54	26.0	49.0
3	30 min	50 min + 0€	63	12	62.7	12.3
4	30 min	60 min + 0€	75	0	73.5	1.5
5	30 min	30 min + 1€	26	49	19.8	55.2
6	30 min	30 min + 2€	50	25	52.5	22.5
7	30 min	30 min + 3€	70	5	70.4	4.6
8	20 min	30 min + 0€	35	40	38.5	36.5
9	15 min	30 min + 0€	64	11	61.7	13.3

Tabelle 3.1: Ergebnis des Stated-Choice-Experiments in der Ökonometrie-Vorlesung im SS 2009. Die letzten beiden Spalten geben die vom kalibrierten Logit-Modell vorausgesagten Entscheidungszahlen an. Das Logit-Modell und seine Parameterschätzung werden in Kapitel 4 besprochen.

(hier: der Weg von zu Hause zur Uni). Das Stated-Choice Experiment besteht nun darin, unter den in der Tabelle dargestellten hypothetischen Situationen (*choice sets*) je eine Entscheidung zu treffen zwischen folgenden Alternativen:

- Alternative 1: zu Fuß oder mit dem Rad, je nach persönlicher Vorliebe,
- Alternative 2: Mit dem ÖPNV.

In den verschiedenen Situationen sind die Alternativen unterschiedlich gut ausgestattet, also mehr oder weniger attraktiv: Als exogene (beeinflussende) Variablen wurden hier lediglich die generischen Variablen

- Komplexe (Haustür-zu-Haustür) Reisezeiten in Minuten
- Zusatzkosten (Ad-Hoc Kosten) einer ÖV Fahrt in €

angenommen. Dabei kann die komplexe Reisezeit, obwohl eine generische Variable, auch *alternativenspezifisch* formuliert werden, indem unterschiedliche Bewertungen für jeden Verkehrsmodus angenommen werden (“eine Minute im ÖV” ist weniger unangenehm als eine zu Fuß”). Mit Zusatzkosten sind die zusätzlichen Kosten gemeint, die unabhängig von ggf. vorhandenen Dauerkarten auf jeden Fall bezahlt werden müssen. Die Probanden müssen sich dies als hypothetische Situation vorstellen, auch wenn in der Realität das Semesterticket alles deckt. Sozioökonomischen Variablen wurden hier nicht erhoben, aber einige sozioökonomische Merkmale sind aufgrund der Zusammensetzung der befragten Personen ohnehin bekannt, z.B. “ist Student” oder “nicht zu alt”.

Die Auswertung direkter Nutzmessungen wird mit der erst später in Kapitel 4 besprochenen **Diskreten Wahltheorie** durchgeführt. Hier wird die Anwendung dieser

3 Datengewinnung: Verkehrserhebungen

mächtigen und für die Verkehrsökonomie zentralen Theorie deshalb nur kurz vorgestellt:

- Zunächst wird für jede Alternative k eine parameterlineare kardinalskalierte Nutzenfunktion $U_k = V_k + \epsilon_k$ formuliert – in Analogie zu den Regressionsmodellen, nur dass der i.i.d. **Zufallsnutzen** ϵ_k nun nicht nur einer Gaußverteilung sondern auch anderen Verteilungen gehorchen kann.
- Die Alternative mit dem höchsten Gesamtnutzen wird gewählt (*Homo Oeconomicus*). Dies resultiert wegen des Zufallsnutzens in *Auswahlwahrscheinlichkeiten* P_k , welche als exogene Variable der diskreten Wahltheorie angesehen werden können.

Für die Erstellung der **deterministischen** Nutzenfunktionen V_k und die Erhebung der geeigneten Merkmale gelten deshalb folgende Prinzipien:

- **Translationsinvarianz:** Man kann zu den Nutzenfunktionen aller Alternativen einen festen Wert hinzuaddieren, ohne dass sich das Ergebnis ändert.
- **Skaleninvarianz:** Wegen dem Prinzip *the winner takes it all*, also einer Entscheidung auf Basis einer Ordinalskalierung (“besser oder schlechter”), ändert sich das Ergebnis nicht, wenn man die Nutzenfunktionen einer beliebigen streng monotonen Funktion unterwirft, solange dies für alle deterministischen *und* stochastischen Nutzenanteile aller Alternativen geschieht. Insbesondere kann man die Nutzenfunktionen mit einer festen positiven Konstante multiplizieren, so dass der deterministische Nutzen *in Einheiten der Standardabweichung des Zufallsnutzens* gemessen wird.

Mit den erhobenen exogenen Variablen T_1 , T_2 (komplexe Reisezeiten für Fuß/Rad bzw. ÖV) und K_2 (ÖV Ad-Hoc-Kosten) ergibt dies bei linearer Formulierung folgende deterministische Nutzenfunktionen:

$$\begin{aligned} V_1 &= \beta_0 + \beta_1 T_1, \\ V_2 &= \beta_2 T_2 + \beta_3 K_2. \end{aligned} \tag{3.1}$$

Die vier Parameter haben folgende Bedeutung:

- β_0 : Globale Bevorzugung der Alternative Fuß/Rad gegenüber der **Referenzalternative** ÖV in Nutzeinheiten (NE),
- β_1 : Zeitsensitivität in NE pro Minute Fuß- oder Radweg,
- β_2 : Zeitsensitivität in NE pro ÖV-Minute
- β_3 : monetäre Sensitivität in NE pro für den ÖV aufgebrauchten Euro.

Wir haben hier die Zeitsensitivität **alternativenspezifisch** formuliert, also unterschiedliche Sensitivitäten für Fuß/Rad und ÖV. Man könnte sie auch **generisch** mit einer Sensitivität für alle Verkehrsmodi beschreiben. Es gilt:

3 Datengewinnung: Verkehrserhebungen

- Wegen der Skaleninvarianz kann man eine NE ohne Einschränkung der Allgemeinheit als eine Schwankungsbreite (Standardabweichung) des Zufallsnutzens festlegen.
- Wegen der Translationsinvarianz bekommt die Alternative 2 als Referenzalternative *keinen* konstanten Teil.

Aus dem binären Logitmodell ergibt sich eine einfache Beziehung zwischen den Auswahlwahrscheinlichkeiten $P(k)$ für die Alternative k und der Ausstattung der Alternativen:

$$\ln\left(\frac{P(1)}{P(2)}\right) = V_1 - V_2 = \beta_0 + \beta_1 T_1 - \beta_2 T_2 - \beta_3 K_2 \quad (3.2)$$

Die Kalibrierung mit der Maximum-Likelihood-Methode (diese wird im Abschnitt 4.7 besprochen werden) ergibt folgendes Ergebnis (geschätzter Wert \pm Standardabweichung der Schätzung):

$$\begin{aligned} \beta_0 &= -1.37 \pm 0.47, & \beta_1 &= -0.228 \pm 0.022, \\ \beta_2 &= -0.195 \pm 0.014, & \beta_3 &= -1.75 \pm 0.12. \end{aligned}$$

Abbildung 3.10 zeigt die Güte des Modells, indem die tatsächlichen Umfrageergebnisse (Tabelle 3.1) mit dem Modell (3.2) verglichen werden.

Wie bei der Regression haben auch die im Wahlmodell geschätzten Parameter Unschärfen und Korrelationen. Abbildung 3.11 zeigt einige davon. Man sieht, dass die Sensitivitäten bezüglich Zeit und Kosten hochsignifikant von null verschieden sind, nicht jedoch der globale Bonus für die Alternative Fuß/Rad (welcher die Rolle des Achsenabschnitts der linearen Regression übernimmt).

Die gefundenen Werte der kalibrierten Modellparameter erlauben auch interessante Aussagen über die hinter den Entscheidungen stehenden Gewichtungen der verschiedenen Kriterien:

- Unschärfe des Zufallsnutzens in Fuß/Rad-Minuten: $-\frac{1}{\beta_1} = 4.4$
- Unschärfe des Zufallsnutzens in ÖPNV-Minuten: $-\frac{1}{\beta_2} = 5.1$
- Unschärfe des Zufallsnutzens in ÖPNV-€: $-\frac{1}{\beta_3} = 0.57$
- Bonus des ÖV in Einheiten der Standardabweichung des Zufallsnutzens: $-\beta_0 = 1.4$
- Bonus des ÖV in ÖV-Minuten: $-\frac{\beta_0}{\beta_2} = 7.0$
- Geldbewertung der ÖV-Kosten in €/ÖV-Stunden: $\frac{60 \beta_2}{\beta_3} = 6.7$

Die Ergebnisse direkter Nutzenmessungen bzw. die dazugehörigen Modellaussagen kann man auch als Kontourplots darstellen: Über einen Unterraum der exogenen Variablen (z.B. Reisezeiten, Kosten) werden Isolinien konstanter Auswahlwahrscheinlichkeiten

3 Datengewinnung: Verkehrserhebungen

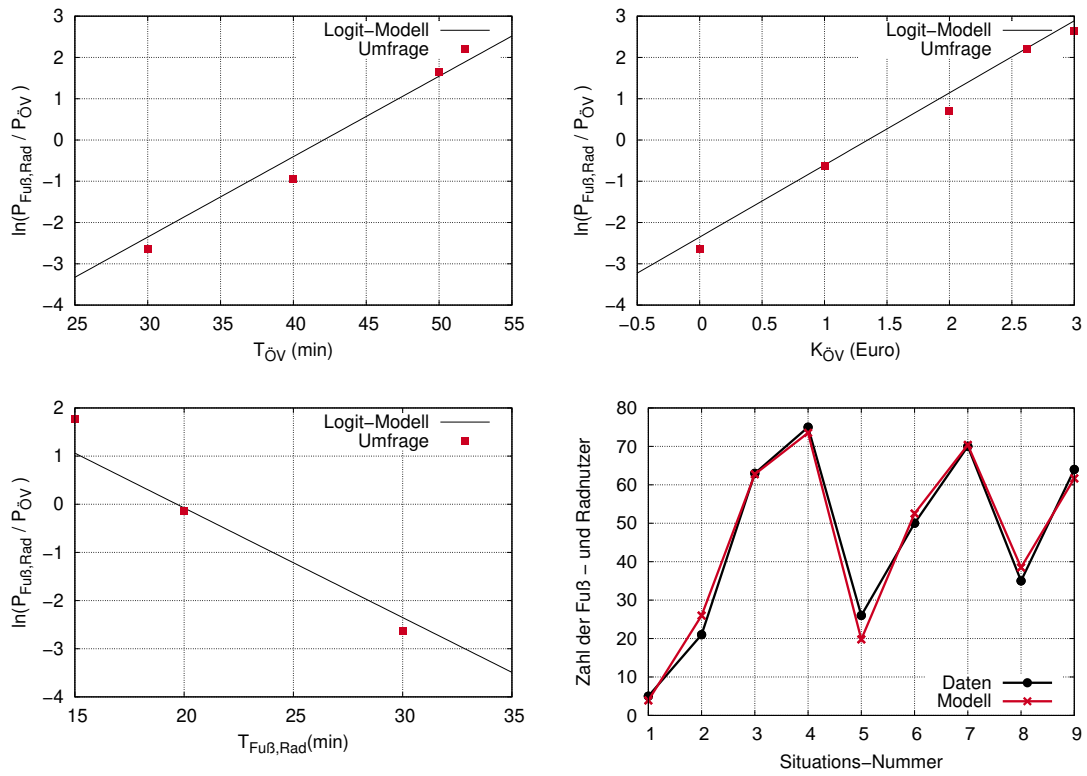


Abbildung 3.10: Ergebnisse der Modellierung des Stated-Choice-Experiments der Tabelle auf S. 129 mit dem (binärem) Logit-Modell. Gezeigt ist die Sensitivität bezüglich der komplexen Reisezeit $T_{\text{ÖV}}$ des ÖPNV (links) und der ÖPNV-Zusatzkosten pro Fahrt (rechts), sowie der Fuß- und Radwegzeit (links unten). Die jeweils nicht veränderten Referenzparameter sind $T_{\text{Fuß}} = T_{\text{ÖV}} = 30$ min und $K_{\text{ÖV}} = 0$ €. Rechts unten sind die tatsächlichen und modellierten absoluten Häufigkeiten der Entscheidung zugunsten Alternative 1 für die sieben Situationen dargestellt.

3 Datengewinnung: Verkehrserhebungen

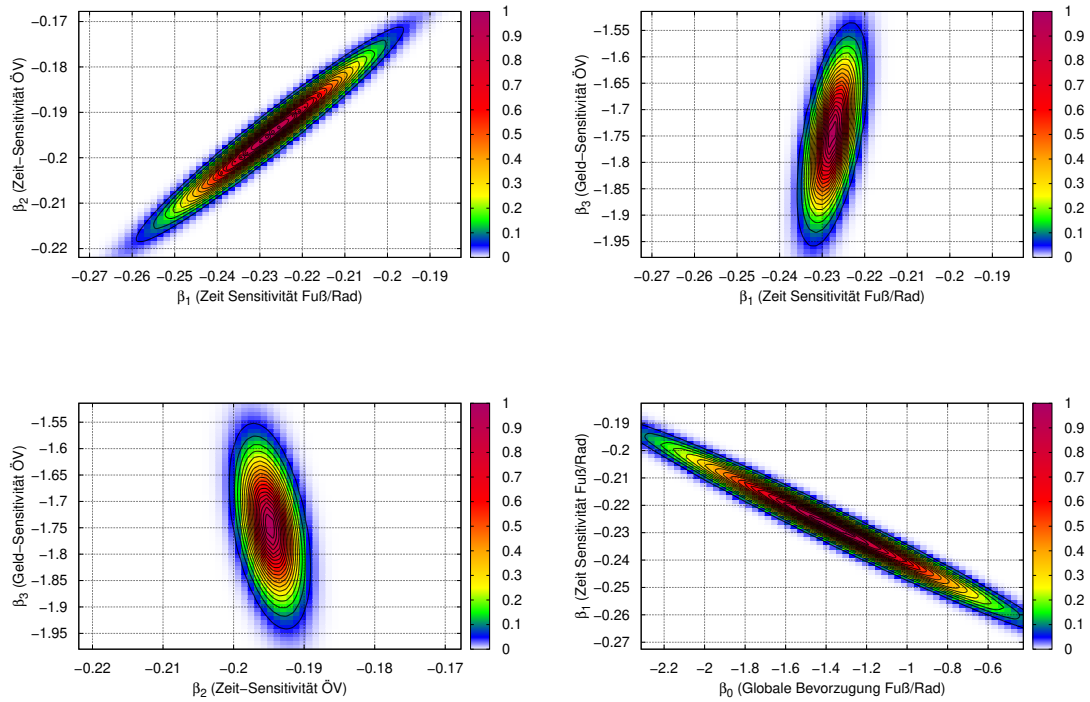


Abbildung 3.11: Likelihoodfunktionen von Parameterkombinationen nach der in Kap. 4.7 besprochenen Methode. Die Bilder zeigen die auf ein Maximum von 1 skalierte Wahrscheinlichkeitsdichte dafür, dass das aus den Nutzenfunktionen (3.1) formulierte Logit-Modell die Daten der Tabelle 3.1 beschreibt. Jeweils zwei Parameter variieren und die anderen beiden werden jeweils auf ihren geschätzten Wert gehalten.

3 Datengewinnung: Verkehrserhebungen

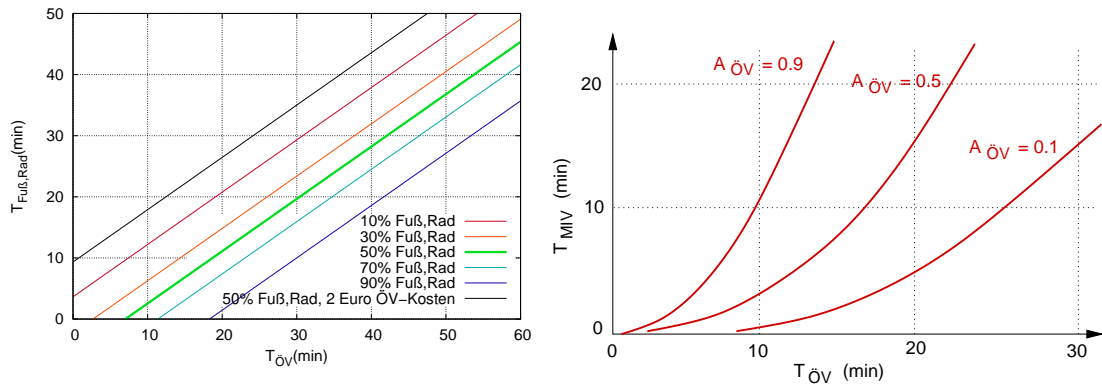


Abbildung 3.12: Kontourplot der Anteilswerte für die Alternative “ÖPNV” in Abhängigkeit der Widerstandsfunktionen (negative Nutzen) der zwei Alternativen. Links: Auf das Umfrageergebnis kalibrierte Logit-Modell; Rechts: So sieht es wohl realistischerweise eher aus.

dargestellt, vgl. Abb. 3.12. Hierbei sieht man die Grenzen des auf konstanter Schwankungsbreite des Zufallsnutzens beruhendem Logit-Modell. In Realität nimmt wohl die Unschärfe des Zufallsnutzens mit den Absolutwerten der Zeiten zu, wodurch die Isolinien qualitativ so wie in Abb. 3.12 rechts aussehen, also insbesondere nicht wie beim Logitmodell parallel zueinander sind (Abb. 3.12 links).

3.6.3 Komplexeres Beispiel aus der Literatur: Modal-Split in der Schweiz

Als Beispiel, was mit dieser Methode alles erfragbar ist, werden hier noch einige Ergebnisse einer Schweizerischen Untersuchung (Intl. Verkehrswesen, 9/2003) wiedergegeben:

Merkmal	Durchschnitt	geschäftlicher Verkehr
Zeitwert MIV	27.7 Sfr/h	74.6 Sfr/h
Zeitwert ÖPNV	18.5 Sfr/h	55.8 Sfr/h
1 × Umsteigen	7.6 Sfr	
Verlässlichkeit	0.1 Sfr/% Verspätungswahrsch.	0.7 Sfr/% Verspätungswahrsch.

Der Zeitwert des ÖPNV ist also, genau wie bei unserer “kleinen Untersuchung”, niedriger als der für IV. Die absoluten Werte sind aber in der Schweizer Untersuchung tendenziell höher, was wohl an den sozioökonomischen Variablen liegt: Schließlich ist die Schweiz teuer und die befragten Personen waren nicht nur Studenten. Diese Abhängigkeit wird auch durch den großen Unterschied der Bewertung von geschäftlichen und sonstigen Verkehr (Wegezweck!) deutlich. Bemerkenswert ist auch, dass einmal Umsteigen mit 7.5 Sfr bewertet wird, zusätzlich zum Zeitverlust, welcher bereits im Zeitwert enthalten ist. Schließlich fällt auf, dass beruflich Tätige die Zuverlässigkeit drastisch höher gewichten als der Durchschnitt.

3 Datengewinnung: Verkehrserhebungen

Aufgabe:

Wie wächst in dem durch (3.1) und (3.2) definierten Entscheidungsmodell die Entscheidungsunschärfe mit den Absolutwerten der Zeitaufwände an? Ist insbesondere die absolute Unschärfe ΔV oder die relative Unschärfe $\Delta V/V$ konstant? Ist das Ergebnis aus Abb. 3.10 und 3.12 konsistent mit diesem Modell?

3.6.4 Experimentelles Design der Choice Sets

Prinzipiell besteht eine wahlbasierte Conjoint-Untersuchung aus vier Schritten

1. Modellspezifikation, insbesondere welche Alternativen und welche exogenen Variablen berücksichtigt werden,
2. Experimentelles Design: Wieviele Choice Sets sollen es sein? Welchen Wertebereich haben die exogenen Variablen? Bekommt jeder Proband dieselbe Menge an Choice Sets? Welche Wertekombinationen werden den Choice Sets zugeordnet?
3. Erstellung der Fragekärtchen (jedes Choice Set auf ein eigenen Kärtchen, bei Pivotstudien ggf. individualisiert für die einzelnen Befragten) und Durchführung der Befragung
4. Auswertung

Das experimentelles Design zielt daraufhin, die Choice Sets so auszuwählen, dass die Analyse des diskreten Wahlmodells bei möglichst geringer Anzahl der Sets (mehr als 15 bis 20 kann man Probanden kaum zumuten) unverzerrte und möglichst scharfe Aussagen erlaubt. Dabei werden, um die Probanden nicht zu verwirren, auch kardinalskalierte Variablen meist nur in wenigen Ausprägungen vorgegeben. Für die Studentenbefragung des Abschnitts 3.6.2 (Wahl zwischen den zwei Alternativen Fuß/Rad und ÖV) könnte man beispielsweise folgende Stufen definieren:

- Komplexe Reisezeit T_1 der Alternative Fuß/Rad: 10, 15, 20, 25 und 30 Minuten,
- Komplexe ÖV-Reisezeit T_2 : 10, 15, 20, 25 und 30 Minuten,
- Ad-Hoc-Kosten K_2 des ÖV: 0.00, 1.50, 3.00 oder 4.50 Euro

Bei Pivot-Designs (Ausgangspunkt sind Revealed Choice Befragungen) könnte man auch feste Prozentstufen bezüglich der tatsächlichen Ausprägungen (z.B. Preis wie bisher, 20 % billiger oder teurer, 40 % teurer) wählen. Es gibt drei wesentliche Strategien zur Auswahl der Choice Sets:

Vollständige Enumeration (*full factorial design*): Es werden alle möglichen Kombinationen angewandt:

3 Datengewinnung: Verkehrserhebungen

i	Fuß-Rad Reisezeit T_1	ÖV-Reisezeit T_2	ÖV Ad-Hoc-Kosten K_2
1	10 min	10 min	0.00 €
2	10 min	10 min	1.50 €
⋮	⋮	⋮	⋮
99	30 min	30 min	3.00 €
100	30 min	30 min	4.50 €

Hier sind es also $n = 5 \times 5 \times 4 = 100$ Choice Sets.

Vorteile:

- Maximale Trennschärfe, alle Arten von Kreuz-Effekten werden so gut wie möglich berücksichtigt

Nachteile:

- Meist prohibitiv viele Choice Sets
- Viele überflüssige Fragen (triviale Sets), bei denen die Wahl offensichtlich ist wie im Set 20, bei dem man die Auswahl hat zwischen einen Fußweg von 10 Minuten oder den ÖV, wenn dieser 30 Minuten dauert sowie zusätzlich 4.50 € kostet.

Orthogonales Design: Die Merkmalskombinationen werden so gewählt, dass es keine deskriptiven Korrelationen gibt. Schreibt man die Tabelle der Choice Sets abstrakt als normalisierte Datenmatrix \mathbf{X} (die Zeilen sind die einzelnen Choice Sets und die Spalten sind $T_1 - \bar{T}_1$, $T_2 - \bar{T}_2$ und $K - \bar{K}$ (bzw. die auf das arithmetische Mittel 0 normalisierten Stufenindices von T_1 , T_2 und K), so enthält die Kovarianzmatrix $(\mathbf{X}'\mathbf{X})/n$ keine Nichtdiagonalelemente. Beispiel: Das Design

i	Fuß-Rad Reisezeit T_1	ÖV-Reisezeit T_2	ÖV Ad-Hoc-Kosten K_2
2	10 min	10 min	1.50 €
18	10 min	30 min	1.50 €
84	30 min	10 min	3.00 €
99	30 min	30 min	3.00 €

ist orthogonal bezüglich der Attributstufen innerhalb einer Alternative, aber nicht zwischen den Alternativen (die Fuß/Rad Zeit T_1 ist positiv mit den ÖV-Kosten K_2 korreliert).

Vorteile:

- Orthogonal

Nachteile:

- Meist immer noch zu viele Choice Sets, wenn man hinreichende Genauigkeit will

3 Datengewinnung: Verkehrserhebungen

- Weiterhin viele triviale Sets, zumindest, wenn man auch zwischen den Alternativen Orthogonalität vorschreibt (dann benötigt man im obigen Beispiel mindestens acht Choice Sets, u.A. Set 19, bei dem man die Auswahl hat zwischen einen Fußweg von 10 Minuten oder ÖV, wenn dieser 30 Minuten dauert sowie zusätzlich 3.00€ kostet.
- Orthogonalität, zumindest zwischen Alternativen, ist eigentlich sinnlos, da die *Differenzen* zwischen den Attributen entscheidend ist.

Effizientes Design: Dies ist die einzige Strategie, bei der man die Zahl n der Choice Sets explizit vorgeben kann. Unter allen Designs (also Mengen mit n Sets, deren Anzahl ist in obigem Beispiel bei $n = 10$ immerhin $\binom{10}{100} > 10^{12}$) ist das effiziente Design dasjenige, welches bei der Auswertung mit dem diskreten Wahlmodell die geringsten Unsicherheiten bei den Parameterschätzern ergibt. Dafür wird üblicherweise das Maß des **D-errors** (*determinantenbasierte Ungenauigkeit*)

$$D_{\text{err}} = \det \left[\left(\mathbf{V}_{\hat{\beta}} \right)^{1/J} \right] \quad (3.3)$$

herangezogen, also das geometrische Mittel der Hauptkomponenten der Varianz-Kovarianz-Matrix der J Parameterschätzer. Die Schätzer, $\hat{\beta}$, ebenso wie die Varianz-Kovarianz-Matrix $\mathbf{V}_{\hat{\beta}}$, werden mit dem *Maximum-Likelihood-Verfahren* gewonnen (welches erst in Abschnitt 4.9 näher betrachtet wird). Das impliziert aber auch, dass man zumindest eine ungefähre Vorstellung der eigentlich erst durch die Befragung zu ermittelnden Parameter haben muss.

Vorteile:

- Frei wählbare Anzahl der Choice Sets
- maximale Trennschärfe

Nachteile:

- Um den *D-error* zu berechnen, benötigt man eine vorläufige Parameterschätzung, welche man aus der Literatur (oder dem *Pre-Test* der Studie) entnehmen muss.
- Die Ermittlung des effizienten Designs ist numerisch aufwändig.

4 Modelle diskreter Entscheidungen

4.1 Einführung

Viele Sachverhalte der Verkehrsökonomie betreffen **diskrete Entscheidungen**, d.h. man muss sich für eine aus üblicherweise wenigen Alternativen entscheiden. Das wird deutlich, wenn man das bereits in Kap 1 besprochene klassische Vierstufenmodell der Verkehrsplanung unter dem Gesichtspunkt der anfallenden Entscheidungen betrachtet (Abb. 4.1):

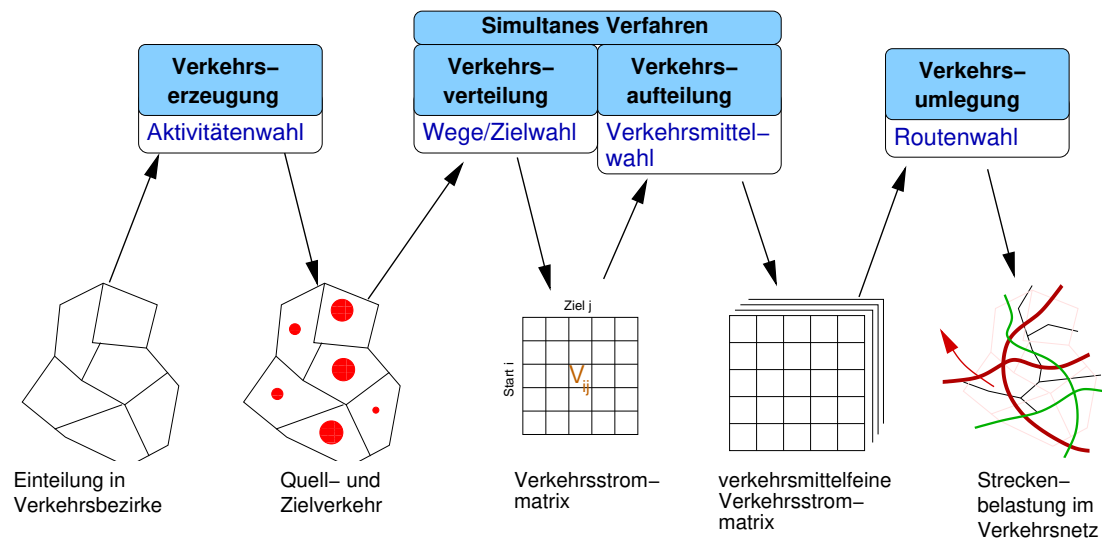


Abbildung 4.1: Das klassische Vierstufenmodell der Verkehrsplanung als Beispiel diskreter Entscheidungen (siehe den Haupttext).

- Zunächst wird in der **Aktivitätenwahl** (frühmorgens oder auch Tage davor) festgelegt, was denn das heutige "Tagewerk" sein soll: Dies betrifft insbesondere wahlfreie Aktivitäten wie Einkaufen gehen, zum Sport gehen usw.
- In der darauffolgenden **Zielwahl** wird entschieden, *wo* die gewählten Aktivitäten stattfinden sollen (z.B. in welchen Läden eingekauft werden soll)
- Danach oder gleichzeitig wird in der **Verkehrsmittelwahl** entschieden, *wie* die Ziele der jeweiligen Aktivitäten erreicht werden sollen: Zu Fuß, mit dem Rad, mit Bus oder Bahn (ÖV), Auto oder Motorrad (MIV), mit dem Flugzeug usw.

4 Modelle diskreter Entscheidungen

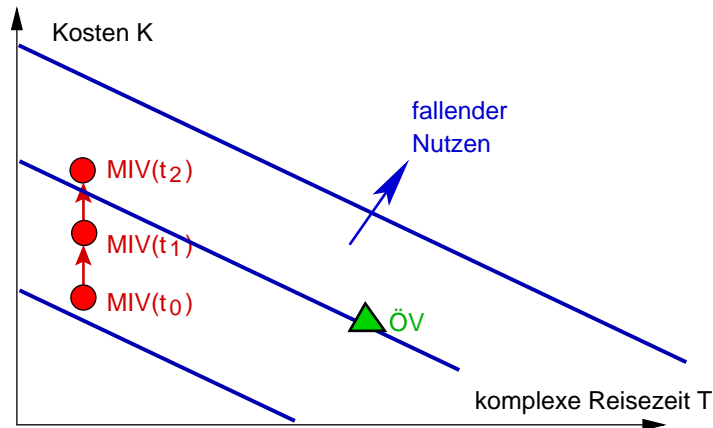


Abbildung 4.2: Indifferenzkurven einer linearen, von zwei exogenen Variablen abhängigen Nutzenfunktion. Die Symbole bezeichnen die Lage bestimmter Alternativen im Raum der exogenen Variablen.

- Schließlich wird in der **Routenwahl** der günstigste Weg für das gewählte Ziel und das gewählte Verkehrsmittel im Streckennetzwerk ausgesucht.

Grundsätzlich sind diskrete Entscheidungen nichtlinear, was nun anhand Abb. 4.2 verdeutlicht wird. In der dargestellten Situation gibt es als maßgebliche exogene Variable (Beeinflussungsfaktoren) die komplexe Reisezeit T und die Kosten K . Diese können jeweils für die beiden in Frage kommenden Alternativen “ÖV” und “MIV” berechnet werden und dienen als Kriterien für die Wahlentscheidung (endogene Variable). Im einfachsten Fall bildet man aus T und K für jede Alternative eine lineare Nutzenfunktion,

$$V_i = \beta_1 T_i + \beta_2 K_i, \quad (4.1)$$

berechnet diese für beide Alternativen und wählt gemäß dem klassischen ökonomischen Modell der Nutzenmaximierung (*Homo Oeconomicus*):

$$i_{\text{gewählt}} = \arg(\max_i (V_i)) \quad (4.2)$$

die Alternative mit dem größeren Nutzen.¹ In der den Raum der exogenen Variablen darstellenden Abbildung 4.2 werden beide Alternativen durch je einen Punkt repräsentiert. Ferner sind **Indifferenzkurven** gleichen Nutzens eingezeichnet, die mit (4.1) natürlich zu Geraden degenerieren. Gewählt wird nun die Alternative, welche sich bezüglich der Indifferenzkurven weiter “links unten” befindet, da in dieser Richtung der Nutzen zunimmt bzw. die generalisierten Kosten abnehmen.

Sei Anfangs ($t = t_0$) die Situation so, dass der MIV bevorzugt wird. Nun steigen die Kraftstoffkosten, so dass für den MIV die Kosten steigen, während alles andere (die

¹Die Argumentfunktion $\arg(\cdot)$ in Gl. (4.2) bezeichnet den zum Maximum bzw. Minimum in der Klammer gehörigen Wert des zu maximierenden (minimierenden) Arguments.

4 Modelle diskreter Entscheidungen

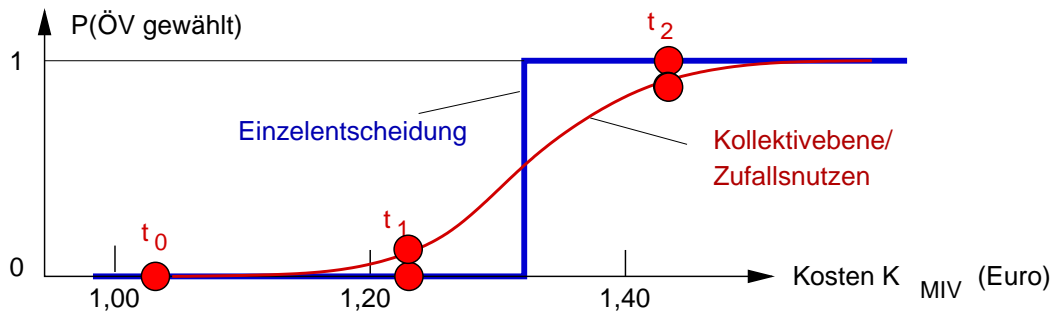


Abbildung 4.3: Die zu der in Abb. 4.2 dargestellten Nutzenfunktion gehörige Wahlentscheidung für den ÖPNV in Abhängigkeit der (Treibstoff-)kosten K_{MIV} für den MIV unter *ceteris paribus* Bedingungen. Auf Individualebene lautet die Entscheidung nur ja (1) oder Nein (0); Auf Kollektivebene bzw. bei Zufallselementen in der Nutzenfunktion ergibt sich eine Wahrscheinlichkeit zwischen 0 und 1. Die drei Zeitpunkte bzw. die dazugehörigen Werte von K_{MIV} entsprechen denen in Abb. 4.2.

Reisezeiten und die Kosten für den ÖV) unverändert bleibt (*ceteris paribus*). Zum Zeitpunkt t_1 ist aber immer noch $K_{\text{MIV}} > K_{\text{ÖV}}$, so dass nach wie vor der MIV gewählt wird. Trotz Änderung der exogenen Variablen K_{MIV} ändert sich also nichts bei der endogenen Variablen der Wahlentscheidung. Erst wenn zum Zeitpunkt t_2 die Kosten K_{MIV} noch einmal um den gleichen Betrag erhöht werden, schlägt die Entscheidung zugunsten des ÖPNV um.

Bei ein und derselben Änderung der exogenen Variablen tut sich also einmal nichts, während es im anderen Fall eine drastische Änderung gibt. Dieses typische Merkmal einer Nichtlinearität bleibt auch bei mehr als zwei Alternativen erhalten. Betrachtet man statt der *Individualebene* die *Kollektivebene*, also unterschiedliche Präferenzen der einzelnen Entscheider, so weicht die "Sprungfunktion" der Alternativenwahl zwar auf, bleibt aber im Wesentlichen nichtlinear (Abb. 4.3). Die im vorhergehenden Kapitel entwickelten mächtigen Werkzeuge der multivariaten linearen Modelle sind also nicht einsetzbar, wenn man nicht den speziellen Mechanismus der **logistische Regression** nutzt, der aber nicht gut verallgemeinerbar ist.

Auch bei linearen Nutzenfunktionen sind die Modelle der diskreten Wahltheorie prinzipiell nichtlinear. Dies gilt auch auf der Kollektivebene.

4.2 Mathematische Beschreibung

Diskrete Wahlentscheidungen werden heutzutage (und auch in diesem Skript) im Allgemeinen durch Maximierung einer aus deterministischen und Zufallsanteilen bestehen-

4 Modelle diskreter Entscheidungen

den Nutzenfunktion (*Homo Oeconomicus*) hergeleitet. Es geht jedoch auch anders: Bereits 1927² wurde ein bestimmtes Modell, das *binomiale Probitmodell* (Abschnitt 4.3.1), anhand der Modellierung psychologischer Stimuli hergeleitet. Erst 1960 wurden diese Stimuli als Ergebnis der Maximierung einer Nutzenfunktion uminterpretiert.³ In systematischer Weise wurde die **diskrete Wahltheorie** (engl. *discrete-choice analysis*), insbesondere das heute am gebräuchlichste *Logit-Modell*, in den Siebziger und Achziger Jahren von Daniel McFadden⁴ (Nobelpreis 2000), Ben-Akiva⁵ und anderen entwickelt.

In jedem Fall werden die Modelle der diskreten Wahltheorie grundsätzlich “mikroskopisch” formuliert, also Einzelentscheidungen individueller Personen i betrachtet. Man kann auch bezüglich der Entscheidung verhaltenshomogene Gruppen zusammenfassen. Ein Beispiel einer solchen Gruppe im Falle der Verkehrsmittelwahl sind alle weiblichen Erwerbstätigen in der Grundgesamtheit bzw. Stichprobe, welche zwischen 30 und 40 Jahren alt sind, kein Kfz aber ein Rad zur Verfügung haben, und bei denen die Entfernung Wohnung-Arbeitsstätte zwischen 5 und 10 km beträgt. Der Personenindex wird dann zum Personengruppenindex, aber die Entscheidung wird nach wie vor individuell (d.h. Person für Person) modelliert.

Modelle der diskreten Wahltheorie haben zwei Hauptkomponenten (Abb. 4.4):

- Für jede Person und jede Alternative gibt es eine deterministische Nutzenfunktion (Abschnitt 4.2.3). Diese hängt von den in Abschnitt 4.2.2 besprochenen exogenen Variablen ab und kann linear, quasilinear oder nichtlinear sein.
- Ein Wahlmodell beschreibt die vom deterministischen Nutzen und einem modellabhängigen Zufallsnutzen abhängige Wahlentscheidung bzw. die Auswahlwahrscheinlichkeiten der möglichen Alternativen. (Abschnitt 4.2.5).

Da grundsätzlich rationale Entscheidungen angenommen werden (*homo oeconomicus*, der Gesamtnutzen wird maximiert), ist das Wahlmodell äquivalent zur Definition des Verteilung des Zufallsnutzens. Wichtig sind normalverteilte Zufallsnutzen, welche zu den **Probit-Modellen** führen (Abschnitt 4.3) und extremwertverteilte (genauer: gumbelverteilte) Zufallsnutzen, welche den in Abschnitt 4.4 behandelten **Logit-Modellen** entsprechen. Beide Ansätze haben Vor- und Nachteile: Probit-Modelle erlauben allgemeine Korrelationen und unterschiedliche Varianzen (**Heteroskedastizität**) und sind statistisch begründet (Zentraler Grenzwertsatz), aber ihre Auswahlwahrscheinlichkeiten sind i.A. nur durch Simulation bestimmbar. Hingegen ist die Logit-Auswahlwahrscheinlichkeit eine einfache Formel, aber es müssen identisch-unabhängige (Gumbel-)Verteilungen der Zufallsnutzen angenommen werden.

²L.L. Thurstone, *The American Journal of Psychology* **38**, 3, pp. 368 (1927).

³Marschak, J. *Binary Choice Constraints and Random Utility Indicators*, Stanford Symposium on Mathematical Methods in Social Science, Edited by K. Arrow. Stanford, CA: University of Stanford 1960.

⁴D. McFadden, *Conditional logit analysis of qualitative choice behavior*, Institute of Urban and Regional Development, University of California (1973).

⁵Moshe Ben-Akiva and Steven Lerman, *Discrete Choice Analysis Theory and Application to Travel Demand*. MIT Press (1985).

4 Modelle diskreter Entscheidungen

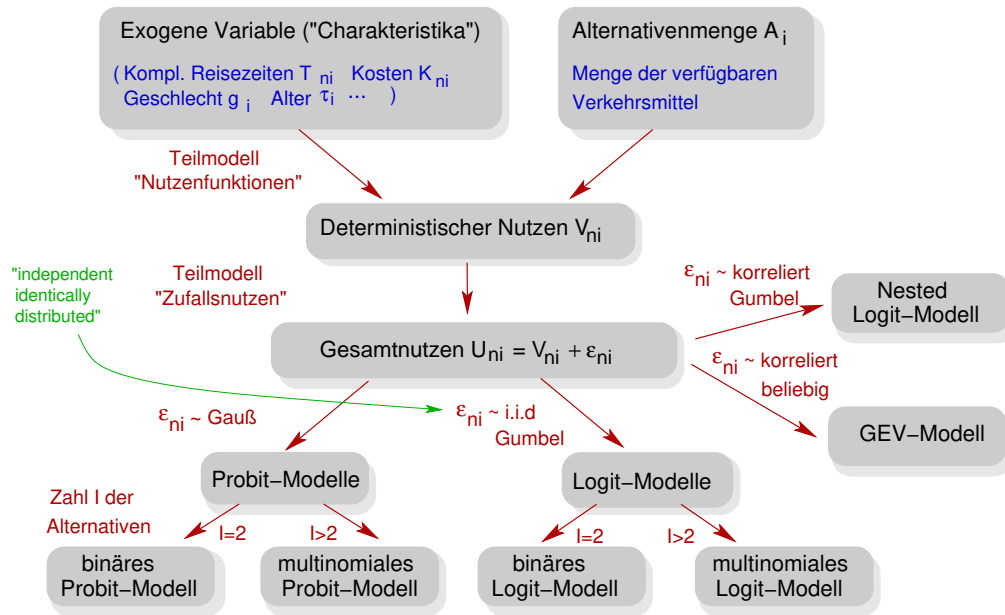


Abbildung 4.4: Die wichtigsten Komponenten von *discrete-choice* Modellen. In blau stehen Beispiele für die Verkehrsmittelwahlentscheidung.

Oftmals ist aber die Annahme unabhängiger und/oder identisch verteilter (i.i.d.) Zufallsnutzen aus dem Sachverhalt heraus nicht zu halten, z.B. bei der kombinierten Verkehrsmittel- und Routenwahl. Will man auf die analytische Berechnung der Auswahlwahrscheinlichkeit nicht verzichten, kommen **Generalized Extreme Value (GEV)** Modelle zum Einsatz (Abschnitt 4.13). Diese erlauben – mit Einschränkungen – korrelierte und heteroskedastische Zufallsnutzen, die einer verallgemeinerten Extremwertverteilung genügen.⁶ Ein besonders wichtiger und anschaulicher Spezialfall der GEV-Modelle sind die **Nested-Logit Modelle** (Abschnitt 4.13.3).

Zunächst gilt es aber, die mathematische Beschreibung der Alternativen und der exogenen Variablen in eine Form zu bringen, auf die Wahlmodelle angewandt werden. Dies wird zunächst betrachtet.

4.2.1 Alternativen

Jede Person bzw. Personengruppe n hat bei der Wahlentscheidung im Allgemeinen unterschiedliche **Alternativenmengen** $A_n = \{a_{ni}\}$ zur Verfügung.⁷ Legt man fest, dass

⁶Die Einschränkung auf eine bestimmte Form der Verteilung, z.B. Gumbel statt Gauß, ist selten relevant, da die Dichtefunktionen ähnlich aussehen. Wichtiger sind die Korrelationen und die Heteroskedastizität.

⁷In der Literatur zur diskreten Wahltheorie werden die einzelnen Datensätze durch die Person mit dem Index n (statt i wie bei der Regressionsrechnung) bezeichnet, während i nun den *Alternativenindex* bezeichnet. Zur leichteren Vergleichbarkeit folgen wir der Konvention (eine Übersicht der Bezeich-

4 Modelle diskreter Entscheidungen

ein bestimmter Index i eine bestimmte feste Alternative bedeutet, entspricht die Menge A_n also einer Menge von natürlichen Zahlen $\{i\}$ im Bereich von 1 bis I . Um die “Buchhaltung” nicht durch unterschiedliche Alternativenmengen zu verkomplizieren, werden für nicht verfügbare Alternativen sehr hohe “virtuelle Strafkosten” bei den deterministischen Nutzenfunktionen veranschlagt, so dass die Wahrscheinlichkeit, solche Alternativen zu wählen, *de facto* gleich Null ist.⁸ Die Alternativen müssen folgenden Bedingungen gehorchen:

1. *exklusiv* bzw. *nicht-häufbar*: Es kann höchstens eine gewählt werden.
2. *vollständig*: Es muss mindestens eine gewählt werden. Berücksichtigt man Punkt 1, muss also genau eine gewählt werden,
3. sie müssen *wesentlich voneinander verschieden* sein,
4. und es darf nur *endlich viele* Alternativen geben.

Beispiel Verkehrsmittelwahl:

Punkt 1 impliziert, dass man nicht fragen sollte “welches Verkehrsmittel verwendeten Sie für Ihren Weg hierher” (mit der möglichen Antwort “kombiniert Rad, ÖV und zu Fuß”, wenn man mit dem Rad zur Haltestelle fährt und an der Zielhaltestelle zu Fuß zum Ziel läuft), sondern beispielsweise “welches Hauptverkehrsmittel verwendeten Sie”.

Punkt 2 impliziert, dass man, speziell bei Stated-Choice Analysen, ggf eine “Nicht-Alternative” (“unter diesen Bedingungen bleibe ich zu Hause”) anbietet, und bei Revealed-Choice Fragen eine Alternative “sonstige Verkehrsmittel” vorsieht, die Motorradfahrer, Hubschrauberflieger, Inline-Skater etc. berücksichtigt.

Punkt 3 ist erforderlich, um Verzerrungen zu vermeiden, speziell bei Stated Choice: Dies wird am deutlichsten am Beispiel des bekannten **Red-Bus Blue-Bus Problems**: Lauten die vier Alternativen bei einem zu Fuß unpraktikablem Weg “Rad”, “Roter Bus”, “Blauer Bus”, “Kfz”, so sind (außer bei Leuten, die rot oder blau hassen) die deterministischen Nutzenfunktionen des roten und blauen Busses gleich. Damit werden vom Modell Busse doppelt gezählt und die Wahrscheinlichkeit, Busse zu wählen, überschätzt (oder, wenn man dieses Modell kalibriert, die Attraktivität von roten oder blauen Bussen bzw. deren Nutzenfunktionen gegenüber sonstigen Verkehrsmitteln unterschätzt). Sind insbesondere alle vier deterministischen Nutzenfunktionen gleich, so würde der ÖV, also einer der Busse, mit einer Wahrscheinlichkeit von $1/2$ gewählt, obwohl eigentlich $1/3$ plausibel wäre.

In der Regel sind nur *Verkehrssysteme* (zu Fuß, Rad, ÖV, MIV) wesentlich verschieden, in speziellen Untersuchungen können aber durchaus Bus und Bahn als eigene Alternativen gelten.

nungen findet sich in Abschnitt 4.17).

⁸Im Logit-Modell führt ein “Malus” von 1000 gegenüber einer anderen Alternative zu einer Auswahlwahrscheinlichkeit der Größenordnung e^{-10^3} (nach dem “Null Komma” stehen etwa 400 Nullen, ehe eine signifikante Ziffer kommt.)

4 Modelle diskreter Entscheidungen

Punkt 4 schließt kontinuierliche Wahlentscheidungen wie die Wahl der Wunschgeschwindigkeit auf freier Strecke aus. Einerseits werden kontinuierliche Wahlentscheidungen aber durch eine metrische endogene Variable beschrieben und sind der Regressionsrechnung zugänglich. Andererseits kann man die Menge der möglichen Wunschgeschwindigkeiten in einen vollständigen Satz einander ausschließender Teilmengen (< 80 km/h, $80-100$ km/h usw.) aufspalten und jede Teilmenge einer Alternative zuordnen, so dass alle vier Punkte erfüllt sind.

Verständnisfrage:

Welche Kriterien sind verletzt, wenn man die Wunschgeschwindigkeiten in fünf Teilmengen < 80 km/h, $80-100$ km/h, $100-120$ km/h, $120-121$ km/h, $121-140$ km/h aufspaltet? Begründung?

4.2.2 Beeinflussungsfaktoren

Wie in allen ökonometrischen Modellen werden Beeinflussungsfaktoren als exogene bzw. Input-Variable modelliert. Man unterscheidet dabei zwei Kategorien:

- Faktoren, welche im wesentlichen Eigenschaften der Alternativen darstellen, werden durch einen Satz von **generischen Variablen**, sog. **Charakteristika** $C_{ni} = \{C_{jni}\}, j = 1, \dots, J_C$ modelliert. Dieser Satz von Faktoren (repräsentiert durch einen Vektor und dargestellt in Fettdruck) hängt naturgemäß von der Alternative i , aber häufig auch von der Personen(gruppe) n ab (z.B. bei der Reisezeit als typischen Vertreter dieser Art von Variablen). Es gibt allerdings auch generische Variable (z. B. Preise von Produkten bei einer Produktwahlentscheidung), welche im wesentlichen für alle Personen denselben Wert aufweisen.⁹

Beispiel Verkehrsmittelwahl ($i = 1$: zu Fuß, 2 : Rad, 3 : ÖV, 4 : MIV):

C_{1ni}	= T_{ni}	Komplexe Reisezeit (in Minuten)
C_{2ni}	= K_{ni}	Kosten (in €)
C_{3ni}	= $\sqrt{V(T_{ni})}$	Zuverlässigkeit (Standardabweichung von T in Minuten, siehe Abschnitt 4.15)
C'_{3ni}	= $T_{0.95}$	Alternativ: 95. Perzentil der Reisezeit
C_{4ni}	= B_{ni}	Quantifizierte Bequemlichkeit ...

Beispielsweise würde $C_{123} = T_{23}$ die komplexe (Haustür-zu-Haustür-) Reisezeit (Kriterium bzw. Faktor $j = 1$) für Person(nengruppe) $n = 2$ bedeuten, wenn sie mit dem ÖV (Alternative $i = 3$) führe.

⁹Wenn man von den immer beliebteren Kundenkarten absieht.

4 Modelle diskreter Entscheidungen

- Rein personenbezogene Faktoren werden durch einen Satz von **sozioökonomischen Variablen** $\mathbf{S}_n = \{S_{jn}\}, j = 1, \dots, J_S$ modelliert. Als personenspezifische Merkmale hängen die S_n *nur* von der Person, nicht aber von der Alternative ab.

Beispiel Verkehrsmittelwahl:

$$\begin{aligned} S_{n1} &= \tau_n && \text{Alter (in Jahren oder als Altersklassenindex)} \\ S_{n2} &= g_n && \text{Geschlecht (z.B. } 0=\text{♂}, 1=\text{♀)} \\ S_{n3} &= e_n && \text{Einkommen (z.B. in €/Monat oder als Klassenindex)} \\ S_{n4} &= b_n && \text{Autobesitz (0=nein, 1=ja) usw.} \end{aligned}$$

Hier sieht man, dass durchaus auch qualitative (nominalskalierte) Variablen wie das Geschlecht oder der Autobesitz als Einflussfaktoren vorkommen können. Dabei gilt dasselbe Prinzip wie bei den Regressionsmodellen (vgl. Abschnitt 2.7):

Um einen nichtmetrischen (nominal- oder ordinalskalierten) nichthäufbaren Einflussfaktor mit k möglichen Ausprägungen in ein Modell diskreter Entscheidungen zu bringen, wird dieser Einflussfaktor durch $k - 1$ dichotome exogene **Dummyvariable** (Pseudovariablen) in der deterministischen Nutzenfunktion modelliert.

- **Externe Variable** wie das Wetter hängen weder von der Person noch von der Alternative ab. Dennoch können auch sie relevant sein, wenn sie unterschiedliche Alternativen unterschiedlich beeinflussen.
- Schließlich gibt es, bewirkt letztendlich durch nichtberücksichtigte Einflüsse, noch **konstante Präferenzen** für verschiedene Alternativen.

4.2.3 Deterministische Nutzenfunktion

Prinzipiell kann man wie bei der Regression zwischen linearen, parameterlinearen (quasilinearen) oder nichtlinearen Formulierungen unterscheiden. Da sich eine parameterlineare Beschreibung genau wie eine lineare handhaben lässt und viel mehr Flexibilität und Aussagekraft erlaubt, werden rein lineare Nutzenfunktionen nicht weiter betrachtet. Andererseits verkomplizieren auch in den Parametern nichtlineare Nutzenfunktionen die Handhabung (insbesondere die Parameterschätzung) erheblich und werden nur herangezogen, wenn dies unvermeidlich ist, wie bei der allgemeinen Modellierung von Schwellen (Abschnitt 4.10).

Der quasilineare Ansatz der deterministischen Nutzenfunktionen wird *de facto* zum linearen Modell, wenn man die Charakteristika und sozioökonomischen Variablen über parameterfreie, i.A. nichtlineare Funktionen zu **exogenen Einflussfaktoren** x_{nmi} zusammenfasst:

$$x_{nmi} = g_{mi}(\mathbf{C}_{n1}, \mathbf{C}_{n2}, \dots, \mathbf{C}_{nI}, \mathbf{S}_n), \quad (4.3)$$

4 Modelle diskreter Entscheidungen

Die Funktionen $g_{mi}(\cdot)$ definieren die Art, wie die Charakteristika, sozioökonomischen Variablen und externe Faktoren in die Einflussfaktoren oder *Kriterien* eingehen.¹⁰ Zusammen mit der noch zu diskutierenden Spezifikation des Zufallsnutzens definieren sie das *discrete-choice* Modell.

Bezüglich der Faktoren x_{nmi} , die die Rolle von transformierten exogenen Variablen spielen, sind die deterministischen Nutzen linear:

$$V_{ni} = \sum_{m=1}^M \beta_m x_{nmi}, \quad (4.4)$$

oder, in Vektornotation:

$$V_{ni} = \beta' \mathbf{x}_{ni}, \quad (4.5)$$

Hierbei beschreibt $\beta_m x_{nmi}$ die durch den Faktor m verursachte *subjektive Nutzenerhöhung*, wie er von Person n für Alternative i wahrgenommen wird. Die exogenen Faktoren x_{nmi} haben i.A. inkommensurable Einheiten (z.B. Zeiteinheiten für die Reisezeit und Geldeinheiten für die Kosten), die durch die Vorfaktoren, also die Modellparameter, kommensurabel gemacht werden. Damit haben Quotienten der Parameter oft eine anschauliche Bedeutung und geben beispielsweise einen *Zeitwert* in Euro pro Minute an.

Für die Formulierung der deterministischen Nutzen gibt es in der diskreten Wahltheorie einige Besonderheiten, die daher rühren, dass nur *Nutzendifferenzen* zwischen Alternativen relevant sind. Dies wird anhand der Verkehrsmittelwahl mit den Verkehrsmodi $i = 1$: zu Fuß, 2: Rad, 3: motorisiert bzw. sonstige veranschaulicht.

- Charakteristika, welche von der Alternative (und ggf. auch von der Person) abhängen, können im einfachsten Fall **generisch**, also direkt modelliert werden. Eine Nutzenänderung durch komplexen Reisezeiten T_{ni} und Ad-Hoc Kosten K_{ni} für Person n bei Wahl der Alternative i wird dann durch

$$\Delta V_{ni} = \beta_1 T_{ni} + \beta_2 K_{ni} \quad (4.6)$$

beschrieben. Üblicherweise sollten die Zeit- und Geldsensitivitäten β_1 bzw. β_2 negativ sein. Der Quotient β_1/β_2 gibt den **impliziten Zeitwert** (engl. **VTTTS** für **value of travel time savings** in den bei T_{ni} und K_{ni} verwendeten Einheiten (z.B. Euro pro Minute) an.

- Sozioökonomische Variable sind per Definition für alle Alternativen dieselben und können deshalb nicht direkt in die Nutzenfunktionen eingehen: Es kommt ja ausschließlich auf *Nutzendifferenzen* an. Sie müssen deshalb durch **alternativenspezifische Variable** bzw. Faktoren formuliert werden, indem man sie an jeweils eine Alternative über **Selektoren** koppelt. Selektoren sind *Dummyvariable* (Dummy=Platzhalter), welche für eine Alternative den Wert 1 und für alle anderen den

¹⁰Da die Entscheidungskriterien mehrere exogene Variable zusammenfassen können, aber auch eine Variable zu mehreren Kriterien führen kann (z.B. ein linearer und quadratischer Anteil in der Reisezeit), ist die Zahl M der Kriterien, also der exogenen Variablen, i.A. nicht gleich der Summe $J_C + J_S$ der Beeinflussungsfaktoren. Sie bekommt deshalb einen eigenen Summationsindex m .

4 Modelle diskreter Entscheidungen

Wert null annehmen. Will man beispielsweise die unterschiedliche Präferenzen der Geschlechter g_n ($0=\sigma$, $1=\phi$) modellieren, ergibt das die Nutzenanteile

$$\Delta V_{ni} = \beta_3 g_n \delta_{i1} + \beta_4 g_n \delta_{i2} \quad (4.7)$$

mit dem Selektor

$$\delta_{ii'} = \begin{cases} 1 & i = i' \\ 0 & i \neq i'. \end{cases} \quad (4.8)$$

Es gilt:

- Bei I Alternativen gibt es bei vollständiger Beschreibung eines zewertigen soziodemographischen Einflusses $I - 1$ exogene Faktoren und dazugehörige Parameter.
 - Die Alternative, welche nicht durch Selektoren ausgewählt wird, ist die **Referenzalternative**. Die Parameter geben die Nutzenänderung durch die soziodemographische Variable *bezüglich dieser Referenz* an. Im obigen Beispiel gibt β_3 an, um wieviele Einheiten sich der Nutzenunterschied des Modus “zu Fuß” gegenüber “motorisiert” bei Frauen gegenüber Männern erhöht.¹¹
- Externe Faktoren werden formal wie sozioökonomische Variable, also alternativen spezifisch formuliert. Schönes ($w = 1$) bzw. schlechtes ($w = 0$) Wetter führt beispielsweise zu den Nutzenänderungen

$$\Delta V_{ni} = w(\beta_5 \delta_{i1} + \beta_6 \delta_{i2}). \quad (4.9)$$

Der Parameter β_5 gibt an, um wieviel sich bei schönem gegenüber schlechtem Wetter der Nutzen von Alternative 1 gegenüber der Referenz 3 erhöht. β_6 bewirkt Selbiges für Alternative 2. Für eine vollständige Beschreibung eines K -wertigen externen Effekts benötigt man ebenfalls $(K - 1)(I - 1)$ exogene Faktoren und zugehörige Parameter.

- Globale Präferenzen für einzelne Alternativen quer durch die gesamte Stichprobe und unabhängig von den einzelnen Eigenschaften der Alternativen, also das Analoge zu den Konstanten der Regressionsrechnung, werden durch **alternativenspezifische Konstanten (ACs)** modelliert. Im obigen Beispiel lautet der vollständige Satz von $I - 1$ alternativenspezifischer Konstanten:

$$\Delta V_{ni} = \beta_7 \delta_{i1} + \beta_8 \delta_{i2} \quad (4.10)$$

Der Parameter β_7 gibt eine globale Bevorzugung der Alternative 1 gegenüber der Referenzalternative 3 und $\beta_7 - \beta_8$ eine Bevorzugung der Alternative 1 gegenüber 2 an.

¹¹Man beachte, dass man zur vollständigen Beschreibung des Einflusses einer nichtmetrischen nichthäufbaren K -wertigen sozioökonomischen Variable bei der Wahlentscheidung zwischen I Alternativen $(K - 1)(I - 1)$ exogene Faktoren und zugehörige Parameter benötigt.

4 Modelle diskreter Entscheidungen

- Generische Variablen kann man auch alternativenspezifisch formulieren: Beim Nutzenanteil

$$\Delta V_{ni} = \beta_1 T_{ni} + \beta_9 T_{ni} \delta_{i1} + \beta_{10} T_{ni} \delta_{i2} \quad (4.11)$$

beschreibt β_1 die bisherige generische Zeitsensitivität, während β_9 den Unterschied der Zeitsensitivität bei Alternative 1 gegenüber 3 und β_{10} den Unterschied bei Alternative 2 gegenüber 3 beschreibt. Damit beschreibt β_1 die Zeitsensitivität der Alternative 3. Alternativ kann man die Faktoren 1, 9 und 10 auch schreiben:

$$\Delta V_{ni} = \beta'_9 T_{ni} \delta_{i1} + \beta'_{10} T_{ni} \delta_{i2} + \beta'_1 T_{ni} \delta_{i3} \quad (4.12)$$

Hier geben $\beta'_9 = \beta_1 + \beta_9$, und $\beta'_{10} = \beta_1 + \beta_{10}$ und $\beta'_1 = \beta_1$ die gesamte Zeitsensitivitäten für die Alternative 1, 2 bzw. 3 an.¹² Die letztere Beschreibung ist aber unvorteilhafter gegenüber der durch β_1 , β_9 und β_{10} gegebenen Formulierung, da die Schätzer der “gestrichenen” β -Werte fast immer eine sehr hohe Korrelation aufweisen: Schließlich enthalten sie alle die gemeinsame (und betragsmäßig fast immer bedeutendste) Größe β_1 . Dies verschlechtert die Trennschärfe und Aussagekraft des Modells (wie in Abschnitt 4.9 näher beschrieben werden wird).

- Verschiedenste Nichtlinearitäten in den exogenen Variablen wie Kopplungen von generischen, sozioökonomischen und/oder externen Variablen sind möglich und können sinnvoll sein:
 - ein Anteil $\Delta V = \beta_{11} T_{ni} / [1 + (T_{ni}/5 \text{ min})^2]$ beschreibt eine betragsmäßig erhöhte (negativere) Zeitsensitivität bei kurzen Wegen ($\beta_{11} < 0$) oder eine weniger negative Sensitivität bis hin zur Indifferenz ($0 < \beta_{11} \leq -\beta_1$),¹³
 - ein Anteil $\Delta V = \beta_{12} T_{ni} E_n$ (Kopplung der Zeit mit der sozioökonomischen Variable “Einkommen” beschreibt eine Änderung der Zeitsensitivität mit den Einkommen. Die gesamte Zeitsensitivität ist nun $\frac{\partial V_{ni}}{\partial T_{ni}} = \beta_1 + \beta_{12} E_n$. (Typischerweise $\beta_{12} < 0$, d.h. bei hohem Einkommen ist die Zeitsensitivität ausgeprägter, also negativer, als bei Geringverdienern).

Zusammenfassend gilt Folgendes:

¹²Man beachte, dass es bei in der Art (4.12) alternativenspezifisch formulierten generischen Variablen, im Gegensatz zu sozioökonomischen Variablen oder Konstanten, keine Referenzalternative gibt und man deshalb reihum selektieren muss.

¹³Will man auch die hier auf 5 min festgelegte Schwellenbreite variabel lassen, kommt man um parameter-nichtlineare Nutzenfunktionen nicht herum, siehe Abschnitt 4.10.

4 Modelle diskreter Entscheidungen

- Von Alternative zu Alternative unterschiedliche Charakteristika wie Kosten und Reisezeit können in die parameterlinearen Nutzenfunktionen entweder direkt als **generische Variable** oder alternativenspezifisch formuliert werden.
- Sozioökonomischer Merkmale oder externe Einflüsse müssen immer alternativenspezifisch, also durch **alternativenspezifische Variable** modelliert werden. Eine Alternative bekommt dabei keinen Nutzenanteil und dient als **Referenzalternative**
- Globalen (nicht durch andere Faktoren beschriebene) Präferenzen werden durch einen vollständigen Satz von $I - 1$ **alternativenspezifische Konstanten** modelliert
- Nichtlinearitäten in den Charakteristika oder sozioökonomischen Variablen sowie Kopplungen von sozioökonomischen und externen Variablen an generische Variable und untereinander sind möglich und können sinnvoll sein.

Verständnisfrage:

Warum wäre in diesem Beispiel die Berücksichtigung einer einkommensabhängigen Zeitsensitivität durch einen exogenen Faktor $K_{ni}E_i$ anstelle von $T_{ni}E_i$ nicht so günstig? Berücksichtigen Sie dazu die Konsequenz für extrem hohen Einkommen bei sinnvollen Parameterwerten (sinnvollen Vorzeichen).

Verständnisfrage:

Könnte man durch entsprechende Wahl der generischen Einflussfaktoren in Gl. (4.4) berücksichtigen, dass einer Person ohne Kfz-Verfügbarkeit die Alternative "MIV" unzugänglich bleibt, ohne die Alternativenmenge zu ändern? (Dies würde die Auswertung mit Ökonometrie- bzw. Statistik-Programmen enorm erleichtern).

Verständnisfrage:

Warum führt die direkte Beschreibung alternativenspezifischer Zeitsensitivitäten durch (4.12) zu hohen positiven Korrelationen der Parameterschätzer, während dies für die Formulierung einer Referenz-Zeitsensitivität (4.6) und Sensitivitätsdifferenzen (4.11) nicht notwendigerweise gilt? Argumentieren Sie vom Sachverhalt ausgehend (nicht zu große Sensitivitätsunterschiede) und berücksichtigen Sie, dass nur Nutzendifferenzen relevant sind!

4 Modelle diskreter Entscheidungen

Verständnisfrage:

Könnte man durch entsprechende Wahl der generischen Einflussfaktoren in Gl. (4.4) berücksichtigen, dass einer Person ohne Kfz-Verfügbarkeit die Alternative "MIV" unzugänglich bleibt, ohne die Alternativenmenge zu ändern? (Dies würde die Auswertung mit Ökonometrie- bzw. Statistik-Programmen enorm erleichtern).

Aufgabe:

Definieren Sie genau, was konkrete Zahlenwerte der in den obigen Beispielen verwendeten Parameter β_1 bis β_{12} aussagen. Geben Sie dazu Änderungen der Nutzenfunktionen in "Nutzeneinheiten" (NE) (entspricht einer Änderung von V_{ni} um 1) an und definieren Sie ggf. geeignete Einheiten der exogenen Faktoren. Falls nur ein bestimmtes Vorzeichen sinnvoll ist, weisen Sie darauf hin und geben Sie es an.

4.2.4 Zufallsnutzen

Grundlage nahezu aller *discrete choice* Modelle ist das Konzept des rationalen Entscheiders:

Der rationale Entscheider oder **Homo Oeconomicus** wählt die Alternative mit dem höchsten Nutzen.

Man könnte auf die Idee kommen, dieses Prinzip direkt auf die deterministische Nutzenfunktion V_{ni} anzuwenden. Man erhält also für die von Person(engruppe) n gewählte Alternative $i_n^{(\text{select})}$ die Bedingung $i_n^{(\text{select})} = \arg(\max_i V_{ni})$. Hierbei ist $\arg(\cdot)$ die bereits in Abschnitt 4.1 eingeführte Argumentfunktion, welche das Ergebnis der Maximierung in der Klammer angibt. Das Problem hierbei ist, dass diese Maximierungsbedingung nur bei vollständig bekannter Nutzenfunktion konsistent mit dem Prinzip des *Homo Oeconomicus* ist. Analog zu den Regressionsmodellen (Kap. 2.3) wird daher ein **Zufallsnutzen** ϵ_{ni} eingeführt, welcher zusammen mit dem deterministischen Nutzen den Gesamtnutzen U_{ni} ergibt:

$$U_{ni} = V_{ni} + \epsilon_{ni}. \quad (4.13)$$

Es gibt mehrere Gründe, warum ein Zufallsnutzen notwendig ist, will man das Prinzip des *Homo Oeconomicus* beibehalten:

- (1) Nicht alle exogenen Variablen werden berücksichtigt und/oder sind bekannt. Damit kann man schreiben (in dieser Aufzählung wird der Personenindex weggelassen):

$$U_i = U(\underbrace{C_i, S}_{\text{bekannt}}, \underbrace{C'_i, S'}_{\text{unbekannt}}) = V(C_i, S) + \epsilon_i^{(1)}.$$

4 Modelle diskreter Entscheidungen

Achtung: Die vernachlässigten Einflussfaktoren dürfen keine systematischen Wirkungen entfalten. Ansonsten sind die Zufallsnutzen mit den deterministischen Nutzen korreliert und das Problem ist fehlspezifiziert. Dann gilt: "Junk in- Junk out"!

- (2) Die exogenen Variablen sind nicht genau messbar:

$$U_i = U(\underbrace{C_i}_{\text{Messfehler}} + \underbrace{\epsilon_i}_{\text{Messfehler}}, \underbrace{S}_{\text{Messfehler}} + \underbrace{\epsilon}_{\text{Messfehler}}) = V(C_i, S) + \epsilon_i^{(2)}.$$

- (3) Die exogenen Variablen sind nicht direkt, sondern nur indirekt und mit unvermeidlichen Fehlern über **Instrumentenvariable** (*observables*) I messbar, z.B. für sozioökonomische Variable:

$$S = f(I_s) + \epsilon.$$

Beispielsweise könnte man das Einkommen über die Instrumentenvariable "Wohnlage" abschätzen. Dies führt zu Zufallsnutzenanteilen $\epsilon_i^{(3)}$ analog zum Fall (2).

- (4) Bei den Entscheidern selbst existiert ein nichtrationales bzw. zufälliges Element. Die Unterscheidung zwischen (1) und (4) ist aber schon nahezu philosophisch, da – gemäß dem Bonmot, dass Zufallsterme nichts anderes als das Eingeständnis von Unwissen darstellen – das scheinbar nichtrationale Verhalten bei Berücksichtigung *aller* Einflussgrößen in Wirklichkeit doch ein Rationales sein könnte. Grundsätzlich wird dies angenommen, da ein Verzicht auf den *Homo Oeconomicus* in diesem Rahmen schlechterdings nicht modellierbar ist.

Die Elemente des Zufallsnutzens gehorchen einer oder mehrerer der folgenden statistischen Eigenschaften, die im Wesentlichen den Gauß-Markow-Annahmen 2.3.4 der Regressionsmodelle entsprechen:

1. Der Zufallsnutzen ist additiv und hat einen verschwindenden Erwartungswert:¹⁴

$$U_{ni} = V_{ni} + \epsilon_{ni}, \quad E(\epsilon_{ni}) = 0. \quad (4.14)$$

2. Unabhängigkeit bezüglich Personen bzw. Einzelentscheidungen:

$$E(\epsilon_{ni}\epsilon_{n'i}) = 0 \quad \text{falls } n \neq n'. \quad (4.15)$$

3. Unabhängigkeit bezüglich Alternativen:

$$E(\epsilon_{ni}\epsilon_{n'i'}) = \sigma_n^2 \delta_{ii'}. \quad (4.16)$$

Hier wurden wieder Selektorfunktionen ($\delta_{ii'} = 0$ für $i \neq i'$ und $=1$ für $i = i'$) eingesetzt.

¹⁴bzw. einen konstanten Erwartungswert. Da es nur auf Nutzendifferenzen ankommt, ist dies gleichwertig.

4 Modelle diskreter Entscheidungen

4. Homoskedastizität (einheitliche Varianz):

$$\sigma_{ni}^2 = E(\epsilon_{ni}^2) = \sigma_\epsilon^2 = \text{konst.} \quad (4.17)$$

Welche dieser Eigenschaften vorausgesetzt werden, hängt von der Modellklasse ab:

- Im **Logit-Modell** sind alle obigen Eigenschaften erfüllt, die Zufallsnutzen sind also i.i.d. Außerdem ist die Verteilung durch die **Gumbel-Verteilung** spezifiziert (siehe Abschnitt 4.4). Manchmal unterscheidet man nach der Zahl der Alternativen:
 - Das **Binomiale Logit Modell** beschreibt genau zwei Alternativen,
 - das **Multinomiale Modell** beschreibt $I \geq 3$ Alternativen.
- Im allgemeinen **Probit-Modell** sind die Zufallsnutzen bezüglich der Alternativen multivariat gaußverteilt, es ist also nur Bedingung (4.14) und die Unabhängigkeit (4.15) zwischen Personen) erfüllt. Im Sonderfall von i.i.d. gaußverteilten Zufallsnutzen (oder allgemein für $I = 2$ Alternativen, da dann Korrelationen nicht messbar und damit nicht relevant sind) sind die Ergebnisse des Probit-Modells nahezu identisch zu denen des Logit-Modells.
- Im **Nested Logit-Modell** werden bei den gumbelverteilten Zufallsnutzen in beschränktem Umfang Korrelationen bezüglich der Alternativen (aber nicht bezüglich der Personen) zugelassen. Gemeinsam mit dem “normalen” Logitmodell ist es die einzige Modellklasse, bei der sich auch im multinomialen Fall die Auswahlwahrscheinlichkeiten analytisch angeben lassen.
- In der Klasse der **GEV-Modelle** (*generalized extreme value*) werden weitere Freiheitsgrade der Korrelationen und auch bei den Verteilungen selbst frei (verallgemeinerte Extremwertverteilungen). Dies sind die allgemeinsten Modelle mit analytischen Auswahlwahrscheinlichkeiten.
- Schließlich haben die Zufallsnutzen des **Mixed-Logit** Modells einen i.i.d. Gumbel- und einen beliebigen Anteil. Sie stellen eine “Mischung” unterschiedlich parametrisierter Logitmodelle dar und benötigen im Allgemeinen numerische Integration oder Monte-Carlo-Verfahren zur Berechnung der Auswahlwahrscheinlichkeiten.

Wichtig ist, dass die Bedingungen der Additivität immer erfüllt sein müssen, was nicht notwendigerweise der Fall ist. Insbesondere führt Fehlerquelle 1 der Aufzählung am Beginn dieses Unterabschnitts häufig zu einer Verletzung. Dies gilt erst recht, wenn die durch die Funktionen $g_{mi}(\cdot)$ spezifizierten deterministischen Nutzenfunktionen falsch sind, so dass kein Parametervektor β das tatsächliche Verhalten abbilden kann.

Prinzipiell modellieren alle obigen Modelle eine einzelne Entscheidung und man kann durchaus Heteroskedastizität zwischen einzelnen Personen/Entscheidungen n ansetzen, solange die Unabhängigkeit gewahrt bleibt (Teilverletzung von Annahme (4.15)). Insbesondere gilt: Ist ein Modell für die Einheitsvarianz analytisch lösbar, dann auch für beliebige Varianzen.

4 Modelle diskreter Entscheidungen

4.2.5 Modellierung der Wahlentscheidung

Mit der Einführung von Zufallsnutzen kann es keine deterministische Entscheidung mehr geben. An dessen Stelle treten **Entscheidungswahrscheinlichkeiten** P_{ni} dafür, dass Person n die Alternative i wählt. Da nach Bedingung (4.15) die Zufallsterme bezüglich der Personen unabhängig sind, hängt die analytische Form der Entscheidungswahrscheinlichkeiten nicht vom Personengruppenindex n ab, so dass dieser im folgenden der Übersichtlichkeit halber weggelassen wird.

Die konkrete Einzelentscheidung läuft auch in den Modellen mit Zufallsnutzen streng deterministisch gemäß dem Prinzip des *Homo Oeconomicus* ab:

$$\xi^{\text{select}} = \arg \left(\max_i (V_i + \epsilon_i) \right). \quad (4.18)$$

Man beachte, dass ξ^{select} als Funktion des Vektors $\epsilon = (\epsilon_1, \dots, \epsilon_I)'$ selbst eine Zufallsvariable darstellt.

Wir betrachten nun den Fall unabhängiger Zufallsnutzen und berechnen, ohne Einschränkung der Allgemeinheit, die Wahrscheinlichkeit dafür, Alternative 1 zu wählen. Hält man zunächst den Zufallsnutzen ϵ_1 fest, ergibt sich die bedingte Auswahlwahrscheinlichkeit zu

$$\begin{aligned} P_1 | \epsilon_1 &= \text{Prob} \left(\xi^{\text{select}} = 1 | \epsilon_1 \right) \\ &= \text{Prob} \left[(V_1 + \epsilon_1 > V_2 + \epsilon_2) \cap \dots \cap (V_1 + \epsilon_1 > V_I + \epsilon_I) \right] \\ &= \prod_{i=2}^I \text{Prob}(V_1 + \epsilon_1 > V_i + \epsilon_i) \\ &= \prod_{i=2}^I \text{Prob}(\epsilon_i < \epsilon_1 + V_1 - V_i) \\ &= \prod_{i=2}^I F_i(\epsilon_1 + V_1 - V_i). \end{aligned}$$

Hierbei wurde die Unabhängigkeit (die Wahrscheinlichkeit, dass mehrere Bedingungen zutreffen ist das Produkt der Einzelwahrscheinlichkeiten) und die Definition der Verteilungsfunktion der Zufallsnutzen, $F_i(x) = \text{Prob}(\epsilon_i < x)$, angewandt.

Nun ist aber ϵ_1 nicht bekannt. Man muss daher, nach dem Satz über totale Wahrscheinlichkeiten, die bedingten Wahrscheinlichkeiten über alle möglichen Bedingungen ϵ_1 , multipliziert mit den Wahrscheinlichkeiten $f_1(\epsilon_1) d\epsilon_1$ des Eintreffens der jeweiligen Bedingung, aufsummieren bzw. integrieren:

$$P_1 = \int_{-\infty}^{\infty} \text{Prob} \left(\xi^{\text{select}} = 1 | \epsilon_1 \right) f_1(\epsilon_1) d\epsilon_1 = \int_{-\infty}^{\infty} \prod_{i=2}^I F_i(\epsilon_1 + V_1 - V_i) f_1(\epsilon_1) d\epsilon_1. \quad (4.19)$$

Diskussion dieses sehr allgemeinen Ausdrucks:

4 Modelle diskreter Entscheidungen

- Zur Herleitung haben wir die Bedingungen (4.14) (Additivität) und (4.16) (Unabhängigkeit), nicht aber Bedingung (4.17) (Homoskedastizität) verwendet. Nimmt man diese und gleich auch noch identische Verteilungen an, kann man die Indices bei den F_i weglassen.
- Ausdruck (4.19) ist ein eindimensionales Integral, welches sich numerisch schnell lösen lässt.
- Ein wichtiger Spezialfall ist das multinomiale Logit Modell mit i.i.d. gumbelverteilten Zufallsnutzen, bei der sich das Integral analytisch lösen lässt.

Im Fall zweier Alternativen (binomiale Modelle) kann man Gl. (4.19) für viele Verteilungsfunktionen direkt ausrechnen. Insbesondere kann man dann die Gleichung mit einem Faltungssatz identifizieren: Allgemein gilt nämlich für die Verteilungsfunktion und die Dichte der Differenz $Y = X_2 - X_1$ zweier stetiger unabhängiger Zufallsvariablen X_i , $i = 1, 2$, mit Verteilungsfunktionen $F_i(x)$ und Dichten $f_i(x)$:

$$Y = X_2 - X_1 \Rightarrow F_y(y) = \int_{-\infty}^{\infty} f_1(x)F_2(y+x) dx, \quad f_y(y) = \int_{-\infty}^{\infty} f_1(x)f_2(y+x) dx \quad (4.20)$$

Vergleicht man dies mit (4.19), sieht man, dass im binomialen Fall die Auswahlwahrscheinlichkeit P_1 durch die Verteilungsfunktion $F_{\epsilon_2 - \epsilon_1}$ der Differenz $\epsilon_2 - \epsilon_1$ der Zufallsnutzen an der Stelle der umgekehrten Differenz der deterministischen Nutzen gegeben ist:

$$P_1 = F_{\epsilon_2 - \epsilon_1}(V_1 - V_2), \quad P_2 = 1 - P_1 \quad (4.21)$$

Dies gilt selbst für korrelierte Zufallsnutzen, die man dann allerdings nicht mehr mit (4.20) berechnen kann:

$$\begin{aligned} P_1 &= \text{Prob}(U_1 > U_2) \\ &= \text{Prob}(V_1 + \epsilon_1 \geq V_2 + \epsilon_2) \\ &= \text{Prob}(\epsilon_2 - \epsilon_1 \leq V_1 - V_2) \\ &= F_{\epsilon_2 - \epsilon_1}(V_1 - V_2). \end{aligned}$$

In Worten:

In binomialen Modellen mit beliebig verteilten und bezüglich der zwei Alternativen beliebig korrelierten Zufallsnutzen ist die Auswahlwahrscheinlichkeit für Alternative 1 durch die Verteilungsfunktion der Zufallsnutzendifferenz $\epsilon_2 - \epsilon_1$ an der Stelle der deterministischen Nutzendifferenz $V_1 - V_2$ gegeben.

4.3 Probit-Modelle

Probit-Modelle sind durch i.i.d oder korrelierte gaußverteilte Zufallsnutzen definiert, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Im Falle korrelierter Zufallsvariablen sind die Koeffizienten der Korrelationsmatrix Teil der Modellparameter.

Ein *Vorteil* der Probit-Modells besteht in der Konsistenz der Herleitung dieser Modelle für viele praktische Fälle: Nach dem **Zentralen Grenzwertsatz der Statistik** ist nämlich eine aus vielen voneinander unabhängigen Einflussgrößen (z.B. denen von Abschnitt 4.2.4) resultierende Zufallsgröße normalverteilt, *unabhängig* von den Verteilungen der einzelnen Einflussgrößen, solange jeder Einflussfaktor für sich nur einen kleinen Anteil hat. Unter sehr allgemeinen und schwachen Annahmen bekommt man daher eine Gaußverteilung für die ϵ_i .

Der *Nachteil* des Probit-Modells ist die für den multinomialen Fall nicht mögliche analytische Berechenbarkeit der Auswahlwahrscheinlichkeiten.

4.3.1 Binomiales Probit-Modell

Die Summe und Differenz zweier gaußverteilter Zufallsvariablen ist wieder gaußverteilt, wobei der Erwartungswert die Summe bzw. Differenz und, bei Unabhängigkeit, die Varianz immer die *Summe* der Einzelvarianzen beträgt (Abb. 4.5). Bezeichnet man mit $\Phi(z)$ die tabellierte Verteilungsfunktion der Standardnormalverteilung bzw. (0,1)-Gaußverteilung, ergibt sich aus (4.21) direkt

$$P_1^{\text{Probit}} = \Phi\left(\frac{V_1 - V_2}{\sqrt{2}\sigma_\epsilon}\right), \quad P_2 = 1 - P_1. \quad (4.22)$$

4.3.2 Multinomiales i.i.d Probit-Modell

Falls man wieder Unabhängigkeit der ϵ_i voraussetzt, also nicht die allgemeinste Form der multivariaten Normalverteilung mit nichttrivialer Kovarianzmatrix (vgl. Abschnitt 2.11.4), erhält man aus (4.19) folgende Auswahlwahrscheinlichkeit für Alternative 1 (f_1 bezeichnet die Dichte der $(0, \sigma_\epsilon)$ -Normalverteilung):

$$P_1 = \int_{-\infty}^{\infty} f_1(\epsilon_1) \Phi\left(\frac{\epsilon_1 + V_1 - V_2}{\sigma_\epsilon}\right) \dots \Phi\left(\frac{\epsilon_1 + V_1 - V_I}{\sigma_\epsilon}\right) d\epsilon_1. \quad (4.23)$$

Der Integrand besteht also aus einem Produkt von (kumulierten) Standardnormalverteilungen (vgl. Abb. 4.6).

4 Modelle diskreter Entscheidungen

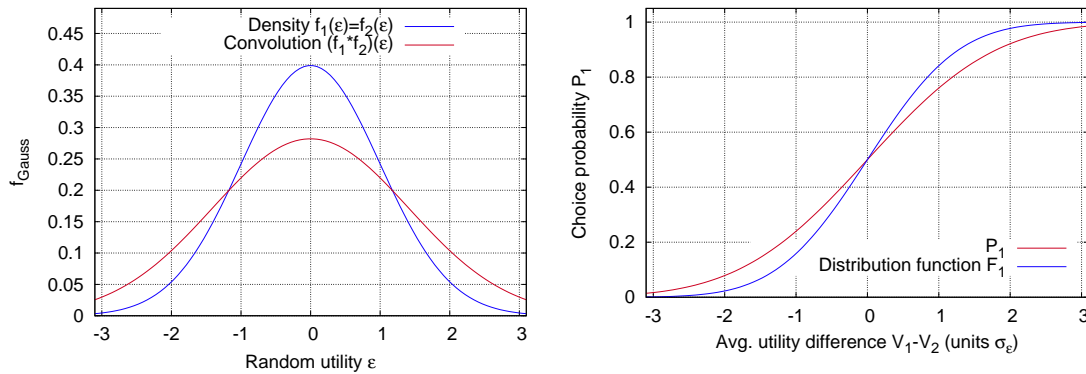


Abbildung 4.5: Probit-Modell: Links: Dichte der normalverteilten Zufallsnutzen ϵ_1 und ϵ_2 sowie die Dichte der Nutzendifferenz $\Delta\epsilon = \epsilon_2 - \epsilon_1$. Rechts: Verteilungsfunktion der Zufallsnutzen und resultierende Auswahlwahrscheinlichkeit für Alternative 1 in Abhängigkeit der deterministischen Nutzendifferenz $V_1 - V_2$.

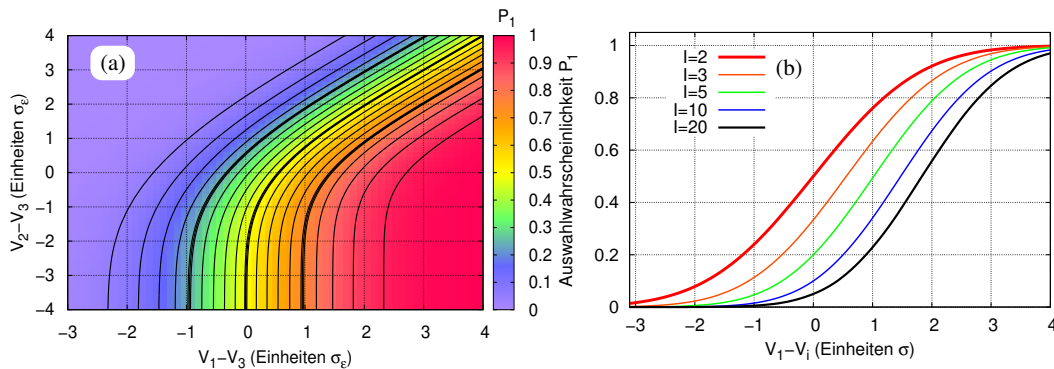


Abbildung 4.6: (a) Auswahlwahrscheinlichkeit P_1 gemäß Gl. (4.23) für Alternative 1 beim trinomialen Probitmodell in Abhängigkeit der deterministischen Nutzendifferenzen zu (Referenz-) Alternative 3; (b) Auswahlwahrscheinlichkeit P_1 im multinomialen Probitmodell als Funktion der Nutzendifferenz $V_1 - V_i$, falls alle anderen Alternativen i im Mittel gleich attraktiv sind ($V_i = V_2$ für $i \geq 2$).

4 Modelle diskreter Entscheidungen

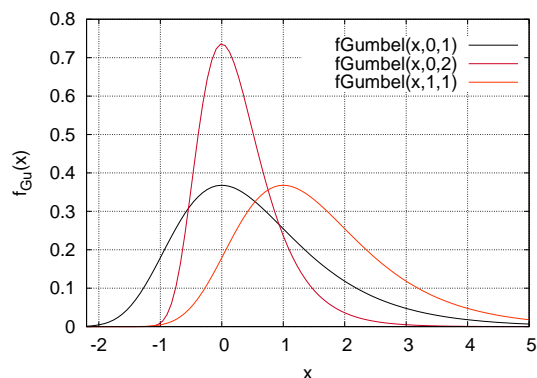


Abbildung 4.7: Dichten verschiedener Extremwert- bzw. Gumbelverteilungen.

4.4 Logit-Modelle

Logit-Modelle sind durch i.i.d. gumbelverteilte Zufallsnutzen definiert.

Die Gumbelverteilung hat die Verteilungsfunktion

$$F_{\text{Gu}}^{(\eta,\lambda)}(x) = \exp \left[-e^{-\lambda(x-\eta)} \right] \quad (4.24)$$

und die Dichtefunktion (vgl. Abb. 4.7)

$$f_{\text{Gu}}^{(\eta,\lambda)}(x) = \frac{dF_{\text{Gu}}^{(\eta,\lambda)}(x)}{dx} = \lambda e^{-\lambda(x-\eta)} \exp \left[-e^{-\lambda(x-\eta)} \right]. \quad (4.25)$$

Für die Anwendung bei diskreten Wahlmodellen wichtige Eigenschaften der Gumbelverteilung sind die folgenden:

- Der Lageparameter η hat die Bedeutung des Modalwerts (Wert mit der höchsten Dichte) und der Skalierungsparameter λ ist proportional zur Inversen der Standardabweichung σ_ϵ :

$$x_{\text{modal}} = \eta, \quad \sigma_\epsilon = \frac{\pi}{\sqrt{6}\lambda} \approx 1.28/\lambda \quad \text{bzw.} \quad \lambda = \frac{\pi}{\sqrt{6}\sigma_\epsilon} \quad (4.26)$$

Da η die Verteilung nur verschiebt und λ nur skaliert, ist auch eine lineare Funktion $Y = a + bX$ einer (η, λ) -gumbelverteilten Zufallsvariablen wieder gumbelverteilt:

$$X \sim \text{Gu}(\eta_i, \lambda) \Rightarrow a + bX \sim \text{Gu}(a + b\eta, \lambda/b). \quad (4.27)$$

Da es bei den Auswahlwahrscheinlichkeiten nur auf Nutzendifferenzen ankommt, bedeutet diese Translationsinvarianz insbesondere, dass der Lageparameter η keinerlei Rolle spielt (und es somit egal ist, dass η den Modalwert und nicht den Erwartungswert angibt).

4 Modelle diskreter Entscheidungen

- Die Gumbelverteilung mit Skalierungsparameter λ ist die Grenzverteilung des Maximums sehr vieler unabhängiger stetiger Zufallsvariablen, solange die Dichte dieser Variablen für große Werte exponentiell mit *demselben* Exponenten λ abfällt¹⁵. Deshalb heißt die Gumbelverteilung auch **(Typ I-) Extremwertverteilung**. Die Abbildung 4.8 und 4.9 zeigen diese Grenzverteilungs-Eigenschaft anschaulich. Für Interessierte befindet sich die Herleitung im Abschnitt 4.16.

Insbesondere ist das Maximum zweier oder mehrerer unabhängiger gumbelverteilter Zufallsgrößen *gleicher Varianz* (bzw. mit identischen Skalierungsparameter λ) wieder gumbelverteilt und zwar *ebenfalls* mit unveränderter Varianz bzw. Skalierungsparameter λ (Herleitung ebenfalls im Abschnitt 4.16):

$$X_i \sim \text{Gu}(\eta_i, \lambda) \Rightarrow Y = \max_i X_i \sim \text{Gu} \left[\frac{1}{\lambda} \ln \left(\sum_{i=1}^I e^{\lambda \eta_i} \right), \lambda \right]. \quad (4.28)$$

Diese Eigenschaft ist auch unmittelbar einsichtig, da nach dem vorhergehenden Satz jede gumbelverteilte Zufallsvariable für sich das Maximum exponentiell abfallender Verteilungen ist. Das Maximum gumbelverteilter Variablen ist damit ebenfalls ein Maximum exponentiell abfallender Verteilungen, also gumbelverteilt. Insbesondere gilt bei i.i.d gumbelverteilten Zufallsnutzen wie im Logit-Modell

$$X_i \sim \text{i.i.d. Gu}(\eta, \lambda) \Rightarrow Y = \max_i X_i \sim \text{Gu} \left(\eta + \frac{\ln I}{\lambda}, \lambda \right). \quad (4.29)$$

- Schließlich ist die *Differenz* $Y = X_2 - X_1$ zweier mit *demselben* Skalierungsparameter λ gumbelverteilten Größen logistisch verteilt:

$$X_1 \sim \text{Gu}(\eta_1, \lambda), X_2 \sim \text{Gu}(\eta_2, \lambda) \Rightarrow P(X_2 - X_1 < y) = \frac{1}{1 + e^{-\lambda(y - \eta_2 + \eta_1)}}. \quad (4.30)$$

4.4.1 Binomiales Logit Modell

Die Auswahlwahrscheinlichkeiten des Logitmodells mit zwei Alternativen bekommt man direkt aus dem allgemeinen Ausdruck (4.21) mit Hilfe der Differenz-Beziehung (4.30) gumbelverteilter Größen:

$$\begin{aligned} P_1^{\text{BNL}} &= F_{\epsilon_2 - \epsilon_1}(V_1 - V_2) = P(\epsilon_2 - \epsilon_1 < V_1 - V_2) = \frac{1}{1 + e^{-\lambda(V_1 - V_2)}}, \\ P_2^{\text{BNL}} &= 1 - P_1^{\text{BNL}}. \end{aligned} \quad (4.31)$$

4.4.2 Multinomiales Logit Modell

Der Vorteil der Gumbel-Verteilung ist, dass man aufgrund der Beziehungen (4.30) mit (4.27) und (4.28) die Auswahlwahrscheinlichkeiten auch für beliebige Alternativenzahlen, einfach analytisch ausdrücken kann (Herleitung für Interessierte in Abschnitt 4.16):

4 Modelle diskreter Entscheidungen

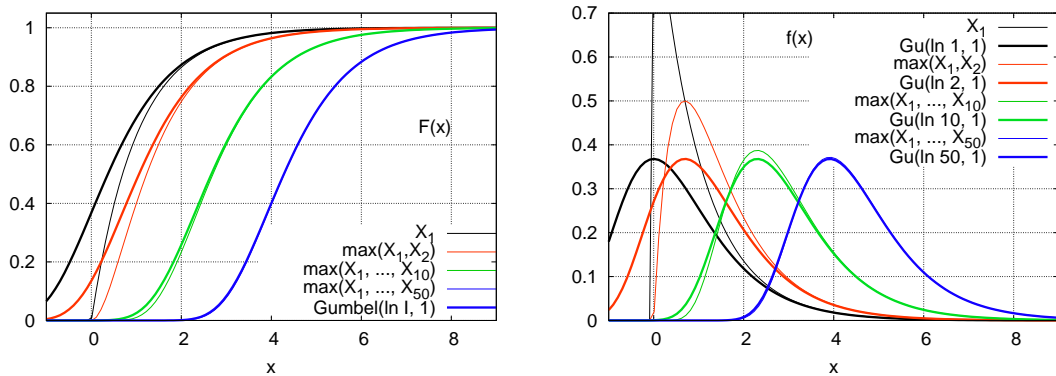


Abbildung 4.8: Die Gumbelverteilung als Grenzverteilung des Maximums sehr vieler unabhängiger Zufallsgrößen mit *exponential tail* am Beispiel exponentialverteilter Zufallsgrößen $X \sim E(\lambda)$ mit $\lambda = 1$. Geplottet ist die Verteilungsfunktion (links) und die dazugehörige Dichte (rechts) jeweils für das Maximum von $I = 1, 2, 5, 20$ und 100 dieser Zufallsgrößen und die dazugehörige Gumbel-Grenzverteilung $Y \sim \text{Gu}(\ln I, 1)$. Bereits bei $I = 20$ gibt es kaum einen sichtbaren Unterschied zwischen der tatsächlichen Verteilung des Maximums und der Gumbelverteilung.

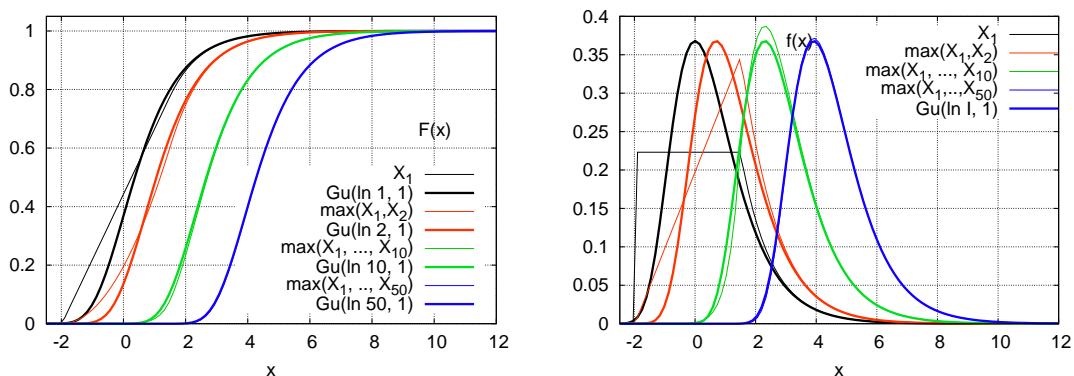


Abbildung 4.9: Die Grenzverteilungseigenschaft funktioniert auch bei anderen Verteilungen, solange ein *exponential tail* existiert. Hier ist das Maximum von unabhängigen Zufallsvariablen gezeigt, welche im Intervall $-1.98 \leq x \leq 1.5$ gleichverteilt sind (konstante Dichte) und erst danach exponentiell (Exponent $\lambda = 1$) abfallen. Die dünnen Kurven geben jeweils die zur Zahl I der exponentiell abfallenden Zufallsvariablen gehörige Extremwertverteilung (Gumbelverteilung) $\text{Gu}(\ln I, 1)$ an. Wieder gibt es bereits bei $I = 20$ kaum einen sichtbaren Unterschied zwischen der tatsächlichen Verteilung des Maximums und der Gumbelverteilung.

4 Modelle diskreter Entscheidungen

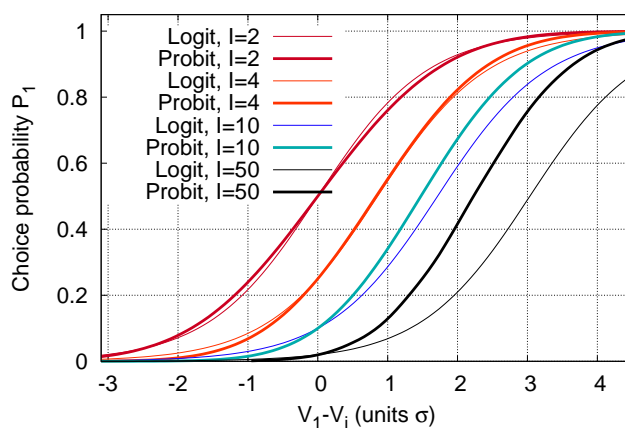


Abbildung 4.10: Auswahlwahrscheinlichkeiten für Alternative 1 im Multinomial-Logitmodell (dick) und Multinomial-Probitmodell (dünn) als Funktion der deterministischen Nutzendifferenz $V_1 - V_i$ dieser Alternativ zu den als gleich hoch angenommenen anderen Nutzen $V_i = V_2$ für $i \geq 2$. Da die Gumbelverteilung rechtsschief ist, die Gaußverteilung aber nicht, sind beim Logitmodell bei vielen Alternativen vergleichsweise höhere Werte von V_1 nötig als beim Logitmodell, um dieselbe Auswahlwahrscheinlichkeit zu erreichen.

$$P_i^{\text{MNL}} = \frac{e^{\lambda V_i}}{\sum_{i'=1}^I e^{\lambda V_{i'}}} = \frac{1}{\sum_{i'=1}^I e^{\lambda(V_{i'} - V_i)}} \quad (4.32)$$

Das resultierende Multinomial-Logit-Modell (MNL) hat folgende wichtige Eigenschaften:

- Es kommt (ebenso wie bei den Probit-Modellen) nur auf die *Differenzen* der deterministischen Nutzen an. Man sagt auch, das Modell ist (ebenso wie die Probit-Modelle) **translationsinvariant**. Man kann also ohne Einschränkung der Allgemeinheit den Nutzen einer Alternative (z.B. V_1 oder V_I) =0 setzen. Dies wurde bereits in Abschnitt 3.6.2 ausgenutzt.
- Die deterministischen Nutzen kommen nur als Produkt mit dem Skalenparameter λ bzw. in Einheiten des $\sqrt{6}/\pi$ -fachen der Zufallsnutzen-Standardabweichung vor:

$$\lambda V_i = \frac{\pi}{\sqrt{6}} \frac{V_i}{\sigma_\epsilon} \quad (4.33)$$

¹⁵Dies ist neben dem Zentralen Grenzwertsatz (die *Summe* sehr vieler beliebig verteilter unabhängiger Zufallsvariablen mit endlicher Varianz ist gaußverteilt) ein weiteres Beispiel eines Grenzwertsatzes.

4 Modelle diskreter Entscheidungen

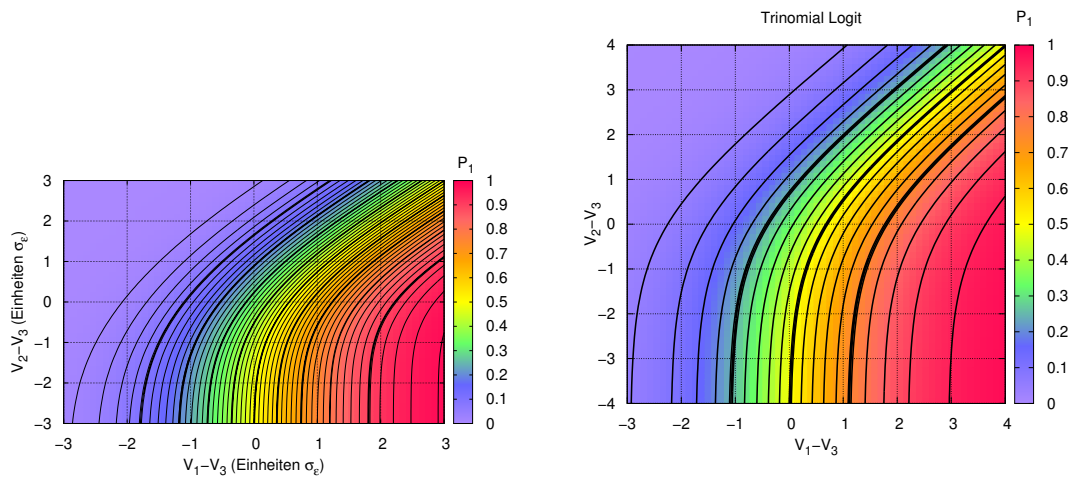


Abbildung 4.11: Auswahlwahrscheinlichkeiten für Alternative 1 beim trinomialen ($I = 3$) Probit-Modell (links) und dem entsprechenden Logit-Modell (rechts) in Abhängigkeit der in Einheiten von σ_ϵ skalierten Nutzenfunktionen. Ohne Einschränkung der Allgemeinheit kann der deterministische Nutzen der Referenzalternative $V_3 = 0$ gesetzt werden.

Man sagt auch, das Logit-Modell ist (ebenso wie die Probit-Modelle) **skaleninvariant**. Dies wurde ebenfalls bereits in Abschnitt 3.6.2 ausgenutzt. Skaliert man den deterministischen *und* den Zufallsnutzen in Einheit des $\sqrt{6}/\pi$ -fachen der Standardabweichung σ_ϵ des Zufallsnutzens, dividiert also die unskalierten Nutzenfunktionen durch $\sigma_\epsilon \sqrt{6}/\pi$, dann ist in den skalierten Größen der Gumbel-Skalenparameter $\lambda = 1$ und man erhält das MNL in seiner gebräuchlichen standardisierten Form:

$$P_i^{\text{MNL}} = \frac{e^{V_i}}{\sum_{i'=1}^I e^{V_{i'}}} = \frac{1}{\sum_{i'=1}^I e^{V_{i'} - V_i}}. \quad (4.34)$$

Zum Schätzen der (in den deterministischen Nutzenfunktionen V_i enthaltenen) Modellparameter β *und* der Standardabweichung des Zufallsnutzens kann man diese Form verwenden (vgl. dazu Abschnitt 4.4.3).

- Das MNL hat, *im Gegensatz zum Probit-Modell*, die sogenannte **IIA-Eigenschaft**. Dieses Kürzel kommt von der englischen Bedeutung *Independence of irrelevant alternatives* und sagt aus, dass die Quotienten zweier Auswahlwahrscheinlichkeiten, also die relative Gewichtung zweier Alternativen, *nicht* von den Nutzen irgendwelcher weiterer Alternativen abhängen. Dies gilt *nur* für das MNL: Man kann sogar das Logitmodell (und daraus die IIA-Eigenschaft) nicht nur aus gumbelverteilten Zufallsnutzen herleiten, sondern umgekehrt auch das Logitmodell aus der

4 Modelle diskreter Entscheidungen

IIA-Eigenschaft:

$$\text{IIA} \iff \text{MNL} \quad (4.35)$$

- Das MNL gilt (näherungsweise) nur für exponentiell abfallende Zufallsnutzen, also insbesondere nicht für gaußverteilte Zufallsnutzen, da diese quadratisch-exponentiell abfallen. Die im Vergleich zur Gaußverteilung höheren Wahrscheinlichkeiten extremer Werte der Zufallsnutzen führen dazu, dass die Wahrscheinlichkeit P_i als Funktion von V_i langsamer ansteigt als beim Probitmodell (Abb. 4.10). Genauer gesagt, ist die Steilheit des Anstiegs von P_i mit V_i unabhängig von der Alternativenzahl (IIA-Eigenschaft), während sie beim Multinomial-Probit-Modell steiler wird. Bis zu $I = 4$ Alternativen, also dem Standardfall der Verkehrsmittelwahl, sind die Unterschiede gering (Abb. 4.10 und 4.11).

4.4.3 Interpretation der Parameter bei Probit- und Logitmodellen

Da weder ein Hinzuftügen eines konstanten Beitrags zu den Nutzenfunktionen aller Alternativen noch eine Multiplikation der deterministischen *und* stochastischen Nutzenanteile mit einem gemeinsamen positiven Faktor die Auswahlwahrscheinlichkeiten ändern (Translations- bzw. Skaleninvarianz)¹⁶ muss man bei der Kalibrierung weder die Lage- noch die Skalenparameter der jeweiligen Verteilung des Zufallsnutzens schätzen. Vielmehr kann beim Logitmodell $\lambda = 1$ bzw. beim Probitmodell $\sigma_\epsilon = 1$ setzen und erhält damit die standardisierten Formulierungen dieser Modelle. Mit (4.33) bedeutet dies, dass die Nutzenfunktionen des standardisierten Probitmodells in den Einheiten σ_ϵ und die des standardisierten Logitmodells in den Einheiten $(\sqrt{6}/\pi) \sigma_\epsilon$ skaliert sind. Die Abb. 4.10 und 4.11 zeigen, dass bei nicht zu großer Alternativenzahl und nicht zu extremen Wahrscheinlichkeiten nahe null oder eins die Aussagen der Logit- und i.i.d. Probitmodelle nahezu identisch sind, *wenn man gleiche Nutzeinheiten, z.B. σ_ϵ , verwendet*. Da nach (4.33) beim standardisierten Logitmodell ($\lambda = 1$) im Vergleich zum standardisierten Probitmodell ($\sigma_\epsilon = 1$) die Relation

$$V_{\text{Logit}} = \frac{\pi}{\sqrt{6}} \frac{V}{\sigma_\epsilon} = \frac{\pi}{\sqrt{6}} V_{\text{Probit}}$$

gilt, erhält man folgendes wichtige Ergebnis:

Standardisierte Logitmodelle liefern weitgehend äquivalente Aussagen wie standardisierte binäre bzw. i.i.d. multinomiale Probitmodelle, wenn alle Logit-Parameter um den Faktor $\pi/\sqrt{6} \approx 1.28$ größer sind als die jeweiligen Probit-Parameter. Teilt man alle linearen Parameter des standardisierten Logitmodells durch 1.28, erhält man die deterministische Nutzenfunktion in Einheiten von σ_ϵ .

¹⁶Man kann sogar die Gesamtnutzen U_i aller Alternativen mit einer streng monoton steigenden aber sonst beliebig nichtlinearen Funktionen behandeln, ohne dass sich die Auswahlwahrscheinlichkeiten ändern, allerdings geht dann die i.i.d.-Eigenschaft der Zufallsnutzen verloren.

4 Modelle diskreter Entscheidungen

Dieses Ergebnis liefert letztendlich eine Begründung, warum man gumbelverteilte Zufallsnutzen annehmen kann, obwohl eine Gaußverteilung (aufgrund des Zentralen Grenzwertsatzes) theoretisch konsistenter ist (vgl. den Anfang von Abschnitt 4.3): Das Ergebnis ist nahezu gleich, die Handhabung der Logitmodelle (analytische Ausrechenbarkeit, einfachere Kalibrierung) aber einfacher.

Man beachte jedoch, dass es auch deutliche Unterschiede zwischen den beiden Modelltypen gibt, vor allem folgt aus (4.35), dass das Probit-Modell selbst dann nicht der IIA-Eigenschaft, wenn man i.i.d - normalverteilte Zufallsnutzen annimmt (vgl. Übungsaufgabe):

$$\epsilon \sim \text{i.i.d} \not\Rightarrow \text{IIA}$$

Beispiel: Zwei Alternativen, Einfluss von Reisezeit und Kosten

Der deterministische Nutzen zweier Alternativen i sei durch

$$V_{ni} = \beta_1 T_{ni} + \beta_2 K_{ni} + \beta_3 \delta_{i1}$$

gegeben. Die komplexen Reisezeiten T_{ni} (in Minuten) sowie die Ad-Hoc-Kosten K_{ni} (in €) werden also als generische Variable formuliert. Zusätzlich gibt es noch einen vollständigen Satz an alternativenspezifischen Konstanten, der hier nur aus dem Beitrag $\beta_3 \delta_{i1}$ besteht. Nach getrennter Schätzung der drei Parameter im Logit- und Probitmodell sind folgende Aussagen möglich:

- Die Standardabweichung des Zufallsnutzens in Minuten ist durch $-1/\beta_1$ (Probitmodell) bzw. $-\pi/(\sqrt{6}\beta_1)$ (Logitmodell) gegeben,
- die Standardabweichung des Zufallsnutzens in € ist durch $-1/\beta_2$ (Probit) bzw. $-\pi/(\sqrt{6}\beta_2)$ (Logit) gegeben,
- der Zeitwert (VTTS, value of travel time savings) in € pro Minute ist gleich β_1/β_2 in beiden Modellen,
- der globale Bonus von Alternative 1 gegenüber der Referenzalternative 2 in Nutzeinheiten ist gleich β_3 ,
- der globale Bonus von Alternative 1 gegenüber 2 ist gleich β_3 Standardabweichungen (Probit) bzw. $(\sqrt{6}/\pi)\beta_3$ Standardabweichungen (Logit),
- der globale Bonus von Alternative 1 gegenüber 2 in Minuten ist gleich $-\beta_3/\beta_1$,
- und der globale Bonus von Alternative 1 gegenüber 2 in Euro ist gleich $-\beta_3/\beta_2$.

4.5 Elastizitäten

Allgemein, z.B. bei Regressionsmodellen, geben Elastizitäten die einheitenlose relative Änderung einer endogenen Variable \hat{y}_i als Folge einer relativer Änderung einer exogenen Variablen x_m an:

$$\epsilon_i^{(m)} = \frac{x_m}{\hat{y}_i} \frac{\partial \hat{y}_i}{\partial x_m} \quad (4.36)$$

4 Modelle diskreter Entscheidungen

Falls es nur eine endogene Variable gibt, wie bei der Regressionsrechnung, haben die Elastizitäten natürlich nur einen Index (bzw. Superskript), nämlich den der exogenen Variablen x_m . Im Gegensatz zur Regressionsrechnung sind in der diskreten Wahltheorie die Definitionen von Elastizitäten jedoch nichttrivialer und vielschichtiger. Die verschiedenen möglichen Definitionen kommen je nach Sachverhalt und Aufgabenstellung zur Anwendung und werden an folgendem Sachverhalt erklärt:

Veranschaulichungs-Beispiel: Flughäfen

Gegeben sind I miteinander konkurrierende Flughafenbetreiber i , z.B. $i = 1$: München (MUC), $i = 2$: Berlin (BER), $i = 3$: Frankfurt (FRA) usw., welche (über entsprechende Konditionen an die verschiedenen Airlines) die Flugpreise eines Fluges (z.B. nach Barcelona) für bestimmte Kundenklassen/Airlines n und damit die Nachfrage y_{ni} beeinflussen können.^a Wir nehmen an, dass es bei der Kundenklasse n die feste Nachfrage nach

$$y_n = \sum_{i=1}^I y_{ni} \quad (4.37)$$

Flügen gibt. Für die deterministischen Nutzenfunktionen der verschiedenen Flughäfen nehmen wir als Einflussfaktoren (neben den immer notwendigen Satz von $I - 1$ alternativenspezifischen Konstanten) nur die Gesamtkosten K_{ni} und die komplexen Reisezeiten T_{ni} (von der Haustür bis zum Flughafen) als generische Variablen an:

$$V_{ni} = \beta_1 K_{ni} + \beta_2 T_{ni} + \sum_{m=1}^{I-1} \beta_{2+m} \delta_{im}. \quad (4.38)$$

Gesucht ist die Preiselastizität der Nachfrage nach Flügen beim Flughafen i , wenn Flughafen j seinen Preis ändert.

^aDer Index n bezieht sich dabei primär auf die Kundenklasse und (bei Auswahlmöglichkeit) auf die Airline, wodurch (über den Kundenwunsch) auch indirekt das Flugziel fest ist. Da die Preise auch von der Flugklasse (First Class, Business Class, Economy) und wegen aktivem *Revenue Management* auch vom Buchungszeitpunkt und ggf individueller Vergünstigungen abhängen, kann eine Kundenklasse n durchaus nur einen einzigen Kunden ($y_n = \sum_i y_{ni} = 1$) enthalten.

4.5.1 Definitionen im Rahmen der diskreten Wahltheorie

Obiges Beispiel zeigt, dass die Definitionsmöglichkeiten von Elastizitäten im Rahmen der diskreten Wahltheorie viel reichhaltiger sind als bei den Regressionsmodellen. Dies wird durch folgende Eigenheiten der Modelle der diskreten Wahltheorie verursacht:

4 Modelle diskreter Entscheidungen

- sie sind *mikroskopisch*,
- sie sind *Mehrgleichungsmodelle*
- und sie betreffen im Allgemeinen ein *nichtteilbares Gut oder Dienstleistung*, was in der diskreten Skalierung der exogenen Variablen $y_{ni} \in \mathbb{N}_0^+$ zum Ausdruck kommt.

Wir unterscheiden drei verschiedene Kategorien, Elastizitäten zu definieren:

1. Die **Elastizität bezüglich Verschiebungen** bzw. die **Substitutions-Elastizität** (engl.: *mode-choice elasticity*) gegenüber der **vollen Elastizität** (engl.: *ordinary elasticity*). Bei der Verschiebungselastizität ist die Entscheidung (“Flug nach Barcelona”) als solches gefallen und man wählt nur den Modus (das “Wie”, also hier, von welchem Flughafen). Solche Substitutionselastizitäten sind vor allem bei *nichtteilbaren Gütern und Dienstleistungen*, wie eben Flügen, relevant. Bei den Gesamt-Elastizitäten hingegen steht als zusätzliche Option ein Flugverzicht bzw. Fahrt mit Auto oder Bahn (*no-choice* Option) zur Debatte. Ferner könne auch bisherige Nichtflieger durch billigere Preise zu Barcelona-Fliegern werden. Unter einigen, oft allerdings nicht erfüllten Annahmen kann man nach *Taplin* einen Zusammenhang zwischen diesen Elastizitäten finden.¹⁷
2. Die **mikroskopische Elastizität**, welche auf eine bestimmte Person und einen bestimmten Flug n bezogen ist, gegenüber der **makroskopischen Elastizität** bezüglich der gesamten Kundschaft und aller angebotenen Flüge. Die Definition einer mikroskopischen Elastizität wird dadurch ermöglicht, dass die Modelle der diskreten Wahltheorie mikroskopische Modelle sind, also *auf Einzelpersonbasis* Auswahrscheinlichkeiten ausgeben.
3. **Eigenelastizitäten** ϵ_{ii} als relative Änderung der Nachfrage bei Flughafen i , wenn sich Flugpreise von/zu diesem Flughafen ändern und die Preise bei allen anderen Flughäfen konstant bleiben, im Gegensatz zu **Kreuzelastizitäten** ϵ_{ij} : Relative Nachfrageänderung bei Flughafen i als Folge von Preisänderungen bei einem anderen Flughafen $j \neq i$.

¹⁷David A. Hensher, Kenneth John Button, “Handbook of transport modelling” (2000); Oum, T.H. and Waters, WG and Yong, J.S., “Concepts of price elasticities of transport demand and recent empirical estimates: an interpretative survey”, *Journal of transport economics and policy* 26, 139–154 (1992).

4 Modelle diskreter Entscheidungen

Verständnisfrage:

Welche Elastizität ist betragsmäßig größer: Die Verschiebungs- oder die volle Elastizität? Wie groß sind Verschiebungs- und volle Elastizitäten, falls es (neben der No-Choice-Option) nur eine Wahlmöglichkeit ($I = 1$) gibt?

Die volle Elastizität ist betragsmäßig größer, da sie auch Entscheidungsänderung "von außen" bzw. "nach außen" mit berücksichtigt. Bei $I = 1$ (Monopol) ist die Verschiebungselastizität $=0$.

4.5.2 Mikroskopische Preiselastizitäten

Sowohl mikroskopische Verschiebungs- als auch volle Elastizitäten können im Rahmen der Wahltheorie modelliert werden: Ordnet man in der Alternativenmenge der No-Choice-Option einen prohibitiv negativen Nutzen zu, modelliert man Verschiebungselastizitäten.¹⁸ Nimmt man jedoch für diese Option eine zu schätzenden alternativen-spezifischen Konstante an, beschreibt man die volle Elastizität.

Berechnung für das MNL bei parameterlinearen Nutzen

Parameterlineare deterministische Nutzenfunktionen lassen sich allgemein als Summe von M exogenen Faktoren darstellen:

$$V_{ni} = \beta' \mathbf{x}_{ni} = \sum_{m=1}^M \beta_m x_{nmi}. \quad (4.39)$$

Hierbei bezeichnet x_{nmi} den m -ten exogenen Faktor für Person bzw. Personengruppe n und Alternative i .

Veranschaulichungs-Beispiel: Flughäfen

Hier bezeichnet $x_{1n1} = K_{n1}$ den Flugpreis von Flughafen MUC für Person n , $x_{2n3} = T_{n3}$ die Reisezeit von der Wohnung dieser Person zum Flughafen FRA usw.

Mikroskopische Substitutionselastizitäten bezüglich des Faktors m sind allgemein definiert durch

$$\epsilon_{nij}^{(\text{mic},m)} = \frac{x_{nmj}}{P_{ni}} \frac{\partial P_{ni}}{\partial x_{nmj}} \quad (4.40)$$

oder in Worten:

¹⁸Erinnern Sie sich, dass die Alternativenmenge immer gegenseitig exklusiv und vollständig sein muss.

4 Modelle diskreter Entscheidungen

Die mikroskopische Substitutionselastizität $\epsilon_{nij}^{(\text{mic},m)}$ bezeichnet die prozentuale Änderung der Auswahlwahrscheinlichkeit bzw. erwartete Nachfrage von Person (bzw. Personengruppe) n für Alternative i , wenn sich der exogene Faktor m der Alternative j für diese Person um 1% ändert.

Die MNL-Auswahlwahrscheinlichkeit ist wie üblich gegeben durch

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{k=1}^I e^{V_{nk}}}.$$

Ihrer partielle Ableitung nach dem jeweiligen exogenen Faktor lässt sich durch Anwendung der Produkt- und Kettenregeln berechnen:

$$\begin{aligned} \frac{\partial P_{ni}}{\partial x_{nmj}} &= \frac{e^{V_{ni}}}{\sum_k e^{V_{nk}}} \frac{\partial V_{ni}}{\partial x_{nmj}} - \frac{e^{V_{ni}}}{(\sum_k e^{V_{nk}})^2} \frac{\partial}{\partial x_{nmj}} \left(\sum_k e^{V_{nk}} \right) \\ &= P_{ni} \frac{\partial V_{ni}}{\partial x_{nmj}} - \sum_k \frac{e^{V_{ni}} e^{V_{nk}}}{(\sum_k e^{V_{nk}})^2} \frac{\partial V_{nk}}{\partial x_{nmj}} \\ &= P_{ni} \left(\frac{\partial V_{ni}}{\partial x_{nmj}} - \sum_k P_{nk} \frac{\partial V_{nk}}{\partial x_{nmj}} \right). \end{aligned}$$

Mit (4.39) ergeben sich die partiellen Ableitungen zu

$$\frac{\partial V_{ni}}{\partial x_{nmj}} = \beta_m \delta_{ij} = \beta_m \begin{cases} 1 & i = j \\ 0 & \text{sonst,} \end{cases} \quad \frac{\partial V_{nk}}{\partial x_{nmj}} = \beta_m \delta_{kj}$$

und damit

$$\frac{\partial P_{ni}}{\partial x_{nmj}} = \beta_m P_{ni} (\delta_{ij} - P_{nj}). \quad (4.41)$$

Mit der Definition (4.40) sind die mikroskopischen Elastizitäten des MNL bei parameterlinearen deterministischen Nutzenfunktionen schließlich gegeben durch

$$\epsilon_{nij}^{(\text{mic},m)} = \beta_m x_{nmj} (\delta_{ij} - P_{nj}). \quad (4.42)$$

Damit ergeben sich

- die mikroskopischen Eigenelastizitäten

$$\epsilon_{nii}^{(\text{mic},m)} = \beta_m x_{nmi} (1 - P_{ni}), \quad (4.43)$$

- die mikroskopischen Kreuzelastizitäten:

$$\epsilon_{nij}^{(\text{mic},m)} = -\beta_m x_{nmj} P_{nj}. \quad (4.44)$$

4 Modelle diskreter Entscheidungen

Aufgrund der IIA-Eigenschaft hängen die Kreuzelastizitäten *nicht* vom Ziel i ab, d.h. alle Alternativen i sind gleichermaßen von einer Änderung der Attribute von j betroffen.

Aufgabe: Substitutionselastizität beschreibt Nullsummenspiel

Zeigen Sie, dass allgemein folgende Summenbedingung gilt:

$$\sum_i P_{ni} \epsilon_{nij}^{(\text{mic},m)} = 0,$$

indem Sie beispielsweise $j = 1$ setzen (bei Alternative 1 wird das Attribut m geändert). Zeigen Sie, dass hinter dieser Beziehung nichts anderes als ein Nullsummenspiel steht: Was einer gewinnt, wird dem bzw. den anderen genommen.

Veranschaulichungs-Beispiel: Flughäfen

Bei den deterministischen Nutzenfunktionen der Flughäfen, Gl. (4.38), haben die Preiselastizitäten folgende Bedeutung:

- Erhöht sich der Preis an Flughafen i um 1%, dann erhöht sich die Wahrscheinlichkeit für Person n , diesen Flughafen zu wählen, um $\epsilon_{nii}^{(\text{mic},K)} = \beta_1 K_{ni} (1 - P_{ni})$ Prozent. Da die Preissensitivität $\beta_1 < 0$ und $1 - P_{ni} > 0$ sind, wird die Wahrscheinlichkeit bei Erhöhung des Preises *reduziert*.
- Erhöht sich der Preis an Flughafen j um 1%, dann erhöht sich die Wahrscheinlichkeit für Person n , Flughafen $i \neq j$ zu wählen, um $\epsilon_{nij}^{(\text{mic},K)} = -\beta_1 K_{nj} P_{nj}$ Prozent. Da die Preise und Wahrscheinlichkeiten positiv sind und $\beta_1 < 0$ ist, führt eine Preiserhöhung bei Flughafen j zu einer prozentualen *Zunahme* der Wahl aller Flughäfen $i \neq j$. Alle Konkurrenzflughäfen i erhalten dabei die gleiche relative Nachfragesteigerung.

4.5.3 Makroskopische Preiselastizitäten

Zunächst ist die makroskopische Elastizität weder gleich der mit dem Wahlmodell wie oben berechnete Elastizität für mittlere exogene Faktoren $\bar{x}_{mi} = 1/N \sum_n x_{nmi}$ noch ist sie ein direkter Mittelwert der mikroskopischen Elastizitäten. Vielmehr muss man bei der Aggregation die persönlichen Einflussfaktoren und Wahlwahrscheinlichkeiten individuell betrachten.

4 Modelle diskreter Entscheidungen

Allgemein ist die makroskopische Elastizität bezüglich des Faktors m definiert durch

$$\epsilon_{ij}^{(\text{mac},m)} = \frac{X_{mj}}{N_i} \frac{\partial N_i}{\partial X_{mj}}. \quad (4.45)$$

Die makroskopischen Merkmalssummenößen (gekennzeichnet durch Großbuchstaben) bedeuten:

- $N_i = \sum_n P_{ni}$ die vom Modell geschätzte Wahlhäufigkeit für Alternative i ,
- $X_{mj} = \sum_n x_{nmj}$ die potenzielle (maximale) Merkmalssumme des Merkmals bzw. Faktors m bei Alternative j , wenn alle diese Alternative wählen, bzw. der Mittelwert $1/N \sum_n x_{nmj}$ des exogenen Faktors über alle Personen (der Unterschied kürzt sich bei der Definition (4.45) heraus),
- $\frac{\partial N_i}{\partial X_{mj}}$ die Änderung der Wahlhäufigkeit der Alternative i bei Änderung des Mittelwertes des Faktors m der Alternative j .

In Worten:

Die makroskopische Substitutionselastizität $\epsilon_{ij}^{(\text{mac},m)}$ bezeichnet die prozentuale Änderung der Wahlhäufigkeit bzw. Nachfrage N_i für Alternative i , wenn sich bei Alternative j der über alle Personen gemittelte exogene Faktor m um 1 % ändert. Ist m beispielsweise der Preis, gibt $\epsilon_{ij}^{(\text{mac},m)}$ die prozentuale Erhöhung der Gesamtnachfrage nach i an, wenn sich die Preise bei j im Mittel um 1 % erhöhen.

Diese Definition ist jedoch nicht eindeutig, da sie offen lässt, wie sich die Änderungen der Merkmalssumme auf die einzelnen Personen aufteilt. Es gibt zwei offensichtliche Aufteilungsmöglichkeiten:

- **Absolute Änderungen** um einen festen Betrag:

$$dx_{nmj} = dX_{mj} \quad (4.46)$$

Hier ändern sich die Attribute bei allen Personen $n = 1, \dots, N$ um denselben *absoluten* Betrag, welcher natürlich dann gleich der Änderung dX_{mj} des Mittelwertes ist.

- **Relative Änderungen:**

$$\frac{dx_{nmj}}{x_{nmj}} = \frac{dX_{mj}}{X_{mj}} \quad (4.47)$$

Hier ändern sich die Attribute bei allen Personen um denselben *relativen* Betrag, welcher dann natürlich gleich der relativen Änderung des Mittelwertes über alle Personen ist.

4 Modelle diskreter Entscheidungen

Formulierung als Funktion der mikroskopischen Elastizitäten

Bei den **absolute Änderungen** gilt mit der Definition (4.40) der makroskopischen Elastizität sowie der Aufteilungsregel (4.46)

$$\frac{\partial N_i}{\partial X_{mj}} = \sum_n \frac{\partial P_{ni}}{\partial X_{mj}} = \sum_n \frac{\partial P_{ni}}{\partial x_{nmj}} = \sum_n \frac{P_{ni}}{x_{nmj}} \epsilon_{nij}^{(\text{mic,m})}$$

und damit mit der Definition (4.45) der makroskopischen Elastizität:

$$\epsilon_{ij}^{(\text{mac,abs,m})} = \frac{X_{mj}}{N_i} \frac{\partial N_i}{\partial X_{mj}} = \frac{X_{mj}}{N_i} \sum_n \frac{P_{ni}}{x_{nmj}} \epsilon_{nij}^{(\text{mic,m})}. \quad (4.48)$$

Bei den **relativen Änderungen** verwendet man hingegen die Aufteilungsregel (4.47) und erhält

$$\frac{\partial N_i}{\partial X_{mj}} = \sum_n \frac{\partial P_{ni}}{\partial X_{mj}} = \sum_n \frac{x_{nmi}}{X_{mj}} \frac{\partial P_{ni}}{\partial x_{nmj}} = \sum_n \frac{P_{ni}}{X_{mj}} \epsilon_{nij}^{(\text{mic,m})}$$

und damit wieder mit der Definition (4.45):

$$\epsilon_{ij}^{(\text{mac,rel,m})} = \frac{1}{N_i} \sum_n P_{ni} \epsilon_{nij}^{(\text{mic,m})}. \quad (4.49)$$

Mit der Gewichtung

$$w_{ni} = \frac{P_{ni}}{N_i} = \frac{P_{ni}}{\sum_n P_{ni}}$$

kann man dies auch schreiben als

$$\epsilon_{ij}^{(\text{mac,rel,m})} = \sum_n w_{ni} \epsilon_{nij}^{(\text{mic,m})}. \quad (4.50)$$

Dies ist ein gewichtetes Mittel der relativen Elastizitäten, wobei die Gewichtung proportional des erwarteten Anteils der Person n an der (bisherigen) Gesamtnachfrage N_i für Alternative i ist.

Veranschaulichungs-Beispiel: Flughäfen

- Die Elastizität $\epsilon_{ij}^{(\text{mac,abs,1})}$ bezeichnet die prozentuale Änderung der Gesamtnachfrage an Flügen von Flughafen i , wenn Flughafen j den Preis (=Faktor bzw. Attribut 1) für alle Flüge um einen gleichen festen Betrag ändert, so dass sich der Gesamtpreis aller angebotenen Flüge um 1 % ändert.
- Die Elastizität $\epsilon_{ij}^{(\text{mac,rel,1})}$ bezeichnet die prozentuale Änderung der Gesamtnachfrage an Flughafen i , wenn Flughafen j auf alle Preise 1 % aufschlägt. Diese Elastizität ist gleich dem gewichteten Mittel der mikroskopischen Elastizitäten, wobei die Wichtung gleich dem erwarteten Anteil des Fluggastes n an der Gesamtnachfrage des Flughafens i ist.

4 Modelle diskreter Entscheidungen

4.5.4 Exkurs: Revenue Management

Siehe [README_RevenueManagementNeuBringen_14](#)

4.6 Einschub: Die Maximum-Likelihood-Methode

Zur Parameterschätzung bzw. Modellkalibrierung kam bei Regressionsmodellen die Least-Squares-Errors (LSE) Methode zum Einsatz. Diese ist jedoch bei den Modellen der diskreten Wahltheorie nicht immer wohl definiert, geschweige denn, dass sie zu unverzerrten oder gar effizienten Schätzern führt. Das Problem ist die Definition der Fehlerquadratsumme: Da die Wahltheorie-Modelle, im Gegensatz zu den Regressionsmodellen, Wahrscheinlichkeitsaussagen machen (bei den Regressionsmodellen ist ϵ ja nur der nicht-modellierte Restanteil), ist eine direkte Bildung der Art “Summe der quadrierten Abweichungen zwischen modellierten und beobachteten Wert” nicht möglich. Allenfalls könnte man versuchen, die modellierten Wahrscheinlichkeiten mit den beobachteten relativen Häufigkeiten möglichst gut in Übereinklang zu bringen. Das führt jedoch im Allgemeinen nicht auf effiziente oder selbst unverzerrte Schätzer.¹⁹

Glücklicherweise gibt es zum Schätzen von Wahlmodellen ein alternatives Verfahren, die **Maximum-Likelihood-Methode**. Bei dieser, oft auch als ML abgekürzte Methode werden die Beobachtungen

$$y_{ni} = \begin{cases} 1 & \text{Person } n \text{ hat Alternative } i \text{ gewählt} \\ 0 & \text{sonst,} \end{cases} \quad (4.51)$$

also die endogenen Variablen, als Realisierungen des stochastischen Modells gesehen, welches durch die Alternativenzahl definiert ist: Dabei betrachten wir den Fall, dass jede Person n nur genau eine Alternative wählen kann,²⁰ also

$$y_n = \sum_{i=1}^I y_{ni} = 1. \quad (4.52)$$

Bei gegebenen Auswahlwahrscheinlichkeiten P_{ni} ergibt das im Modell bei zwei Alternativen die $(1, P_{n1})$ - Binomialverteilung. Der erste Parameter besagt, dass es sich nur um eine Ja-Nein-Entscheidung (*Bernoulli-Experiment*) handelt, während der zweite Parameter die Wahrscheinlichkeit für die Wahl der Alternative 1 (“ja”) angibt. Bei mehr als zwei Alternativen verallgemeinert man die Binomialverteilung durch eine durch eine durch die Wahrscheinlichkeiten $P_{n1}, \dots, P_{n,I-1}$ charakterisierte *Multinomialverteilung*.²¹ Falls, wie hier, nur jeweils eine Entscheidung betrachtet wird ($\sum_i y_{ni} = 1$), ergeben sowohl die

¹⁹Die Kalibrierungsbedingungen (4.72) des Logitmodells könnte man jedoch so interpretieren.

²⁰Falls es für eine Person mehrere Wahl-Situationen gibt, oder eine homogene Personengruppe, welche als eine Einheit betrachtet wird, wiederholt man die Wahlentscheidung entsprechend häufig. Man geht also in der Modellierung, zumindest konzeptionell, auf jedem Fall mikroskopisch vor.

²¹Wegen $\sum_i P_{ni} = 1$ ist jeweils die letzte Wahrscheinlichkeit P_{nI} der Binomial- und Multinomialverteilung nicht unabhängig und deshalb kein Parameter.

4 Modelle diskreter Entscheidungen

Binomial- als auch die Multinomialverteilung mit der Wahrscheinlichkeit $P_{ni_n}(\boldsymbol{\beta})$ den "richtigen" Vektor

$$\mathbf{y}_n = (0, \dots, 0, 1, 0, \dots, 0)',$$

wobei die 1 an der Position i_n der tatsächlich bei Wahl n gewählten Alternative steht (vgl. Gl. (4.60) im nächsten Abschnitt).

Nun wird, in Abhängigkeit des Parametervektors $\boldsymbol{\beta}$, die kombinierte Wahrscheinlichkeit dafür berechnet, dass die Messung 1 den Wert \mathbf{y}_1, \dots , die Messung N den Wert \mathbf{y}_N liefert:

$$L(\boldsymbol{\beta}) = \text{Prob}(\mathbf{Y}_1(\boldsymbol{\beta}) = \mathbf{y}_1, \mathbf{Y}_2(\boldsymbol{\beta}) = \mathbf{y}_2, \dots, \mathbf{Y}_N(\boldsymbol{\beta}) = \mathbf{y}_N) = \max(\boldsymbol{\beta}). \quad (4.53)$$

Bei festen Messdaten, aber in Abhängigkeit der Parameter, wird diese Funktion **Likelihoodfunktion** genannt.²² Die im Parametervektor $\boldsymbol{\beta}$ enthaltenen Parameter werden so bestimmt, dass die *Likelihood*, also die Wahrscheinlichkeit dafür, dass das stochastische Modell die Messungen liefert, maximal ist.

Die Anwendungsgebiete der LSE und der ML-Methode überschneiden sich, da man letztere auch auf stetige endogene Variable anwenden kann.²³ Es gilt:

- Die LSE-Methode kann man bei stetigen endogenen Variablen und beliebiger, auch nicht explizit vorgegebener, Verteilung der Zufallsgrößen anwenden.
- Die ML-Methode kann man bei stetigen und diskreten endogenen Variablen anwenden, aber die Form der Verteilung der Zufallsgrößen muss explizit gegeben sein.

4.6.1 Regressionsmodelle: Vergleich mit der LSE-Methode

Als erstes Beispiel wenden wir die ML-Methode auf das vollständig nach Gauß-Markow spezifizierte lineare multivariate Regressionsmodell (2.2) an.²⁴ Das Modell lautet also

$$Y(\mathbf{x}; \boldsymbol{\beta}) = \sum_{m=0}^M \beta_m x_m + \epsilon = \boldsymbol{\beta} \mathbf{x} + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

²²Im Englischen wird unterschieden zwischen *probability* (Wahrscheinlichkeit, dass eine Zufallsgröße mit fester Verteilung, also festen Parametern, einen bestimmten Wert liefert) und *Likelihood* (Wahrscheinlichkeit, mit der in Abhängigkeit der Parameter eine Zufallsgröße einen vorgegebenen Messwert trifft). Insbesondere ist bei diskreten Zufallsgrößen die Wahrscheinlichkeitsfunktion diskret, während die *Likelihoodfunktion* immer kontinuierlich ist: Im Gegensatz zu den exogenen oder endogenen Variablen werden die Modellparameter nämlich immer kontinuierlich angenommen.

²³Bei stetigen Größen ist die Wahrscheinlichkeit des Eintreffens eines festen Wertes i. A. exakt gleich Null. Deshalb wird in diesem Fall, genau genommen, die *Wahrscheinlichkeitsdichte* maximiert. Die ML-Schätzer entsprechen also der Stelle, bei welcher die Dichte der kombinierten Wahrscheinlichkeitsfunktion an der Stelle der gemessenen endogenen Variablen bezüglich der Parameter maximal ist. Bezüglich der Parameter, also als Likelihoodfunktion, ist die Verteilung aber in beiden Fällen stetig, so dass eine Unterscheidung nicht nötig ist.

²⁴In diesem Beispiel wird wieder die Indexkonvention der Regressionsrechnung verwendet, also i für den i -ten Satz an Messwerten (\mathbf{x}_i, y_i) .

4 Modelle diskreter Entscheidungen

und die dazugehörigen Systemgleichungen

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{mit} \quad \boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_i, \dots, \epsilon_n)', \quad \epsilon_i \sim \text{i.i.d } N(0, \sigma_\epsilon^2).$$

Die Wahrscheinlichkeit, dass das Modell die Messung i beschreibt, oder genauer, die Wahrscheinlichkeitsdichte des Modells mit den exogenen Variablenvektor \mathbf{x}_i an der Stelle y_i lautet

$$f_i(y_i) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left[-\frac{(y_i - \boldsymbol{\beta}\mathbf{x}_i)^2}{2\sigma_\epsilon^2} \right].$$

Da die Zufallsgröße ϵ_i der Messung i von den Zufallsgrößen anderer Messungen unabhängig ist (i.i.d-Eigenschaft), ist die Likelihooddichte des verbundenen Ereignisses, dass das Modell *alle* Messungen y_i , $i = 1, \dots, n$ beschreibt, durch das Produkt gegeben:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left[-\frac{(y_i - \boldsymbol{\beta}\mathbf{x}_i)^2}{2\sigma_\epsilon^2} \right]. \quad (4.54)$$

Das Rechnen und Ableiten mit Produkten ist umständlich. Deshalb berücksichtigt man, dass sich die Stelle $\hat{\boldsymbol{\beta}}$, an der die Likelihood-Dichte ihr Maximum aufweist, durch Anwenden einer streng monoton steigenden Funktion auf L nicht ändert. Insbesondere ist der Logarithmus eine solche Funktion und er erweist sich als günstig, da er Produkte in Summen umwandelt. Dies ergibt hier die **Log-Likelihood**(dichte)

$$\begin{aligned} \tilde{L}(\boldsymbol{\beta}) &= \ln L(\boldsymbol{\beta}) \\ &= \sum_{i=1}^n \left\{ -\frac{1}{2}(\ln 2\pi + \ln \sigma_\epsilon^2) - \left[\frac{(y_i - \boldsymbol{\beta}\mathbf{x}_i)^2}{2\sigma_\epsilon^2} \right] \right\} \\ &= -\frac{n}{2}(\ln 2\pi + \ln \sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned} \quad (4.55)$$

Wie der Name schon sagt, maximiert man bei der ML-Methode die *Likelihood*. Als notwendige (und hier hinreichende) Bedingungen muss der Gradient, d.h. alle partiellen Ableitungen, verschwinden. Mit den Ableitungsregeln in Kap. 2.11.3 ergibt dies

$$\frac{\partial \tilde{L}}{\partial \boldsymbol{\beta}} = \frac{1}{2\sigma_\epsilon^2}(2\mathbf{X}'\mathbf{y} - 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) \stackrel{!}{=} 0 \quad \Rightarrow \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (4.56)$$

Man erhält also dasselbe Ergebnis wie bei der LSE-Methode! Ableiten nach σ_ϵ^2 (nicht nach σ_ϵ) liefert

$$\frac{\partial \tilde{L}}{\partial \sigma_\epsilon^2} = -\frac{n}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \stackrel{!}{=} 0$$

und daraus

$$\sigma_\epsilon^2 = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n} = \frac{S}{n}$$

mit der LSE-Fehlerquadratsumme S . Setzt man (4.56) ein, erhält man

$$\hat{\sigma}_\epsilon^2 = \frac{S_{\min}}{n}. \quad (4.57)$$

4 Modelle diskreter Entscheidungen

Dies ist jedoch *kein* unverzerrter Schätzer der Residualvarianz, da dieser nach (2.24) $(n - M - 1)$ und nicht n im Nenner hat. Der Schätzer $\hat{\beta}$ selbst ist jedoch unverzerrt (erwartungstreu), da identisch mit dem der LSE-Schätzung. Wir halten fest:

Beim linearen oder quasilinearen Modell mit i.i.d. gaußverteilten Störgrößen ϵ liefert die ML-Methode für die Parameter β exakt denselben Schätzer wie die LSE-Methode. Der ML-Schätzer der Residualvarianz ist jedoch verzerrt (aber immerhin konsistent).

Schließlich wird noch begründet, warum die Dichte der mehrdimensionalen Gaußverteilung (2.47) der Schwankungen der Parameterschätzer auch “Likelihood-Dichte” genannt wird: Spaltet man β in den Schätzer $\hat{\beta}$ und der Abweichung $\Delta\hat{\beta} = \beta - \hat{\beta}$ des wahren Wertes vom Schätzer auf und setzt dies in die Log-Likelihood-Dichte (4.55) ein, so ergibt sich

$$\tilde{L}(\Delta\hat{\beta}) = \text{const.} - \frac{1}{2\sigma_\epsilon^2} \left[\mathbf{y}^2 - \mathbf{y}\mathbf{X}(\hat{\beta} + \Delta\hat{\beta}) - (\hat{\beta}' + \Delta\hat{\beta}')\mathbf{X}'\mathbf{y} + (\hat{\beta}' + \Delta\hat{\beta}')\mathbf{X}'\mathbf{X}(\hat{\beta} + \Delta\hat{\beta}) \right]$$

Nach geschicktem Umformen unter Verwendung von $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ergibt sich

$$\tilde{L}(\Delta\hat{\beta}) = \text{const.} - \frac{1}{2\sigma_\epsilon^2} \Delta\hat{\beta}'\mathbf{X}'\mathbf{X}\Delta\hat{\beta} \quad (4.58)$$

Dies ist genau der Logarithmus von (2.47)! Damit ist die Bezeichnung “Likelihood-Dichte” für die multivariate Dichtefunktion der Parameterschätzer gerechtfertigt. *Ein- und dieselbe Verteilung* bezeichnet also sowohl die Schwankungen der Schätzer um den wahren Wert, als auch (mit anderer Normierung) die Wahrscheinlichkeitsdichte (bezüglich des Parametervektors) dafür, dass das Modell mit den geschätzten Parametern die Daten erklären kann!

4.6.2 Anwendung auf das einfachste Wahlmodell

Ehe im nächsten Abschnitt die eigentlichen Modelle der diskreten Wahltheorie mit der ML-Methode geschätzt werden, erfolgt nun “zum Kennenlernen” eine Anwendung auf das wohl einfachstmögliche Modell diskreter Entscheidungen: N Personen antworten unabhängig voneinander eine Ja-Nein Frage. Ordnet man der Antwort “Ja” die Alternative 1 und “Nein” Alternative 2 zu, ergeben sich $N_1 = \sum_n y_{n1} \leq N$ Ja-Antworten. Wie das “Trivialmodell” der linearen Regression hat das Modell keine exogenen Variablen, d.h., alle befragten Personen kommen aus einer verhaltenshomogenen Gruppe. Damit gilt $P_{n1} = P_1$ für alle Personen n und wir können anstelle eines expliziten Wahlmodells direkt die $(1, P_1)$ -Binomialverteilung als “Modell” heranziehen. Dieses Modell liefert für jede befragte Person die endogene Variable y_{n1} (0: “nein”; 1: “ja”):

$$Y_{ni} = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } P_1 \\ 0 & \text{sonst.} \end{cases} \quad (4.59)$$

4 Modelle diskreter Entscheidungen

Da genau eine Alternative gewählt wird, ist $Y_{n2} = 1 - Y_{n1}$ abhängig und braucht nicht separat betrachtet werden: Wenn die Beobachtung y_{n1} richtig vorausgesagt wurde, dann auch y_{n2} . Die Anwendung des Modells auf N Personen liefert für die Gesamtzahl $N_1 = \sum_n y_{n1}$ an Ja-Antworten die (N, P_1) -Binomialverteilung, da eine Summe von unabhängigen $(1, P_1)$ binomialverteilten Zufallsgrößen (N, P_1) binomialverteilt ist. Damit ergibt sich folgende Likelihoodfunktion:

$$L(P_1) = \binom{N}{N_1} P_1^{N_1} (1 - P_1)^{N - N_1}$$

und damit die Log-Likelihood

$$\tilde{L}(P_1) = \ln L(P_1) = \ln \binom{N}{N_1} + N_1 \ln P_1 + (N - N_1) \ln(1 - P_1).$$

Ableiten und Nullsetzen liefert

$$\frac{d\tilde{L}(P_1)}{dP_1} = \frac{N_1}{P_1} + \frac{N - N_1}{1 - P_1} (-1) \stackrel{!}{=} 0,$$

also

$$N_1(1 - P_1) = P_1(N - N_1) \Rightarrow P_1 = \frac{N_1}{N}.$$

Der nach der ML-Methode geschätzte Parameter \hat{P}_1 ist also gleich der relativen Häufigkeit der erhaltenen Ja-Antworten. Anders gesagt: Die modellierte Wahrscheinlichkeit für "Ja" ist gleich der beobachteten relativen Häufigkeit für "Ja".

Falls man diesen einfachen Wahlprozess nicht direkt mit der Binomialverteilung sondern z.B. mit einem Logitmodell formuliert, erhält man, wegen des Fehlens exogener Variablen, für die Nutzenfunktionen nur einen vollständigen Satz an alternativenspezifischen Konstanten, der im binärem Fall wie hier nur einen einzigen Parameter enthält:

$$V_{n1} = \beta_1, \quad V_{n2} = 0.$$

Daraus ergibt sich $P_1(\beta_1) = \frac{e^{\beta_1}}{1 + e^{\beta_1}}$ und, durch Umstellung nach β_1 und Einsetzen von $P_1 = N_1/N$ der ML-Schätzer

$$\hat{\beta}_1 = \ln \left(\frac{P_1}{1 - P_1} \right) = \ln \left(\frac{N_1}{N - N_1} \right).$$

Im nächsten Abschnitt wird dies für nichttriviale Nutzenfunktionen systematisch untersucht.

4.7 Parameterschätzung mit der Maximum-Likelihood-Methode

In diesem Abschnitt wird die ML-Methode zunächst auf allgemeine Modelle der diskreten Wahltheorie angewandt und dann die Kalibrierung des Multinomial-Logit (MNL) Modells als dem praktikabelsten Modell vertieft behandelt. Folgende Informationen bzw. Daten sind für die Kalibrierung notwendig:

4 Modelle diskreter Entscheidungen

1. Die **Ergebnisse der empirischen Befragung** bzw. der realisierten Entscheidungen als absolute Häufigkeiten y_{ni} mit $\sum_i y_{ni} = 1$, bzw. die bei jeder Entscheidung n gewählte Alternative i_n ,
2. die **modellierten Entscheidungswahrscheinlichkeiten** $P_{ni}(\boldsymbol{\beta})$ dafür, dass bei Entscheidung n die Alternative i gewählt wird und insbesondere die Wahrscheinlichkeit $P_{ni_n}(\boldsymbol{\beta})$ der tatsächlich gewählten Alternative.

Wegen der Exklusivitätseigenschaft der Alternativen (nur ein $y_{ni_n} = 1$, alle anderen $y_{ni} = 0$ für $i \neq i_n$) ist die Wahrscheinlichkeit, dass der vom Modell bei Entscheidung n vorausgesagte Entscheidungsvektor gleich dem tatsächlichen ist, sowohl für Binomial- als auch für Multinomialmodelle direkt durch die Auswahlwahrscheinlichkeit für die tatsächlich gewählte Alternative $i = i_n$ gegeben:

$$\text{Prob}(\hat{\mathbf{Y}}_n = \mathbf{y}_n) = \text{Prob}(\hat{Y}_{n1} = y_{n1}, \dots, \hat{Y}_{nI} = y_{nI}) = \prod_{i=1}^I [P_{ni}(\boldsymbol{\beta})]^{y_{ni}} = P_{ni_n}(\boldsymbol{\beta}). \quad (4.60)$$

Während der rechte Ausdruck dieser Formel der einfachste ist, eignet sich der Ausdruck davor mit dem Produkt besser für die "Weiterverarbeitung". Zudem ist dieser Ausdruck, bis auf einen nicht von $\boldsymbol{\beta}$ abhängigen und irrelevanten *Multinomialkoeffizienten* $(\sum_i y_{ni})! / (\prod_i (y_{ni})!)$ als Vorfaktor, auch für den Fall $\sum_i y_{ni} > 1$ gültig. Damit können mehrere Beobachtungen bzw. Entscheidungen zugunsten derselben Alternative zusammengefasst werden, wenn jeweils auch alle exogenen Faktoren gleich sind.

4.7.1 Maximum-Likelihood Funktionen

Bei den Modellen der diskreten Wahltheorie kann man die allgemeine Bedingung (4.53) folgendermaßen formulieren:

$$\hat{\boldsymbol{\beta}} = \arg \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \arg \max_{\boldsymbol{\beta}} [R_{\text{komb}}(\mathbf{Y}_1(\boldsymbol{\beta}) = \mathbf{y}_1, \dots, \mathbf{Y}_N(\boldsymbol{\beta}) = \mathbf{y}_N)]. \quad (4.61)$$

Der Parametervektor $\boldsymbol{\beta}$ wird also so geschätzt (Schätzwert $\hat{\boldsymbol{\beta}}$), dass die *kombinierte Wahrscheinlichkeit* R_{komb} dafür, dass alle prognostizierten absoluten Häufigkeiten $\hat{Y}_{ni}(\boldsymbol{\beta})$ gleich den beobachteten Häufigkeiten y_{ni} sind, maximal wird.

Bei MNL-Modellen oder allgemein bei von n unabhängigen Zufallsnutzen läuft dies auf eine Multiplikation der einzelnen Multinomialverteilungen hinaus und resultiert mit (4.60) in folgender **Likelihoodfunktion**:

$$L(\boldsymbol{\beta}) = \prod_{n=1}^N \prod_{i=1}^I [P_{ni}(\boldsymbol{\beta})]^{y_{ni}} = \prod_{n=1}^N P_{ni_n}(\boldsymbol{\beta}), \quad (4.62)$$

wobei i_n die bei Person n beobachtete Wahlentscheidung aus den Daten ist. Wichtig ist, dass diese Wahrscheinlichkeit *bei festen, durch die Daten gegebenen Werten* y_{ni} als Funktion der Modellparameter $\boldsymbol{\beta}$ maximiert wird.

4 Modelle diskreter Entscheidungen

4.7.2 Log-Likelihood des allgemeinen Entscheidungsmodells

Mit (4.62) ergibt sich die Log-Likelihood des diskreten Wahlmodells zu

$$\tilde{L}(\boldsymbol{\beta}) = \ln L(\boldsymbol{\beta}) = \sum_{n=1}^N \sum_{i=1}^I y_{ni} \ln P_{ni}(\boldsymbol{\beta}) = \sum_{n=1}^N \ln P_{ni_n}(\boldsymbol{\beta}) \quad (4.63)$$

wobei i_n die bei Person n beobachtete Wahlentscheidung aus den Daten ist und man anstelle von L auch \tilde{L} bezüglich $\boldsymbol{\beta}$ maximieren kann.

4.7.3 Grafische Lösung

Durch direktes Plotten der Log-Likelihood kann man bereits grafisch die Parameter schätzen, wie nun anhand zweier Beispiele gezeigt wird.

Beispiel 1: Revealed-Choice-Befragung mit Reisezeiten

Erhebung: Es werden 30 Personen gefragt, mit welchen Verkehrsmittelkategorien sie am Befragungstag den jeweils ersten Weg von Zuhause zurückgelegt haben: Nichtmotorisiert zu Fuß/Rad (Alternative 1) oder mit den motorisierten Verkehrsmitteln ÖPNV/Kfz (Alternative 2). Zu jeder dieser Alternativen wurden auch die Haustür-zu-Haustür Reisezeiten des innerhalb der Alternativen jeweils bevorzugten Verkehrsmittels erfragt und in Tabelle 4.1 eingetragen. Zur einfacheren Darstellung wurden in dieser Tabelle die 30 Personen in 6 Gruppen zu jeweils 5 Personen mit gleichen Reisezeiten klassiert. Wie bereits erwähnt, gilt die Form (4.62) der Likelihoodfunktion bzw. der Ausdruck (4.63) der Log-Likelihood (4.63) bis auf konstante Multiplikatoren bzw. Summanden auch für den Fall $\sum_i y_i > 1$. Da solche Faktoren bzw. Summanden nicht die Stelle $\hat{\boldsymbol{\beta}}$ des Maximums verändern, kann man (4.63) direkt auf die Zeilen der Tabelle anwenden. Konzeptionell verläuft die Rechnung jedoch "mikroskopisch" für jede Person getrennt, es wird also eine Tabelle mit 30 statt 6 Zeilen (jede Person bzw. Entscheidung in einer eigenen Zeile) abgearbeitet.²⁵

Modellierung: Die komplexen Reisezeiten T_{ni} in Minuten werden als generische Variable (also nicht alternativenspezifisch) angesetzt. Außerdem gibt es einen vollständigen Satz von alternativenspezifischen Konstanten, der hier aus einem Element besteht, wobei als Referenz die Alternative 2 definiert wurde:

$$V_{n1}(\beta_1, \beta_2) = \beta_1 + \beta_2 T_{n1}, \quad V_{n2} = \beta_2 T_{n2}. \quad (4.64)$$

Die binomialen Probit- und Logitmodelle werden jeweils in ihrer standardisierten Form, also $\sigma_\epsilon = 1$ beim Probit- und $\lambda = 1$ beim Logitmodell, angesetzt. Beim Logitmodell werden die resultierenden Parameter aber mit den Faktor $\sqrt{6}/\pi$ multipliziert dargestellt, so dass in beiden Modellen die Nutzenfunktionen in Einheiten von σ_ϵ gemessen werden und die Parameter direkt vergleichbar werden (vgl. Abschnitt 4.4.3).

²⁵Die meisten Softwarepakete wie die Open-Source-Software *Biogeme* verlangen in der Tat, dass jede Entscheidung eine eigene Zeile in der Input-Datei enthält.

4 Modelle diskreter Entscheidungen

Personenklasse	Zeit Alternative 1 (min)	Zeit Alternative 2 (min)	Wahl Alt. 1	Wahl Alt. 2
1	15	30	3	2
2	10	15	2	3
3	20	20	1	4
4	30	25	1	4
5	30	20	0	5
6	60	30	0	5

Tabelle 4.1: Ein einfaches Beispiel einer *Revealed Choice*-Befragung

Analyse: Die aus der Befragung gewonnenen *empirischen* Häufigkeiten y_{ni} stehen in den beiden letzten Spalten der Tabelle. Für die Klasse 1 muss man also 3 Personen mit $y_{n1} = \delta_{i1}$ rechnen, für die restlichen beiden Personen dieser Klasse $y_{ni} = \delta_{i2}$ usw. Die *modellierten* Auswahlwahrscheinlichkeiten des Probit-Modells (4.22) und des Logit-Modells (4.31) mit (4.33) sind in den beiden Modellen gegeben durch

$$\begin{aligned}
 P_{n1}^{\text{Probit}}(\beta_1, \beta_2) &= \Phi\left(\frac{V_{n1}(\beta_1, \beta_2) - V_{n2}(\beta_1, \beta_2)}{\sqrt{2}}\right), \\
 P_{n1}^{\text{Logit}}(\beta_1, \beta_2) &= \frac{1}{1 + e^{-[V_1(\beta_1, \beta_2) - V_2(\beta_1, \beta_2)]}}, \\
 P_{n2}(\beta_1, \beta_2) &= 1 - P_{n1}(\beta_1, \beta_2).
 \end{aligned}$$

Damit sind die Angaben für die Berechnung der Log-Likelihoodfunktion (4.63) gegeben:

$$\tilde{L} = \sum_{n=1}^6 (y_{n1} \ln P_{n1} + y_{n2} \ln P_{n2}) \quad (\text{direkt aus Tabelle; } \sum_i y_{ni} = 5)$$

bzw.

$$\tilde{L} = \sum_{n=1}^{30} (y_{n1} \ln P_{n1} + y_{n2} \ln P_{n2}) \quad (\text{jede Entscheidung einzeln; } \sum_i y_{ni} = 1).$$

Abbildung 4.12 zeigt, dass in beiden Modellen die geschätzten Werte $\hat{\beta}_1 \approx -1.1$ und $\hat{\beta}_2 \approx -0.11$ in etwa gleich sind.²⁶ Dies ist konsistent mit Abb 4.10, da man aus ihr ablesen kann, dass bis etwa $I = 4$ Alternativen und bei Auswahlwahrscheinlichkeiten, welche nicht zu dicht an null oder eins liegen, die Probit- und Logit-Auswahlwahrscheinlichkeiten sich nur geringfügig unterscheiden.²⁷ Diese angenäherte Gleichheit der Parameterwerte gilt allerdings nur, wenn die Nutzenfunktionen gleich skaliert sind, in der Regel auf

²⁶Siehe dazu entsprechende Beispiele im Lehrbuch von Meier/Weiss.

²⁷Dann sind Abweichungen des Zufallsnutzens vom Erwartungswert um mehrere Standardabweichungen relevant: Nur bei derartigen Abweichungen unterscheiden sich die Normal- und die Gumbelverteilungen deutlich.

4 Modelle diskreter Entscheidungen

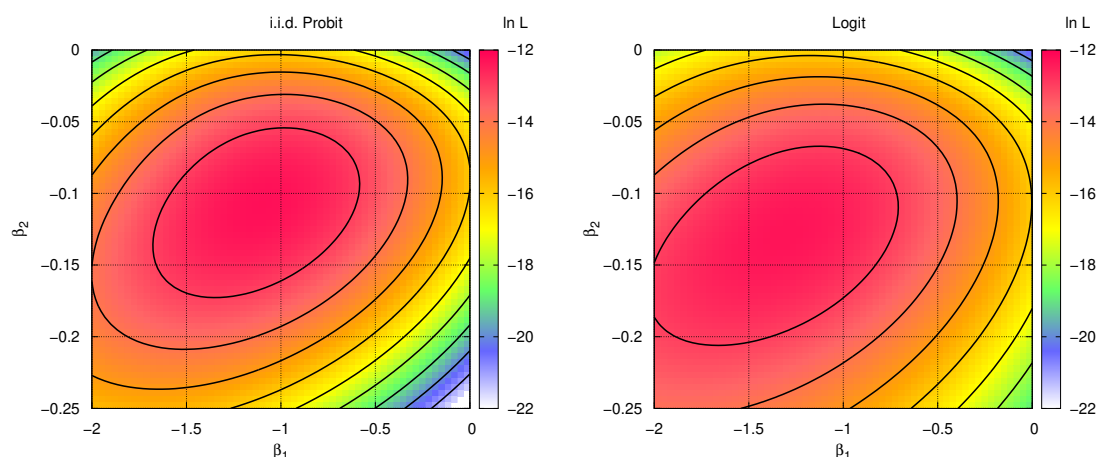


Abbildung 4.12: Log-Likelihood (ohne den konstanten Summanden $\sum_i \ln B_i$) des binomiale Probit-Modells (links) und des entsprechende Logit-Modells (rechts) für die deterministische Nutzenfunktion $V_{ni} = \beta_1 \delta_{i1} + \beta_2 T_{ni}$ zu den in Tabelle 4.1 gegebenen Daten. Die deterministische Nutzenfunktion ist jeweils in Einheiten von σ_ϵ skaliert. Der Abstand einer ‘Höhenlinie’ entspricht einer Log-Likelihood-Differenz von 1.

die Standardabweichung σ_ϵ des Zufallsnutzens. Andernfalls, insbesondere wenn man den Nenner $\sqrt{2}$ im binären Probitmodell weglässt oder die Parameter des standardisierten Logit-Modells nicht nachträglich mit $\sqrt{6}/\pi$ multipliziert, erhält man verschiedene Zahlenwerte der Parameter. Meist sind jedoch nur *Quotienten* der Parameter relevant, da diese (wegen der Skaleninvarianz) Auswahlwahrscheinlichkeiten und abgeleitete Größen wie marginale Konsumneigungen, eine Zahlungsbereitschaft (*willingness to pay*, *WTP*) für gewisse Nutzenänderungen, Elastizitäten, subjektive Zeitwerte (*value of travel time savings*, *VTTS*) und dergleichen angeben. Bei Quotienten jedoch spielen unterschiedliche Skalierungen keine Rolle, da sich diese weghebt, so dass wir folgendes wichtige Ergebnis erhalten:

Für $I \leq 4$ Alternativen und nicht zu extremen Auswahlwahrscheinlichkeiten, $P_{ni} \in [0.01, 0.99]$, sind die Quotienten geschätzter Parameterwerte und damit marginale Kosten, Zeitwerte, Elastizitäten und andere gewonnene ökonometrische Aussagen in Probitmodellen mit unkorrelierten Zufallsnutzen und in Logit-Modellen nahezu gleich.

Im Beispiel bleibt der einzige Quotient $\hat{\beta}_1/\hat{\beta}_2$ unabhängig von der Skalierung und man

4 Modelle diskreter Entscheidungen

Klasse \ Modus	$i = 1$ (Fuß)	$i = 2$ (Rad)	$i = 3$ (ÖPNV)	$i = 4$ (Kfz)	y_n
$n = 1$: 0-1 km	17	16	10	0	43
$n = 2$: 1-2 km	9	23	20	2	54
$n = 3$: 2-5 km	2	27	55	4	88
$n = 4$: 5-10 km	0	7	42	7	56
$n = 5$: 10-20 km	0	0	18	7	25

Tabelle 4.2: Aggregierte Daten der Revealed-Choice-Befragungen in den Ökonometrie-Vorlesungen 2007-2011.

erhält in beiden Modellen nahezu denselben Wert:

$$\left(\frac{\hat{\beta}_1}{\hat{\beta}_2} \right)_{\text{Probit}} = 10, \quad \left(\frac{\hat{\beta}_1}{\hat{\beta}_2} \right)_{\text{Logit}} = 10.2$$

Er gibt an, dass die globale Bevorzugung für Alternative 2 (=globale Benachteiligung von Alternative 1) etwa 10 min entspricht, Alternative 2 also im Mittel gleich häufig wie Alternative 1 gewählt wird, wenn sie eine um 10 min längere komplexe Reisezeit aufweisen würde. Im Falle definierter skalierten Nutzenfunktionen (wie hier) gewinnt man zusätzlich Information über die Standardabweichung des Zufallsnutzens, hier $-1/\beta_1 \approx 9$ min.

Beispiel 2: Revealed-Choice Vorlesungsbefragung mit Entfernungen

Das folgende Beispiel basiert auf einer unter Studenten in den Ökonometrie-Vorlesungen von 2007 bis 2011 durchgeführten Umfrage, welches Verkehrsmittel (zu Fuß, Rad, ÖV, MIV) am Befragungstag für den Weg "Wohnung-Uni" gewählt wurde. Die Ergebnisse sind in Tabelle 4.2 dargestellt: Zur einfacheren Datenerhebung in der Vorlesung wurden die Daten aggregiert erfasst und anstelle der individuellen Entfernungen nur Entfernungsklassen berücksichtigt. Der Index n bedeutet damit keine einzelne Person, sondern eine Klasse von Personen mit y_n Mitgliedern. Es sei darauf hingewiesen, dass die Aggregation nur aufgrund der schnelleren Datenerhebung sowie aus Gründen der Übersichtlichkeit gemacht wurde. Abgesehen davon verursacht sie nur unnötige Genauigkeitseinbußen, da dadurch identische exogene Faktoren bei den Personen innerhalb jeder Gruppe angenommen werden.

Um die Darstellung zu vereinfachen, werden zunächst nur zwei Alternativen unterschieden und folgende Aggregation durchgeführt:

- Die "langsamen" nichtmotorisierten Verkehrsmittel Fuß und Rad werden zu Alternative $i = 1$ zusammengefasst,
- die "schnellen" motorisierten Alternativen ÖPNV und MIV (motorisierter Individualverkehr) zu Alternative 2.

4 Modelle diskreter Entscheidungen

Klasse n	Klassenmitte r_n	$i = 1$ (Fuß/Rad)	$i = 2$ (ÖPNV/Kfz)	y_n
$n = 1$: 0-1 km	0.5 km	33	10	43
$n = 2$: 1-2 km	1.5 km	32	22	54
$n = 3$: 2-5 km	3.5 km	29	59	88
$n = 4$: 5-10 km	7.5 km	7	49	56
$n = 5$: 10-20 km	15.0 km	0	25	25

Tabelle 4.3: Aggregierte Daten für die vereinfachte binomiale Formulierung der Revealed-Choice-Befragungen in den Vorlesungen.

Ferner wird für die Entfernung – mangels detaillierter Informationen – jeweils die Klassenmitte r_n genommen. Damit resultiert aus den ursprünglichen Daten die Daten der Tabelle 4.3. Die aggregierte deterministische Nutzenfunktion sei gegeben durch

$$V_{ni}(\beta_1, \beta_2) = \beta_1 r_n \delta_{i1} + \beta_2 \delta_{i1}. \quad (4.65)$$

Sie enthält als Faktoren lediglich die alternativenspezifisch formulierte Entfernung $x_{1ni} = r_n \delta_{i1}$ (Bewertung der Entfernung bei Fuß/Rad gegenüber der Referenz ÖV/MIV) und die alternativenspezifische Konstante $x_{2ni} = \delta_{i1}$ (globale Präferenz Fuß/Rad gegenüber der Referenz ÖV/MIV).

Da die Entfernungen nicht nach Verkehrsmitteln unterschieden wurden, $r_{ni} = r_n$, muss man die eigentlich generische Variable “Entfernung” formal wie eine soziodemographische Variable behandeln: Bei ein und derselben Person sind ja in der Tat die Entfernungen kaum unterschiedlich. Insbesondere wäre, im Gegensatz zum allgemeinen Fall alternativenspezifisch formulierter generischer Variablen, das Modell durch einen weiteren Faktor $\beta_3 r_n \delta_{i2}$ fehlspezifiziert.

Da es nur zwei Parameter gibt, kann die Schätzung direkt durch Plotten der Log-Likelihood, Abb. 4.13, vorgenommen werden. Der aus der Abbildung entnehmbare Schätzer $\hat{\beta}_1 = -0.46$ bezieht sich auf die “normale” Logit-Skalierung in Vielfachem von $\pi/\sqrt{6}$ des Zufallsnutzens. Dies bedeutet, dass mit jedem zusätzlichen Kilometer die Attraktivität der Fuß/Rad Wege um $0.46\sqrt{6}/\pi \approx 0.35$ Zufallsnutzenunschärfen gegenüber den ÖV/MIV-Wege abnimmt. Der Schätzer $\hat{\beta}_2 = 1.1$ bedeutet, dass bei der Entfernung $r = 0$ die Option Fuß/Rad gegenüber ÖV/MIV einen Bonus von $1.1\sqrt{6}/\pi \approx 0.9$ Zufallsnutzenunschärfen hat: Man muss schon ein Amerikaner sein, um bei einer Distanz von, z.B., 100 m das Auto anzuwerfen. Die aus der Abb. 4.13 entnehmbare negative Korrelation zwischen $\hat{\beta}_1$ und $\hat{\beta}_2$ ist auch anschaulich. Im Wesentlichen müssen die Wahrscheinlichkeiten, also die Nutzendifferenzen, für die am häufigsten vorkommenden mittleren Entfernungen richtig geschätzt werden. Dort kann man einen negativeren Entfernungsbeitrag für Fuß/Rad-Wege (β_1 negativer) durch einen positiveren globalen Bonus β_2 bei Entfernung null (β_2 positiver) teilweise kompensieren

Abbildung 4.14 stellt die Aussagekraft des Modells durch einen Vergleich der beobachteten und geschätzten relativen Häufigkeiten dar (Quotient y_{n1}/y_n bzw. Wahrscheinlich-

4 Modelle diskreter Entscheidungen

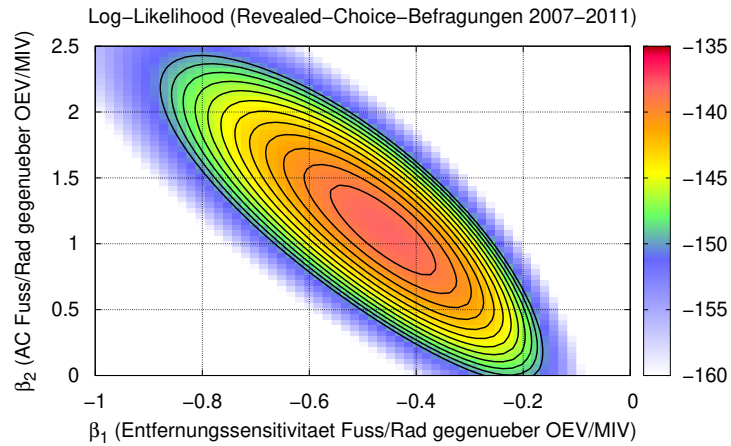


Abbildung 4.13: Log-Likelihood des durch (4.65) spezifizierten Binomial-Logit Modells bezüglich der Daten von Tabelle 4.3. Die geschätzten Parameter sind $\beta_1 = -0.46 \pm 0.05$ und $\beta_2 = 1.1 \pm 0.2$

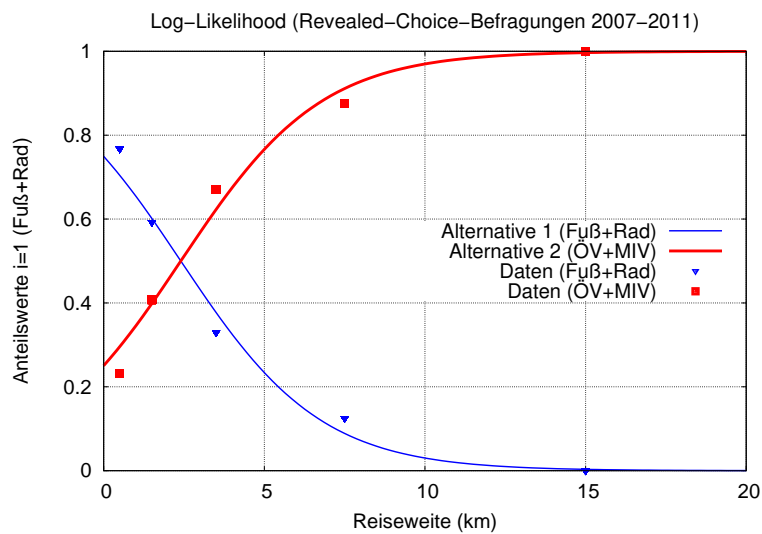


Abbildung 4.14: Nach Tabelle 4.3 beobachtete und nach (4.65) modellierte relative Häufigkeiten der Fuß/Rad-Wege in den verschiedenen Entfernungsklassen für die Schätzwerte $\hat{\beta}_1 = -0.46$ und $\hat{\beta}_2 = 1.1$.

4 Modelle diskreter Entscheidungen

keit $P_{n1} = P_1(r_n)$). Man sieht, dass die Aussagegüte ganz ordentlich ist. Die Nutzendifferenz $V_1 - V_2 = V_1$ ist null, falls $\beta_1 r + \beta_2 = 0$, also bei einer Reiseweite $r = -\beta_2/\beta_1 = 2.4$ km. Bei dieser Entfernung sind also die modellierten Anteilswerte bzw. Entscheidungswahrscheinlichkeiten $=0.5$, was auch in der Abbildung zu sehen ist.

Man beachte, dass die modellierten Wahrscheinlichkeitskurven *nicht* das Ergebnis einer LSE-Schätzung sind. Letztere würde hier verzerrte Schätzer liefern!

4.7.4 Kalibrierungsbedingungen für das allgemeine Entscheidungsmodell

Die *grafische* Methode der Parameterschätzung unter direkter Verwendung der Log-Likelihood (4.63) des allgemeinen Entscheidungsmodells stößt, unabhängig von der Zahl der Alternativen, bei mehr als zwei Parametern an Grenzen: Die Log-Likelihoodfunktion $\tilde{L}(\boldsymbol{\beta})$ kann man dann einfach nicht mehr vollständig plotten. Die *funktionale* Schätzung durch Maximierung von $\tilde{L}(\boldsymbol{\beta})$ bzw. Minimierung von $-\tilde{L}(\boldsymbol{\beta})$ ergibt als notwendige Bedingungen (Extremalwert) M im Allgemeinen nichtlineare Gleichungen für die M Parameter:

$$\frac{\partial \tilde{L}(\boldsymbol{\beta})}{\partial \beta_m} = \sum_{n=1}^N \sum_{i=1}^I \frac{y_{ni}}{P_{ni}(\boldsymbol{\beta})} \frac{\partial P_{ni}(\boldsymbol{\beta})}{\partial \beta_m} = 0. \quad (4.66)$$

Kompakt als Ableitung nach dem Parametervektor $\boldsymbol{\beta}$ geschrieben (vgl. die Definitionen und Rechenregeln der Abschnitte 2.11.2 und 2.11.3), lauten die notwendigen Kalibrierungsbedingungen

$$\frac{\partial \tilde{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{n=1}^N \sum_{i=1}^I \frac{y_{ni}}{P_{ni}(\boldsymbol{\beta})} \frac{\partial P_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}. \quad (4.67)$$

Man kann für das Logitmodell und das i.i.d. Probitmodell zeigen, dass bei parameterlinearen Nutzenfunktionen das so gefundene Extremum tatsächlich ein globales Maximum darstellt. Es gilt dann sogar weitergehend, dass die Likelihood- und Loglikelihood-Funktionen global konkav, also überall "nach unten gekrümmt" sind²⁸ Damit sind (4.66) bzw. (4.67) nicht nur notwendig sondern auch hinreichend.

Die nichtlinearen Gleichungen (4.67) sind, abgesehen von Trivialmodellen, nur numerisch lösbar. Dies geschieht mit Methoden der **nichtlinearen Optimierung**. Bekannte Verfahren sind

- das konzeptionell einfachste, aber bei Nichterfüllung der globalen Konvexität häufig instabile **Newton-Verfahren** (vgl. Abschnitt 4.8),
- das *Gradient-Descent* Verfahren (langsam, aber sicher),
- *Quasi-Newton-Verfahren*, welche ausnutzen, dass $\tilde{L}(\boldsymbol{\beta})$ aus einer Summe von i.A. vielen Summanden besteht,

²⁸McFadden, *Conditional logit analysis of qualitative choice behavior*, Institute of Urban and Regional Development, University of California (1973).

4 Modelle diskreter Entscheidungen

- Verfahren wie der *Broyden-Fletcher-Goldfarb-Shanno (BFGS) Algorithmus* oder der *Levenberg-Marquardt-Algorithmus*, welche die besten Eigenschaften des *Gradient-Descent* und der *Quasi-Newton-Verfahren* kombinieren. Speziell der BFGS-Algorithmus wird in vielen ökonomischen Softwareprodukten angewandt,
- *Genetische Algorithmen*²⁹.

Häufig wird das BFGS-Verfahren angewandt. Schwieriger wird es bei parameter-nichtlinearen Nutzenfunktionen, welche beispielsweise bei der Modellierung von *Wahrnehmungsschwellen* benötigt (Abschnitt 4.10). Dann gilt folgende Warnung:

Parameter-nichtlineare Nutzenfunktionen oder korrelierte Zufallsnutzen können nicht nur zu nicht-konkaven Log-Likelihoods führen sondern auch zu welchen mit Nebenmaxima (vgl. Abb. 4.19). Dann läuft jede der obigen Methoden Gefahr, sich in ein solches zu “verrennen”. Das Risiko wird durch einen möglichst guten *initial guess* sowie ggf durch die Wahl von *genetische Algorithmen* minimiert.

Eine wichtige Rolle wird später auch die symmetrische $M \times M$ **Hesse-Matrix** der zweiten Ableitungen von $-\tilde{L}$ spielen.³⁰ Die negative Hesse-Matrix an der Stelle $\beta = \hat{\beta}$ wird, auch aus später klar werdenden Gründen, **Informationsmatrix** genannt:

$$H_{jm} = -\frac{\partial^2 \tilde{L}(\hat{\beta})}{\partial \beta_j \partial \beta_m} \quad \text{bzw.} \quad \mathbf{H} = -\frac{\partial^2 \tilde{L}(\hat{\beta})}{\partial \beta \partial \beta'}. \quad (4.68)$$

Im folgenden wird die Maximum-Likelihood-Methode für die bisher vorgestellten Modelle spezialisiert.

4.7.5 Schätzung des binomialen Probit-Modells mit quasilinearen Nutzen

Skaliert man die Nutzenfunktionen in Einheiten der Standardabweichung σ_ϵ des Zufallsnutzens, ergeben sich nach (4.22) die Auswahlwahrscheinlichkeiten des Probitmodells zu

$$P_{n1}(\beta) = \Phi \left(\frac{V_{n1}(\beta) - V_{n2}(\beta)}{\sqrt{2}} \right), \quad P_{n2}(\beta) = 1 - P_{n1}(\beta).$$

²⁹Mit denen schießt man aber hier i.A. “mit Kanonen auf Spatzen”, zumindest, wenn man a priori weiß, dass es ein einziges globales Maximum gibt.

³⁰Damit kann man die Eigenschaft, dass die Log-Likelihoodfunktion global konkav bzw. $-\tilde{L}$ global konvex ist, genauer definieren: die Hesse-Matrix der zweiten Ableitungen von $-\tilde{L}$ ist überall positiv definit.

4 Modelle diskreter Entscheidungen

Die Ableitung der Wahlwahrscheinlichkeiten $P_{ni}(\boldsymbol{\beta})$ ergeben für parameterlineare Nutzenfunktionen der Art (4.4), also $V_{ni} = \boldsymbol{\beta}' \mathbf{x}_{ni}$, die Bedingungen

$$\begin{aligned} \frac{\partial P_{n1}}{\partial \boldsymbol{\beta}} &= \frac{1}{\sqrt{2}} f_N \left(\frac{V_{n1} - V_{n2}}{\sqrt{2}} \right) \frac{\partial}{\partial \boldsymbol{\beta}} (V_{n1} - V_{n2}) \\ &= \frac{1}{\sqrt{2}} f_N \left(\frac{V_{n1} - V_{n2}}{\sqrt{2}} \right) (\mathbf{x}_{n1} - \mathbf{x}_{n2}), \\ \frac{\partial P_{n2}}{\partial \boldsymbol{\beta}} &= -\frac{\partial P_{n1}}{\partial \boldsymbol{\beta}}. \end{aligned}$$

Hierbei ist $f_N(x) = \Phi'(x) = 1/\sqrt{2\pi}e^{-x^2/2}$ die Dichtefunktion der Standardnormalverteilung. Setzt man dies in (4.67) ein, erhält man

$\frac{\partial \tilde{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{\sqrt{2}} \sum_{n=1}^N \left(\frac{y_{n1}}{P_{n1}} - \frac{y_{n2}}{P_{n2}} \right) f_N \left(\frac{V_{n1} - V_{n2}}{\sqrt{2}} \right) (\mathbf{x}_{n1} - \mathbf{x}_{n2}) = 0.$	Kalibrierung binomiales Probitmodell.
(4.69)	

4.7.6 Schätzen der Logit-Modelle mit quasilinearen Nutzen

Für die MNL-Auswahlwahrscheinlichkeiten (4.34), also $P_{ni} = e^{V_{ni}} / \sum_{i'} e^{V_{ni'}}$ erhält man für die Log-Likelihood $\tilde{L} = \sum_{n,i} y_{ni} \ln P_{ni}$ direkt

$$\tilde{L}(\boldsymbol{\beta}) = \sum_{n,i} \left[y_{ni} \left(V_{ni} - \ln \sum_{i'} e^{V_{ni'}} \right) \right] \quad (4.70)$$

und durch Ableiten und Nullsetzen (nichttrivial, Interessierte sehen die Herleitung in Abschnitt 4.16 nach!) die Kalibrierungsgleichungen

$$\frac{\partial \tilde{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{n,i} \left[y_{ni} - P_{ni}(\boldsymbol{\beta}) \right] \frac{\partial V_{ni}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}. \quad (4.71)$$

Hierbei wurde die Doppelsumme verkürzt geschrieben: $\sum_{n=1}^N \sum_{i=1}^I = \sum_{n,i}$. Dies sind M nichtlineare Gleichungen für die M Parameter β_m . In Worten lautet die m -te Bedingung:

Die Summe (über alle Alternativen und Personen bzw. Choice Sets) der mit dem Grenznutzen bezüglich β_m gewichteten Differenzen zwischen den gemessenen und vom Modell erwarteten absoluten Häufigkeiten muss für alle $m = 1, \dots, M$ verschwinden.

Der praktisch wichtigste Fall sind quasilineare deterministische Nutzenfunktionen gemäß Gl. (4.4), also $V_{ni} = \boldsymbol{\beta}' \mathbf{x}_{ni}$. Dann wird (4.71) zum (allerdings immer noch nichtlinearen) Gleichungssystem

4 Modelle diskreter Entscheidungen

$$\frac{\partial \tilde{L}(\hat{\beta})}{\partial \beta} = \sum_{n,i} [y_{ni} - P_{ni}(\hat{\beta})] \mathbf{x}_{ni} = \mathbf{0}. \quad \begin{array}{l} \text{Kalibrierung} \\ \text{MNL-Modell.} \end{array} \quad (4.72)$$

Dies kann man anschaulicher ausdrücken, wenn man **Merkmalssummen** als Summe der exogenen Faktoren x_{nmi} über alle Alternativen und Personen definiert:

$$X_m = \sum_{n,i} y_{ni} x_{nmi}, \quad \text{bzw.} \quad \mathbf{X} = \sum_{n,i} y_{ni} \mathbf{x}_{ni} \quad (4.73)$$

Dann lauten die MNL-Kalibrierungsbedingungen für quasilineare deterministische Nutzen $V_{ni} = \beta' \mathbf{x}_{ni}$ schlicht und ergreifend

Der vom MNL geschätzte Vektor \mathbf{X}^{MNL} der Merkmalssummen muss gleich dem beobachteten Merkmalssummen-Vektor \mathbf{X}^{data} sein, also muss für alle Komponenten m gelten

$$X_m^{\text{MNL}} = X_m^{\text{data}}, \quad X_m^{\text{MNL}} = \sum_{n,i} P_{ni}(\hat{\beta}) x_{nmi}, \quad X_m^{\text{data}} = \sum_{n,i} y_{ni} x_{nmi} = \sum_n x_{mni_n}.$$

Prinzip-Beispiel für vier Einflussfaktoren

Der quasilineare deterministische Nutzen zweier Verkehrsmodi i sei gemäß folgender Gleichung aus vier Kriterien bzw. Einflussfaktoren zusammengesetzt:

$$V_{ni} = \beta' \mathbf{x}_{ni} = \sum_{m=1}^4 \beta_m x_{nmi} = \beta_1 T_{ni} + \beta_2 K_{ni} + \beta_3 g_i \delta_{i1} + \beta_4 \delta_{i1}.$$

mit

- $x_{1ni} = T_{ni}$ der komplexen Reisezeit für Person n bei Wahl des Verkehrsmodus i ,
- $x_{2ni} = K_{ni}$ die zugehörigen (Grenz-)Kosten,
- $x_{3ni} = g_n \delta_{i1}$ der exogene Faktor für das Geschlecht mit $g_{\text{♂}} = 0$ und $g_{\text{♀}} = 1$,
- $x_{4ni} = \delta_{i1}$ der alternativen spezifischen Konstante (Referenz ist Alternative 2).

Die MNL-Schätz- bzw. Kalibrierungsbedingungen der vier zugehörigen Parameter lauten

Faktor 1: Komplexe Reisezeit. Die *Merkmalssumme* X_1 ist gleich der Gesamtreisezeit T bei Umsetzung aller Entscheidungen. Die zugehörige *Schätzbedingung* lautet

$$T^{\text{data}} = T^{\text{MNL}}(\beta)$$

mit

$$\begin{aligned} T^{\text{data}} &= \sum_{n,i} y_{ni} T_{ni} && \text{Realisierte Gesamtreisezeit,} \\ T^{\text{MNL}}(\beta) &= \sum_{n,i} P_{ni}(\beta) T_{ni} && \text{modellierte Gesamtreisezeit.} \end{aligned}$$

4 Modelle diskreter Entscheidungen

Faktor 2: (Grenz-)Kosten. Die *Merkmalssumme* X_2 ist gleich den Gesamtkosten K bei Umsetzung aller Entscheidungen. Die zugehörige *Schätzbedingung* lautet in Analogie zu den Reisezeiten

$$K^{\text{data}} = K^{\text{MNL}}(\boldsymbol{\beta})$$

mit

$$\begin{aligned} K^{\text{data}} &= \sum_{n,i} y_{ni} K_{ni} && \text{Realisierte Gesamtkosten,} \\ K^{\text{MNL}}(\boldsymbol{\beta}) &= \sum_{n,i} P_{ni}(\boldsymbol{\beta}) K_{ni} && \text{modellierte Gesamtkosten.} \end{aligned}$$

Faktor 3: Bevorzugung der Alternative 1 bei Frauen. Dieser Faktor ist, wie bei sozioökonomischer Variable gefordert, alternativenspezifisch formuliert. Die *Merkmalssumme* $X_3 = \sum_{n,i} y_{ni} g_n \delta_{i1} = \sum_n y_{n1} g_n = N_{1,\text{♀}}$ ist die Anzahl der von Frauen getroffenen Entscheidungen zugunsten Alternative 1. Die zugehörige *Schätzbedingung* lautet

$$N_{1,\text{♀}}^{\text{data}} = N_{1,\text{♀}}^{\text{MNL}}(\boldsymbol{\beta})$$

mit

$$\begin{aligned} N_{1,\text{♀}}^{\text{data}} &= \sum_n y_{n1} g_n && \text{Realisierte Entscheidungszahl für Alt. 1 bei Frauen,} \\ N_{1,\text{♀}}^{\text{MNL}}(\boldsymbol{\beta}) &= \sum_n P_{n1}(\boldsymbol{\beta}) g_n && \text{modellierte Entscheidungszahl für Alt. 1 bei Frauen.} \end{aligned}$$

Faktor 4: Globale Bevorzugung der Alternative 1. Die Schätzung der alternativenspezifischen Konstante geht analog wie die der alternativenspezifischen sozioökonomischen Variable. Die *Merkmalssumme* $X_4 = \sum_{n,i} y_{ni} \delta_{i1} = \sum_n y_{n1} = N_1$ ist die Gesamtzahl der zugunsten Alternative 1 getroffenen Entscheidungen. Die zugehörige *Schätzbedingung* lautet

$$N_1^{\text{data}} = N_1^{\text{MNL}}(\boldsymbol{\beta})$$

mit

$$\begin{aligned} N_1^{\text{data}} &= \sum_n y_{n1} && \text{Realisierte Gesamt-Entscheidungszahl für Alt. 1} \\ N_1^{\text{MNL}}(\boldsymbol{\beta}) &= \sum_n P_{n1}(\boldsymbol{\beta}) && \text{modellierte Gesamt-Entscheidungszahl für Alt. 1.} \end{aligned}$$

4.7.7 Schätzung des Konstantenmodells

Das Konstantenmodell hat keine personenabhängigen exogenen Variablen, nur einen vollständigen Satz an alternativenspezifischen Konstanten (ACs). Mit der Referenzalternative $i = I$ lauten seine deterministischen Nutzenfunktionen

$$V_{ni} = \sum_{j=1}^{I-1} \beta_j \delta_{ji}, \quad (4.74)$$

Neben dem Nullmodell (alle Nutzenfunktionen sind =0) ist das Konstantenmodell ein wichtiges Referenzmodell zur Definition globaler Gütemaße wie dem *Likelihood-Ratio-Index* (vgl. Abschnitt 4.12 weiter unten).

4 Modelle diskreter Entscheidungen

Schätzung auf der Ebene der vorausgesagten Wahrscheinlichkeiten

Für beliebige Alternativenzahl und beliebige Spezifikationen des Zufallsnutzens (Logit, Probit und andere, mit oder ohne Korrelationen) kann die Schätzung auf der Ebene der vorausgesagten Wahrscheinlichkeiten P_{ni} allgemein durchgeführt werden. Hierzu benötigen wir nur die Tatsache, dass aufgrund des Fehlens personenspezifischer Merkmale die modellierte Wahrscheinlichkeit $P_{ni} = P_i$ nicht von der Person n abhängt. Damit ist die Log-Likelihood-Funktion

$$\begin{aligned}\tilde{L}(\mathbf{P}) &= \sum_{ni} y_{ni} \ln P_{ni} \\ &= \sum_{ni} y_{ni} \ln P_i \\ &= \sum_i \left(\sum_n y_{ni} \right) \ln P_i \\ &= \sum_i N_i \ln P_i.\end{aligned}$$

mit $N_i = \sum_n y_{ni}$ der Zahl der realisierten Entscheidungen für Alternative i .

Um dies jetzt nun bezüglich der Wahrscheinlichkeiten zu maximieren, muss man die Nebenbedingung $\sum_i P_i = 1$ berücksichtigen.³¹ ³² Dies wird durch die Technik der sogenannten *Lagrange-Multiplikatoren* realisiert:

Maximierung/Minimierung einer Funktion $f(\mathbf{x})$ mit Nebenbedingung $g(\mathbf{x}) = 0$: Leite nicht nur die Funktion ab und setze sie null, sondern die Funktion abzüglich der mit dem *Lagrange-Multiplikator* λ multiplizierten Nebenbedingung. Die Nebenbedingung muss dabei in der obigen Form $g(\mathbf{x}) = 0$ vorliegen. Bei mehreren Nebenbedingungen gibt es analog mehrere Lagrangeparameter

Hier ergibt dies

$$\frac{\partial}{\partial P_i} \left(\sum_i N_i \ln P_i - \lambda \sum_i P_i \right) = \frac{N_i}{P_i} - \lambda \stackrel{!}{=} 0,$$

also $P_i = N_i/\lambda$. Der noch unbekannte Lagrange-Parameter wird durch Einsetzen in die Nebenbedingung bestimmt:

$$\sum_i P_i = \sum_i \frac{N_i}{\lambda} = \frac{N}{\lambda} = 1 \Rightarrow \lambda = N$$

³¹Andernfalls ist das Ergebnis unbeschränkt und alle $P_i \rightarrow \infty$ wären das "Optimum"

³²Streng genommen müsste man auch $P_i \geq 0$ und $P_i \leq 1$ fordern. Das Ergebnis erfüllt jedoch auch ohne diese Restriktionen diese Nebenbedingungen, so dass diese nicht *relevant* sind.

4 Modelle diskreter Entscheidungen

und damit

$$P_i^{\text{Konstantenmodell}} = \frac{N_i}{N} = f_i \quad (4.75)$$

In Worten: Die prognostizierten Wahrscheinlichkeiten des kalibrierten Konstantenmodells sind gleich den realisierten relativen Häufigkeiten.

Direktschätzung der Parameter des Multinomial-Logit-Modells

Setzt man die MNL-Formel (4.34) mit den deterministischen Nutzenfunktionen (4.74) in (4.75) ein, erhält man für die Quotienten (nicht alle Fälle sind aufgeführt)

$$\frac{P_i}{P_j} = \frac{f_i}{f_j} = e^{V_i - V_j} = \begin{cases} e^{\beta_i - \beta_j} & i \neq I, j \neq I \\ e^{\beta_i} & i \neq I, j = I \end{cases}$$

Aus dem letzteren Fall direkt

$$\beta_i^{\text{MNL, Konstantenmodell}} = \ln\left(\frac{f_i}{f_I}\right) = \ln\left(\frac{N_i}{N_I}\right). \quad (4.76)$$

Direktschätzung der Parameter des binomialen Probit-Modells

Für das binomialen Probit (BP) Modell gilt (mit der Referenzalternative $I = 2$, also $V_1 = \beta_1, V_2 = 0$):

$$P_1 = \Phi\left(\frac{V_1 - V_2}{\sqrt{2}}\right) = \Phi\left(\frac{\beta_1}{\sqrt{2}}\right)$$

also

$$\beta_i^{\text{BP, Konstantenmodell}} = \sqrt{2}\Phi^{-1}\left(\frac{N_1}{N}\right), \quad (4.77)$$

wobei $\Phi^{-1}(\cdot)$ die Umkehrfunktion (Quantilsfunktion) der Standardnormalverteilung bezeichnet.

4.8 Numerische Lösung der Schätzgleichungen mit der Newton-Methode

Das Vorgehen wird für den praktisch wichtigsten Fall der Logit-Modelle mit parameterlinearen Nutzen gezeigt, ist aber analog auch auf die Probitmodelle anwendbar. Es wird das konzeptionell einfache **Newton'sche Iterationsverfahren** vorgestellt, welches meist in wenigen Schritten konvergiert, im Allgemeinen aber auch instabil sein kann (Abschnitt 4.7.4). Für Logitmodelle mit parameterlinearen Nutzen ist \tilde{L} jedoch global konkav, was gleichbedeutend damit ist, dass das Newton'sche Verfahren *immer* konvergiert.

Die Bedingung (4.72) wird zunächst als Standardform einer nichtlinearen multivariaten Nullstellensuche geschrieben: Der Gradient \mathbf{g} der negativen Log-Likelihood³³ muss

³³Der positive Gradient $\partial\tilde{L}/\partial\boldsymbol{\beta}$ wird in der englischsprachigen Literatur oft als *score* bezeichnet. Mit der üblichen Bedeutung ("Ergebnis") hat das aber nichts zu tun.

4 Modelle diskreter Entscheidungen

gleich null sein:

$$\mathbf{g}(\boldsymbol{\beta}) = -\frac{\partial \tilde{L}}{\partial \boldsymbol{\beta}} = \sum_{n,i} [P_{ni}(\boldsymbol{\beta}) - y_{ni}] \mathbf{x}_{ni} \stackrel{!}{=} \mathbf{0}. \quad (4.78)$$

Der Gradient der Logit-Modelle enthält als Komponenten also genau die Merkmalssummen-Differenzen! Für die Probitmodelle ergeben sich der Gradient \mathbf{g} aus (4.67). Hinreichend nahe der gesuchten Nullstelle $\hat{\boldsymbol{\beta}}$ mit $\mathbf{g}(\hat{\boldsymbol{\beta}}) = 0$ gilt

$$\mathbf{g}(\boldsymbol{\beta}) \approx \mathbf{g}(\hat{\boldsymbol{\beta}}) + \mathbf{H} \cdot (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = \mathbf{H} \cdot (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}). \quad (4.79)$$

wobei die Hesse- bzw. Informationsmatrix $\mathbf{H} = \frac{\partial \mathbf{g}}{\partial \boldsymbol{\beta}} = -\frac{\partial^2 \tilde{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}}$ durch (4.68) gegeben ist. Für die Logitmodelle ergeben sich (vgl. wieder Abschnitt 4.16 für eine Herleitung)

- im allgemeinen multinomialen Fall ($I \geq 2$ Alternativen)

$$\mathbf{H} = \sum_{n,i} P_{ni} \mathbf{x}_{ni} \left(\mathbf{x}'_{ni} - \sum_{i'} \mathbf{x}'_{i'} P_{i'} \right), \quad (4.80)$$

- im binomialen Fall ($I = 2$ Alternativen) bei Verlagerung des gesamten deterministischen Nutzens auf Alternative 1

$$\mathbf{H} = \sum_n P_{n1} P_{n2} \mathbf{x}_{n1} \mathbf{x}'_{n1}. \quad (4.81)$$

Um einen Iterationsschritt zu formulieren, löst man die Gleichung (4.79) nach dem gesuchten Schätzwert $\hat{\boldsymbol{\beta}}$ des Koeffizientenvektors auf, wozu eine Multiplikation mit \mathbf{H}^{-1} notwendig ist:

$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta} - \mathbf{H}^{-1} \mathbf{g}(\boldsymbol{\beta}). \quad (4.82)$$

Da (4.79) nur eine Näherung ist, gilt dies auch für die gerade hergeleitete Beziehung. Ist man jedoch nicht zu weit von $\hat{\boldsymbol{\beta}}$ entfernt oder ist \tilde{L} global konkav, konvergiert fortlaufendes Einsetzen der rechten Seite in die linke meist sehr rasch. Häufig kann man als Startbedingung $\boldsymbol{\beta} = \mathbf{0}$ ansetzen. Die numerische Berechnung der Kalibrierungsbedingungen des MNL mit quasilinearer Nutzenfunktion ist also durch folgendes "Rezept" gegeben:

$$\begin{aligned} \text{Startbedingung} & : \boldsymbol{\beta}^{(0)} = \mathbf{0}, \\ i\text{-te Iteration} & : \boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)} - \left(\mathbf{H}(\boldsymbol{\beta}^{(i)}) \right)^{-1} \mathbf{g}(\boldsymbol{\beta}^{(i)}), \\ \text{Ergebnis nach } I\text{-ter Iteration} & : \hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta}^{(I+1)} \end{aligned} \quad (4.83)$$

Die Abbruchbedingung für die letzte Iteration I kann man beispielsweise daran festmachen, dass sich in diesem Iterationsschritt alle Schätzer relativ um nicht mehr als δ (z.B. $\delta = 10^{-3}$) ändern:

$$\frac{\beta_m^{(I+1)} - \beta_m^{(I)}}{\beta_m^{(I+1)}} < \delta \quad \forall m = 1, \dots, M. \quad (4.84)$$

4 Modelle diskreter Entscheidungen

Verständnisfrage:

Wie schnell konvergiert das Newton-Verfahren, wenn die in \mathbf{g} zusammengefassten Funktionen $g_m = -\frac{\partial \tilde{L}}{\partial \beta_m}$ linear in den Parametern β sind? Antwort: Eine Iteration liefert das exakte Ergebnis

4.8.1 Beispiel: Revealed-Choice-Befragung mit zwei Alternativen

Es wird wieder die im zweiten Beispiel des Abschnitts 4.7.3 beschriebene Vorlesungsbefragung der Verkehrsmittelwahl in Abhängigkeit der Entfernungen betrachtet. Ferner wird wie dort die Situation gemäß Tabelle 4.3 auf zwei Alternativen reduziert. Das Binomial-Logit-Modell mit den Nutzenfunktionen (4.65),

$$V_{ni}(\beta_1, \beta_2) = \beta_1 r_n \delta_{i1} + \beta_2 \delta_{i1},$$

ist dasselbe wie bei der grafischen Lösung, das Modell wird aber nun numerisch mit dem Newton-Verfahren geschätzt.

Wir nehmen als Anfangsbedingung $\beta' = (0, 0)$ an. Es ist zu vermuten, dass dies nicht die kalibrierten Werte sind, so dass sich die realisierten und vom Modell geschätzten Merkmalssummen wohl unterscheiden. Dies wird nun untersucht:

Faktor 1: Entfernung zu Fuß oder mit dem Rad. Die Merkmalssumme X_1 ist gleich der Gesamtentfernung R_1 zu Fuß oder mit dem Rad. Die zugehörige *Schätzbedingung* lautet

$$R_1^{\text{data}} = R_1^{\text{MNL}}(\beta).$$

Die Realisierung mit den Daten und die Schätzung durch das Modell mit $\beta' = (0, 0)$, also $P_{n1} = P_{n2} = 0.5$, ergibt

$$\begin{aligned} R_1^{\text{data}} &= \sum_n y_{n1} r_n = \underline{218.5 \text{ km}} && \text{realisierte Gesamtentfernung Fuß/Rad,} \\ R_1^{\text{MNL}}(0, 0) &= \frac{1}{2} \sum_n y_n r_n = \underline{608.25 \text{ km}} && \text{modellierte Gesamtentfernung Fuß/Rad.} \end{aligned}$$

Man beachte, dass hier jeweils $y_n = \sum_i y_{ni}$ gleichartige Entscheidungssituationen (für die verschiedenen Entfernungsklassen) zusammengefasst wurden, was den zusätzlichen Faktor y_n bei der modellierten Merkmalssumme begründet.

Faktor 2: Alternativenspezifische Konstante (AC) der Alternative 1. Die Merkmalssumme X_2 ist die Gesamtzahl der Wahlentscheidungen für Alternative 1: Die zugehörige *Schätzbedingung* lautet

$$N_1^{\text{data}} = N_1^{\text{MNL}}(\beta)$$

mit

$$\begin{aligned} N_1^{\text{data}} &= \sum_n y_{n1} = \underline{101} && \text{realisierte Entscheidungszahl Fuß/Rad,} \\ N_1^{\text{MNL}}(0, 0) &= \frac{1}{2} \sum_n y_n = \underline{144} && \text{modellierte Entscheidungszahl Fuß/Rad.} \end{aligned}$$

4 Modelle diskreter Entscheidungen

Die realisierten und modellierten Merkmalssummen differieren stark, so dass offensichtlich $\beta' = (0, 0)$ nicht der richtige Schätzvektor ist. Insbesondere wird zu Fuß und mit dem Rad eine viel zu hohe Entfernung zurückgelegt, da bei $\beta = \mathbf{0}$ die Auswahlwahrscheinlichkeiten immer $P_{ni} = 1/2$ betragen, auch bei den langen Wegen.

Um das Newton-Verfahren anzuwenden, benötigen wir noch das Inverse der Hessematrix:

$$\mathbf{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \Rightarrow \mathbf{H}^{-1} = \frac{1}{\text{Det}(\mathbf{H})} \begin{pmatrix} H_{22} & -H_{12} \\ -H_{21} & H_{11} \end{pmatrix}.$$

Damit kann man die Newton-Iterationsgleichungen (4.83) schreiben als

$$\begin{aligned} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{(i+1)} &= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{(i)} - \frac{1}{\text{Det}(\mathbf{H}^{(i)})} \begin{pmatrix} H_{22} & -H_{12} \\ -H_{21} & H_{11} \end{pmatrix}^{(i)} \cdot \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}^{(i)} \\ &= \left[\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \frac{1}{\text{Det}(\mathbf{H})} \begin{pmatrix} H_{22}g_1 - H_{12}g_2 \\ -H_{21}g_1 + H_{11}g_2 \end{pmatrix} \right]^{(i)}. \end{aligned} \quad (4.85)$$

Die Matrixelemente werden mit (4.81) berechnet:

$$\begin{aligned} H_{11} &= \sum_n r_n^2 P_{n1} P_{n2}, \\ H_{12} = H_{21} &= \sum_n r_n P_{n1} P_{n2}, \\ H_{22} &= \sum_n P_{n1} P_{n2}. \end{aligned}$$

Startet man die Newton-Iteration mit $\beta = \mathbf{0}$, ergibt sich folgender Verlauf, wobei als Kontrolle auch die beiden Merkmalssummendifferenzen (Komponenten des Gradienten \mathbf{g}) aufgeführt sind:

Iteration	β_1	β_2	$g_1 = X_1^{\text{MNL}} - X_1^{\text{data}}$	$g_2 = X_2^{\text{MNL}} - X_2^{\text{data}}$
0	0.000000	0.000000	384.25	32
1	-0.211613	0.47782	99.70	7.33
2	-0.356599	0.85944	30.21	2.570
3	-0.439557	1.05873	4.70891	0.48637
4	-0.456050	1.09507	0.144063	0.0163218
5	-0.456565	1.09616	$1.37 * 10^{-4}$	$1.60 * 10^{-5}$

Man sieht, dass bereits nach fünf Iterationen das Ergebnis mit bei Weitem hinreichender Genauigkeit gefunden wurde und sich die Merkmalssummen-Differenzen g_1 und g_2 dem Wert null angenähert haben. Damit kann man die Parameterwerte der letzten Zeile mit den Schätzern $\hat{\beta}_1$ bzw. $\hat{\beta}_2$ identifizieren. Diese Schätzer sind konsistent mit der grafischen Lösung der Abb. 4.13

4 Modelle diskreter Entscheidungen

Schließlich ergibt das Newton-Verfahren noch die Informationsmatrix (Hessematrix an der Stelle der geschätzten Parameter):

$$\mathbf{H}(\boldsymbol{\beta}) = \begin{pmatrix} 557 & 131 \\ 131 & 47.2 \end{pmatrix}$$

und, im Vorgriff auf Abschnitt 4.9, die Varianz-Kovarianz-Matrix der Schätzer nach Gl. (4.94) bzw. (4.99)³⁴

$$\mathbf{V} = \mathbf{H}^{-1} = \frac{1}{\text{Det}(\mathbf{H})} \begin{pmatrix} H_{22} & -H_{12} \\ -H_{21} & H_{11} \end{pmatrix} = \begin{pmatrix} 0.00523 & -0.0146 \\ -0.0146 & 0.0618 \end{pmatrix}$$

und damit die Standardabweichungen

$$\sigma(\hat{\beta}_1) = \sqrt{V_{11}} = 0.0723, \quad \sigma(\hat{\beta}_2) = \sqrt{V_{22}} = 0.249,$$

sowie die Kreuzkorrelation (4.100)

$$\rho_{12} = \frac{-H_{12}}{\sqrt{H_{11}H_{22}}} = -0.810.$$

Die negative Korrelation ist konsistent mit der graphischen Schätzung.

4.8.2 Beispiel 2: Revealed-Choice-Befragung mit vier Alternativen

In diesem Beispiel wird die volle Information der Tabelle 4.2 ausgenutzt und alle vier Alternativen separat modelliert. Wieder werden der Einfachheit halber die jeweils $y_n = \sum_i y_{ni}$ Entscheidungssituationen einer ENtfernungsklasse zusammengefasst. In Verallgemeinerung des binomialen Modells (4.65) sind die deterministischen Nutzenfunktionen des multinomialen Modells mit den Alternativen $i = 1$ (Fuß), $i = 2$ (Rad), $i = 3$ (ÖV) und $i = 4$ (MIV) gegeben durch

$$V_{ni}(\boldsymbol{\beta}) = \sum_{m=1}^3 \beta_m r_n \delta_{im} + \sum_{m=1}^3 \beta_{3+m} \delta_{im}. \quad (4.86)$$

Die Kalibrierung der sechs Parameter mit der Newton-Methode (oder irgendeiner anderen Methode) ergibt

Entfernungssensitivität Fuß-MIV	$\hat{\beta}_1 = -1.425 \pm 0.185$	(4.87)
Entfernungssensitivität Rad-MIV	$\hat{\beta}_2 = -0.484 \pm 0.059$	
Entfernungssensitivität ÖV-MIV	$\hat{\beta}_3 = -0.138 \pm 0.035$	
Alternativenspez. Konstante Fuß-MIV	$\hat{\beta}_4 = 4.10 \pm 0.42$	
Alternativenspez. Konstante Rad-MIV	$\hat{\beta}_5 = 3.56 \pm 0.36$	
Alternativenspez. Konstante ÖV-MIV	$\hat{\beta}_6 = 2.95 \pm 0.33$	

³⁴ Alle Formeln gelten nur asymptotisch, d.h. für große Fallzahlen $\sum_{n,i} y_{ni}$. Deshalb wurde auch auf ein Dach zur Markierung von \mathbf{V} oder $\sigma(\hat{\beta}_m)$ verzichtet.

4 Modelle diskreter Entscheidungen

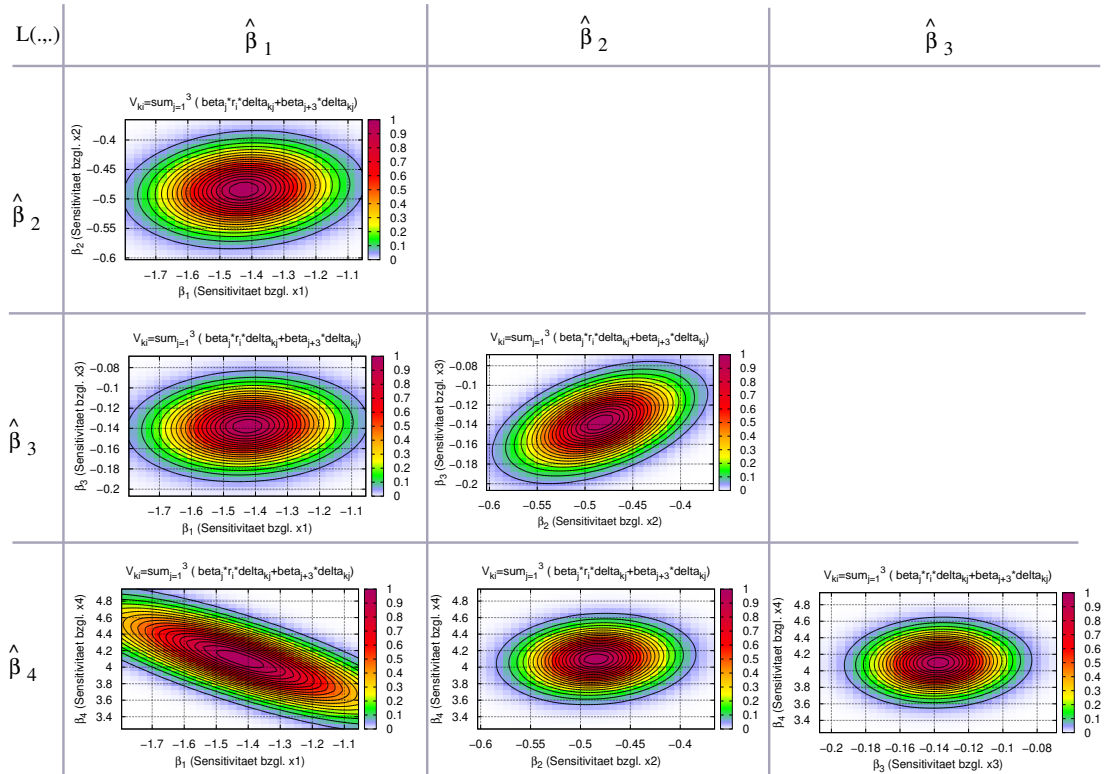


Abbildung 4.15: Likelihoodfunktion $L(\beta)$ für das multinomiale Logit-Modell (4.86) bezüglich der Daten der Tabelle 4.2. Die sechsdimensionale Likelihoodfunktion ist in Form einer graphischen Korrelationsmatrix bezüglich der ersten vier Parameter dargestellt. Die jeweils nicht variierten Parameter werden auf die geschätzten Werte gesetzt.

4 Modelle diskreter Entscheidungen

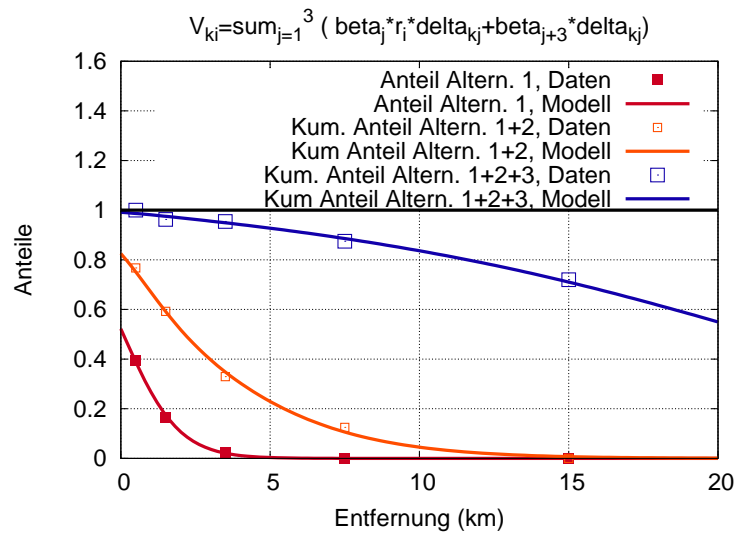


Abbildung 4.16: Kumulierte Auswahlwahrscheinlichkeiten (Kurven) und aus den Daten gewonnene kumulierte bedingte relative Häufigkeiten (Symbole) der Revealed-Choice Befragung in den Vorlesungen. Das MNL-Modells (4.86) wurde dabei mit den Daten der Tabelle 4.2) geschätzt (Ergebnis in Gl. (4.87)).

Ein graphischer Test in Form der Likelihoodfunktion $L(\beta)$, geplottet jeweils in Abhängigkeit von zwei Parametern (die jeweils verbleibenden vier Parameter wurden auf die Schätzwerte gesetzt) ergibt Übereinstimmung sowie einen groben Überblick über die Korrelationen der Schätzer der ersten vier Parameter: Zwischen den Entfernungssensitivitäten ergaben sich kaum Korrelationen, während die Entfernungssensitivitäten und die zugehörigen alternativenspezifischen Konstanten wie beim zwei-Alternativen-Fall stark negativ korreliert sind (und auch die dortige Begründung hier zutrifft).

Die Entfernungssensitivitäten sind konsistent mit dem binärem Modell: Der Wert $\hat{\beta}_1$ des binären Modells entspricht im multinomialen Modell in etwa den Differenzen zwischen den mit den Wahlhäufigkeiten gewichteten Mittelwerten von $\hat{\beta}_1$ und $\hat{\beta}_2$ (Muskelkraft-Alternativen) und $\hat{\beta}_3$ und 0 (motorgetriebene Alternativen). Die alternativenspezifischen Konstanten hingegen sind viel höher. Dies liegt aber einfach daran, dass die Referenzalternative $i = 4$ (MIV) so selten gewählt wurde. Ein Grund dafür ist die nicht abgefragte PKW-Verfügbarkeit, welche einen wichtigen exogenen Faktor darstellt.

Schließlich zeigt der Vergleich zwischen den MNL-Auswahlwahrscheinlichkeiten und den aus den Daten gewonnenen bedingten relativen Häufigkeiten

$$f_n|i = \frac{y_{ni}}{y_n}, \quad (4.88)$$

dass die Modellgüte (im Sinne der Reproduktion beobachteter Daten) erstaunlich gut ist. Schließlich sind alle Parameter signifikant von null verschieden.

4 Modelle diskreter Entscheidungen

Choice Set	Attribute Alternative 1	Attribute Alternative 2	Attribute Alternative 3	Wahl Alt. 1	Wahl Alt. 2	Wahl Alt. 3
1	30 min	30 min	30 min	6	8	19
2	40 min	30 min	30 min	2	10	21
3	50 min	30 min	30 min	0	12	21
4	30 min	40 min	30 min	8	3	22
5	30 min	50 min	30 min	9	2	22
6	30 min	30 min	30 min+1€	10	15	8
7	30 min	30 min	30 min+2€	15	18	0
8	50 min	50 min	30 min+2€	7	14	12

Tabelle 4.4: *Stated Choice*-Befragung. Die *Choice Sets* beziehen sich auf den Weg zur Vorlesung.

4.8.3 Beispiel 3: Stated Choice Conjoint-Analyse

Gegeben sind folgende Ergebnisse einer Stated-Choice Befragung in einer Ökonometrie-Vorlesung: Den Befragten wurde mitgeteilt, dass sich die hypothetischen Auswahl-situationen, also die **Choice Sets**, auf den Weg von zu Hause in den Hörsaal beziehen sollen. Wie bei den vorherigen Beispielen werden alle je $y_n = 33$ Antworten eines Choice Sets n zusammengefasst.³⁵ Jede der drei Alternativen Fuß, Rad und ÖV/MIV³⁶ hat als Attribute (exogene Faktoren) die komplexe Reisezeit T_{ni} und die Ad-Hoc-Kosten K_{ni} , die sich auf den einfachen Weg (nur Hinweg) beziehen:

Diese Ergebnisse werden nun mit Hilfe der **wahlbasierten Conjoint-Analyse** analysiert, was äquivalent zur Formulierung, Schätzung und Diskussion eines multinomialen Wahlmodells ist. Im Gegensatz zu den vorhergehenden beiden Beispielen haben aufgrund der vordefinierten Choice-Sets die generischen Variablen T_{ni} und K_{ni} für jede Alternative im Allgemeinen verschiedene Werte, so dass ihr Effekt generisch anstelle alternativenspezifisch modelliert werden kann. Konkret wird das MNL mit folgenden deterministischen Nutzenfunktionen angesetzt:

$$V_{ni} = \beta_1 T_{ni} + \beta_2 K_{ni} + \beta_3 \delta_{i1} + \beta_4 \delta_{i2}. \quad (4.89)$$

Aus den Daten erhält man folgende Merkmalssummen:

- Merkmalssumme 1: Gesamte Reisezeit

$$X_1^{\text{data}} = T^{\text{data}} = \sum_{i=1}^8 \sum_{n=1}^3 y_{ni} T_{ni} = 8410 \text{ min}$$

³⁵In einer echten Untersuchung wird das natürlich getrennt, also mikroskopisch behandelt; dann kann man auch noch sozioökonomische Variablen berücksichtigen, was hier bei der summarischen Betrachtung nicht möglich ist.

³⁶Da MIV bei den Studenten i.A. sehr selten gewählt wird, wurde er mit dem ÖV zusammengefasst. Im Allgemeinen sollte er separat als Alternative angeboten werden.

4 Modelle diskreter Entscheidungen

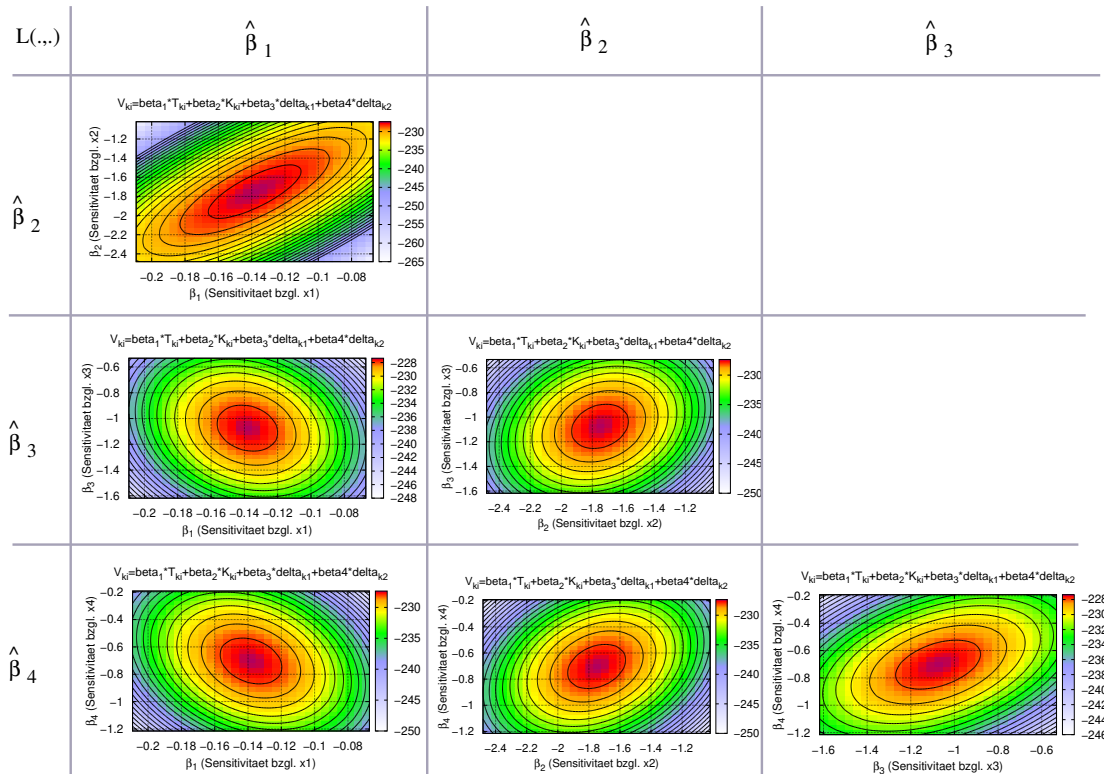


Abbildung 4.17: Log-Likelihoodfunktion $\ln L(\beta)$ für das multinomiale Logit-Modell (4.89) bezüglich der Daten der Tabelle 4.4. Die vierdimensionale Log-Likelihoodfunktion ist in Form einer graphischen Korrelationsmatrix bezüglich der ersten vier Parameter dargestellt. Die jeweils nicht variierten Parameter werden auf die geschätzten Werte gesetzt.

4 Modelle diskreter Entscheidungen

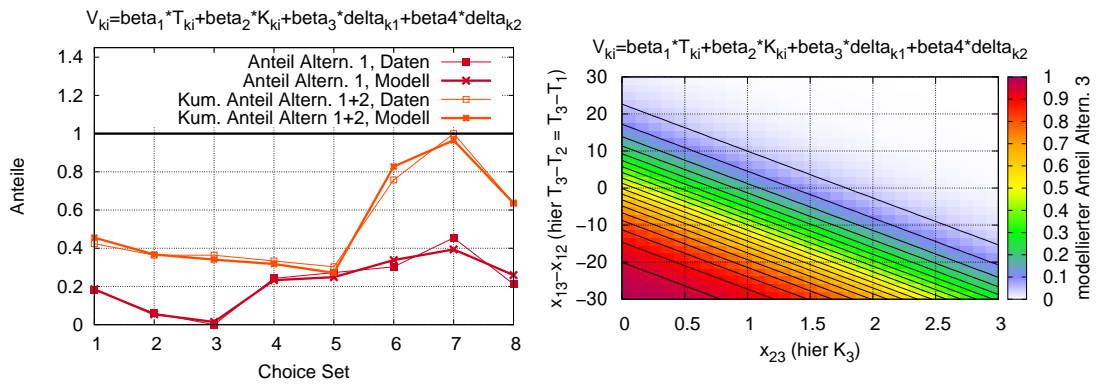


Abbildung 4.18: Kumulierte Auswahlwahrscheinlichkeiten des anhand der Daten der Tabelle 4.4 kalibrierten MNL-Modells (4.89). Zum Vergleich sind die aus den Daten gewonnenen (kumulierten) bedingten relativen Häufigkeiten $f_n|i = y_{ni}/y_n$ geplottet.

- Merkmalssumme 2: Gesamtkosten

$$X_2^{\text{data}} = C^{\text{data}} = \sum_{i=1}^8 \sum_{n=1}^3 y_{ni} C_{ni} = 32 \text{ €}$$

- Merkmalssumme 3: Zahl der Entscheidungen für Alternative 1

$$X_3^{\text{data}} = N_1^{\text{data}} = \sum_{i=1}^8 y_{n1} = 57$$

- Merkmalssumme 4: Zahl der Entscheidungen für Alternative 2

$$X_4^{\text{data}} = N_2^{\text{data}} = \sum_{i=1}^8 y_{n2} = 81$$

Die vier Schätzbedingungen $X_j^{\text{MNL}}(\beta) = X_j^{\text{data}}$ führen auf die Schätzer

$$\begin{array}{ll}
 \text{Zeitsensitivität (alle)} & \hat{\beta}_1 = -0.138 \pm 0.018 \\
 \text{Kostensitivität (alle)} & \hat{\beta}_2 = -1.75 \pm 0.18 \\
 \text{Alternativenspez. Konstante Fuß-ÖV/MIV} & \hat{\beta}_3 = -1.07 \pm 0.14 \\
 \text{Alternativenspez. Konstante Rad-ÖV/MIV} & \hat{\beta}_4 = -0.703 \pm 0.128
 \end{array} \tag{4.90}$$

Die Ergebnisse sind plausibel und ergeben beispielsweise eine (mit $\sqrt{6}/\pi$ multiplizierte) Unschärfe des Zufallsnutzens von $1/((\pi/\sqrt{6}) * 0.138) \approx 6$ min bzw. $1/((\pi/\sqrt{6}) * 1.75) = 0.5$ €, einen Zeitwert von $60 * 0.138 / 1.75 = 4.7$ €/h und einen Anfangsnachteil der muskelkraftgetriebenen Alternativen in der Größenordnung eines Zufallsnutzens.

4 Modelle diskreter Entscheidungen

Die Korrelationen der Schätzer halten sich in Grenzen (Abb. 4.17) und die Aussagekraft ist vergleichsweise hoch (Abb. 4.18 links). Schließlich gibt Abb. 4.18 rechts eine "Prognose", wie sich bei anderen Attributen der Alternativen die Auswahlwahrscheinlichkeit für die Alternative 3 "ÖV/MIV" ändert: Beispielsweise müssen 2 € Kosten durch etwa 25 min Reisezeitersparnis ausgeglichen werden, damit man wieder auf die durch $\hat{\beta}_3$ und $\hat{\beta}_4$ festgelegte Auswahlwahrscheinlichkeit von etwa 50% bei gleichen Kosten und Reisezeiten kommt. Dies kann man auch direkt aus den Geraden gleichen Nutzens ablesen: Die negative Steigung ergibt den Kehrwert des Zeitwertes in Minuten pro Euro.

Verständnisfrage:

Was passiert, wenn in Tabelle 4.4 noch 10 weitere Choice Sets wie das Set 7 mit null Prozent relativer Häufigkeit für den ÖV zur Entscheidung gestanden hätten und man das erwartete Ergebnis "jeweils keine Entscheidung für ÖV/MIV" bekommen hätte? Ändert sich an der Likelihoodfunktion und am Schätzergebnis Wesentliches?

Am Schätzergebnis ändert sich kaum etwas, die neuen Antworten liefern ja auch kaum Information: Zwar verändern die neuen Daten die Likelihoodfunktion an vielen Stellen (insbesondere wird sie bei $\beta = 0$ viel negativer), aber in der Nähe der geschätzten Parameter $\hat{\beta}$ ist sie fast gleich und folgerichtig auch die Schätzer weitgehend unverändert.

4.8.4 Diskussion

Die vorhergehenden Beispiele lassen sich zu folgenden allgemeinen Prinzipien der Modellformulierung und Kalibrierung zusammenfassen:

- Die Qualität der Modelle steht und fällt mit einer sinnvollen Definition der Nutzenfunktion V_{ni} , welche alle wesentlichen Einflussfaktoren enthält, aber keine überflüssigen.
- Ein Kriterium zur Messung der Qualität des Modells ist die Signifikanz der Modellparameter, d.h., ob sie wesentlich von Null verschieden sind (ist in den Beispielen der Fall) und ob die Korrelationen ρ_{lm} der Parameterschätzer nicht zu extrem sind: $|\rho_{lm}| < 0.8$ ist OKeps).
- Wie bei den Regressionsmodellen erhöht sich die Trennschärfe (das Inverse der Standardabweichung der Schätzer) tendenziell mit der Wurzel der Zahl der abgefragten Entscheidungen.
- Ist die alternativenspezifische Konstante einer Alternative ungewöhnlich negativ, ist häufig der Grund dafür, dass diese Alternative bei den meisten Befragten nicht verfügbar ist und dennoch die Verfügbarkeit nicht als exogener Faktor berücksichtigt wurde. Ein Grund dafür ist bei den "MIV"-Alternativen der Beispiele die PKW-Verfügbarkeit, welche einen wichtigen exogenen Faktor darstellt.

4 Modelle diskreter Entscheidungen

- Die Maximalzahl an unabhängigen Entscheidungen in den Daten ist $N(I - 1)$ wobei bei Zusammenfassen von Personengruppen N nicht die Personenzahl sondern die Zahl verschiedener Personengruppen ist. Beispielsweise gilt im Beispiel des Abschnitts 4.8.2 nach Tabelle 4.2 $N(I - 1) = 15$. Ein Modell mit 15 Parametern könnte alle Entscheidungen exakt wiedergeben. Seine Prognosefähigkeit geht dagegen gegen null.³⁷
- Ein überflüssiger Modellparameter kann die Modellqualität drastisch *verschlechtern*, obwohl sich die Likelihood und die “Fitgüte” natürlich verbessern. Nimmt man zur Nutzenfunktion (4.86) zusätzlich einen Anteil $\beta_7 \sqrt{r_n} \delta_{i1}$, so steigt die Log-Likelihood und die Fitgüte marginal, aber die Werte einiger der anderen Parameter verändern sich und ihre Unsicherheiten steigen drastisch:

$$\begin{aligned}\beta_1 &= -1.17 \pm 1.01, \\ \beta_2 &= -0.48 \pm 0.06, \\ \beta_3 &= -0.138 \pm 0.035, \\ \beta_4 &= 4.44 \pm 1.41, \\ \beta_5 &= 3.56 \pm 0.36, \\ \beta_6 &= 2.95 \pm 0.33, \\ \beta_7 &= -0.62 \pm 2.46.\end{aligned}$$

Zunächst einmal ist β_7 überflüssig, da nicht signifikant von null verschieden. Zusätzlich vergrößert dieser Beitrag die Unsicherheiten der Parameter β_1 und β_4 , während die anderen im Wesentlichen unbeeinflusst bleiben. Dies liegt an den extremen Korrelationen $\text{Corr}(\hat{\beta}_0, \hat{\beta}_7) = -0.98$ sowie $\text{Corr}(\hat{\beta}_4, \hat{\beta}_7) = -0.95$.

4.9 Statistische Eigenschaften der ML-Schätzer

4.9.1 Kovarianzmatrix der Parameterschätzer

Die zur Durchführung der Schätzung mit dem Newton-Verfahren sowieso benötigten Größen erlauben eine quantitative Bestimmung der statistischen Eigenschaften der Modellparameter. Diese Analyse ist für eine seriöse Darstellung der Ergebnisse unerlässlich.

Die Analyse wird unter der Annahme durchgeführt, dass innerhalb der noch zu bestimmenden Konfidenzregion, also für alle interessanten Fälle, der Gradient $\mathbf{g}(\boldsymbol{\beta})$ der (negativen) Log-Likelihood linear bezüglich der in $\boldsymbol{\beta}$ enthaltenen Parameter ist. Dies hat mehrere Konsequenzen:

- Da $g_m(\boldsymbol{\beta}) = -\frac{\partial \bar{L}}{\partial \beta_m}$ und an der geschätzten Stelle $\hat{\boldsymbol{\beta}}$ der Gradient $\mathbf{g}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ verschwindet, ist die Log-Likelihoodfunktion selbst in der Nähe von $\hat{\boldsymbol{\beta}}$ quadratisch in

³⁷Man sagt, “mit fünf Parametern kann man einen ganzen Elefanten modellieren und mit sechs sogar das Wackeln seines Schwanzes”. Für die Ökonometrie, deren Modelle oft wesentlich mehr Parametern enthalten, ist dies sicherlich übertrieben, Der Spruch enthält aber dennoch ein Körnchen Wahrheit.

4 Modelle diskreter Entscheidungen

den Parametern und es gilt

$$\frac{d\tilde{L}}{d\boldsymbol{\beta}} = 0 \Rightarrow \tilde{L}(\boldsymbol{\beta}) \approx \tilde{L}_{\max} - \frac{1}{2} \Delta\boldsymbol{\beta}^T \cdot \mathbf{H} \cdot \Delta\boldsymbol{\beta} \quad (4.91)$$

wobei $\tilde{L}_{\max} = \tilde{L}(\hat{\boldsymbol{\beta}})$ und $\Delta\boldsymbol{\beta} = \boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$. Die Hessematrix (Informationsmatrix) \mathbf{H} der zweiten Ableitungen von $-\tilde{L}$ wird ebenfalls an der Stelle $\hat{\boldsymbol{\beta}}$ genommen (vgl. Gl. (4.95)).

- Wegen $\tilde{L} = \ln L$ kann die Likelihood-Funktion *selbst* als Dichtefunktion einer multivariaten Gaußverteilung aufgefasst werden:³⁸

$$L(\boldsymbol{\beta}) = L_{\max} \exp \left[-\frac{1}{2} (\Delta\boldsymbol{\beta}^T \cdot \mathbf{H} \cdot \Delta\boldsymbol{\beta}) \right]. \quad (4.92)$$

- Vergleicht man dies mit der Dichte der allgemeinen multivariaten Gaußverteilung,

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{J+1} \text{Det} \boldsymbol{\Sigma}}} \exp \left[-\frac{1}{2} \mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{x} \right], \quad (4.93)$$

und identifiziert \mathbf{x} mit $\Delta\boldsymbol{\beta}$ sowie die allgemeine Kovarianzmatrix $\boldsymbol{\Sigma}$ mit der Parameter-Kovarianzmatrix \mathbf{V} , erhält man für diese Matrix die wichtige Beziehung

$$\mathbf{V} = \text{Cov}(\hat{\boldsymbol{\beta}}) = E \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \right] \approx \mathbf{H}^{-1}(\hat{\boldsymbol{\beta}}). \quad (4.94)$$

Damit wird auch deutlich, warum $\mathbf{H}(\hat{\boldsymbol{\beta}})$ mit den Komponenten

$$H_{lm} = -\frac{\partial^2 \tilde{L}(\hat{\boldsymbol{\beta}})}{\partial \beta_l \partial \beta_m} \quad (4.95)$$

manchmal als **(Fisher'sche) Informationsmatrix** bezeichnet wird: Je größer die Komponenten von \mathbf{H} , desto kleiner die durch die Kovarianzen ausgedrückte "Informationslücke". Das Ungefähr-Zeichen ist der Näherung der lokalen Linearität der Funktion $\mathbf{g}(\boldsymbol{\beta})$ geschuldet. Die Beziehung wird asymptotisch (also für eine Zahl der Messpunkte gegen unendlich) exakt, deshalb spricht man auch von der **asymptotischen Kovarianzmatrix** der Parameterschätzer.

- Harald Cramér und Calyampudi Radhakrishna Rao haben einen wichtigen Satz, die **Cramér-Rao Ungleichung** hergeleitet: Eine untere Grenze der Varianzen und Kovarianzen der Parameterschätzer ist das Inverse der Fisher'schen Informationsmatrix. Das heißt, der Maximum-Likelihood-Schätzer für *Discrete-Choice* Modelle ist **effizient**.

³⁸ Allerdings ist L nach (4.92), wenn man es als Dichtefunktion von $\hat{\boldsymbol{\beta}}$ auffasst, nicht korrekt normiert. Letztendlich steckt der Satz von Bayes bei gegebenen A-Priori-Wahrscheinlichkeiten dahinter: $L = P(\mathbf{y}|\boldsymbol{\beta})$, während bei gegebenen Daten \mathbf{y} die Parameter-Dichtefunktion $f(\boldsymbol{\beta})$ (salopp ausgedrückt) durch die Bedingung $P(\boldsymbol{\beta} \in \hat{\boldsymbol{\beta}} \pm (\frac{1}{2} \Delta\boldsymbol{\beta})^M | \mathbf{y}) = f(\boldsymbol{\beta}) (\frac{1}{2} \Delta\boldsymbol{\beta})^M$ gegeben ist. Der Satz von Bayes erlaubt, die Dichte f durch L auszudrücken: $f(\boldsymbol{\beta}) = P(\boldsymbol{\beta} \dots | \mathbf{y}) = P(\mathbf{y}|\boldsymbol{\beta}) C = L(\boldsymbol{\beta}) C$ wobei C eine aus den A-Priori-Wahrscheinlichkeiten bestehenden Konstante ist.

4 Modelle diskreter Entscheidungen

- Bei der Regressionsrechnung hatten wir die Varianz-Kovarianzmatrix analog zu (4.94) durch das Inverse der Hessematrix \mathbf{H}_S der Fehlerquadratsumme S am Kalibrierungspunkt schreiben können: $\mathbf{V}_{\text{Regr}} = 2\sigma_\epsilon^2 \mathbf{H}_S^{-1}$. Formal entspricht also die Log-Likelihood-Funktion einem Vielfachen der Fehlerquadratsumme der Regression,

$$\tilde{L}(\boldsymbol{\beta}) \hat{=} -S(\boldsymbol{\beta})/(2\sigma_\epsilon^2),$$

wie auch der Vergleich von (4.58) mit (2.48) nahelegt.

4.9.2 Konfidenzintervalle und Tests

Die Komponenten V_{jm} der Kovarianzmatrix beinhalten die gesuchten Varianzen und Kovarianzen der geschätzten Parameter, woraus sich alle statistischen Eigenschaften wie Schwankungsbreiten, Korrelationen, Konfidenzbereiche etc. genau wie bei den Regressionsmodellen (Abschnitt 2.6) ergeben. Insbesondere sind die Varianzen der Schätzer,

$$V(\hat{\beta}_m) = V_{mm}, \quad (4.96)$$

durch die Diagonalelemente gegeben. Ferner hat die Korrelationsmatrix die Komponenten

$$\rho_{jm} = \frac{V_{jm}}{\sqrt{V_{jj}V_{mm}}}. \quad (4.97)$$

Konfidenzintervalle einzelner Parameter zur Fehlerwahrscheinlichkeit α bekommt man analog zu Gl. (2.65), nur dass wegen der sowieso angenommenen Gauß-Näherung der Likelihoodfunktion die Schätz- bzw. Testvariable gaußverteilt (und nicht Student-t-verteilt³⁹) angenommen wird:⁴⁰

$$K_{\beta_m}^{(\alpha)} : \beta_m \in \left[\hat{\beta}_m - \Delta\hat{\beta}_m, \hat{\beta}_m + \Delta\hat{\beta}_m \right], \quad \Delta\hat{\beta}_m = z_{1-\alpha/2} \sqrt{V_{jj}}. \quad (4.98)$$

Konfidenzintervalle von Parameterkombinationen und mehrdimensionale Konfidenzregionen sowie statistische Tests werden ebenfalls wie bei den Regressionsmodellen (Abschnitte 2.6.4 und 2.6.7) bestimmt, wobei ggf. die Student-t-Verteilung durch eine Standardnormalverteilung ersetzt wird.

Dies alles gilt für Logit-, Probit- und andere Modelle der diskreten Wahltheorie. Die Matrix \mathbf{H} ist allgemein durch die zweiten Ableitungen der negativen Log-Likelihoodfunktion an der Stelle $\hat{\boldsymbol{\beta}}$ gegeben (Gl. (4.68) bzw. (4.95)) und auch ihr Inverses \mathbf{H}^{-1} wurde bei der Newton-Methode bereits benötigt. Für die MNL-Modelle gilt für \mathbf{H} der explizite Ausdruck (4.80).

³⁹In Lehrbüchern und Statistik-Software wird dennoch gerne die Student- t -Verteilung bemüht. Ausgerechnet wird sie aber immer im Grenzfall unendlich vieler Freiheitsgrade, also im Grenzfall der Standardnormalverteilung.

⁴⁰Achtung: Bitte nicht die Elemente der Varianz-Kovarianz-Matrix der Parameterschätzer mit dem deterministischen Nutzen V_{ni} verwechseln!

4 Modelle diskreter Entscheidungen

4.9.3 Spezialfall zweier Modellparameter

Explizit erhält man für zwei Parameter die beiden Varianzschätzer und die Kovarianz gemäß

$$\text{Cov}(\hat{\beta}) = \begin{pmatrix} \hat{V}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \hat{V}(\hat{\beta}_2) \end{pmatrix} = \frac{1}{\text{Det}(\mathbf{H})} \begin{pmatrix} H_{22} & -H_{12} \\ -H_{21} & H_{11} \end{pmatrix} \quad (4.99)$$

sowie die Korrelation

$$\rho_{12} = \frac{-H_{12}}{\sqrt{H_{11}H_{22}}} \quad (4.100)$$

wobei $\mathbf{H} = \mathbf{H}(\hat{\beta})$ an der Stelle der geschätzten Parameterwerte ausgewertet wird.

4.10 Parameter-nichtlineare Nutzenfunktionen



Die meisten Problemstellungen erlauben es, auch nichtlineare Sachverhalte mit parameterlinearen Nutzenfunktionen zu modellieren. Dies stößt jedoch an Grenzen, wenn der nichtlineare Effekt nur innerhalb eines gewissen Bereichs der exogenen Variablen eine Rolle spielt:

- In einigen Zusammenhängen wurde, unter anderem von *Lohse*, ein indifferentes Verhalten innerhalb bestimmter **Schwellen** vorgeschlagen: Es wird angenommen, dass es egal ist, ob absolute Zeiten (oder Zeitdifferenzen) null oder z.B. 5 Minuten betragen (vgl. Tabelle 4.5).
- Umgekehrt wurde von *Kahnemann* und *Tversky* eine **Sensitivitätserhöhung** in der Nähe von Bezugspunkten und außerdem betragsmäßig höhere Sensitivitäten für negative gegenüber positive Abweichungen vom Bezugspunkt vorgeschlagen.⁴¹ (Tabelle 4.6).

Prinzipiell kann man die *Höhe* von Schwellen (d.h. die Änderung der marginalen Sensitivitäten innerhalb des Schwellenbereichs) parameterlinear beschreiben. Kennt man hingegen die *Breite* der Schwellen bzw. den *Bereich* der Sensitivitätsänderung nicht, ist eine im eigentlichen Sinne nichtlineare Beschreibung nötig.

4.10.1 Prinzip-Beispiel

Als einfachstmögliches Beispiel dient Tabelle 4.5: Die (extra konstruierten!) Entscheidungshäufigkeiten spiegeln wider, dass Zeitdifferenzen von 10 Minuten oder weniger kaum als entscheidungsrelevant wahrgenommen werden. Eine nichtlineare Nutzenfunktionen,

⁴¹Letzteres Verhaltensmuster ist auch als **loss aversion** bekannt.

4 Modelle diskreter Entscheidungen

Personen- klasse	Zeit Alternative 1 (min)	Zeit Alternative 2 (min)	Wahl Alt. 1	Wahl Alt. 2
1	25	30	11	10
2	30	30	10	10
3	35	30	10	10
4	40	30	9	11
5	45	30	5	15
6	50	30	2	15
7	55	30	1	15
8	60	30	0	15

Tabelle 4.5: Hypothetische Wahlentscheidungen zwischen zwei Alternativen, welche eine Schwelle in der Nähe gleicher Nutzen repräsentieren. Die einzige (generisch formulierte) exogene Variable ist die komplexe Reisezeit.

Personen- klasse	Zeit Alternative 1 (min)	Zeit Alternative 2 (min)	Wahl Alt. 1	Wahl Alt. 2
1	25	30	16	7
2	30	30	10	10
3	35	30	7	20
4	40	30	3	20
5	45	30	3	25
6	50	30	2	30
7	55	30	1	17
8	60	30	2	50

Tabelle 4.6: Hypothetische Wahlentscheidungen zwischen zwei Alternativen, welche eine erhöhte Zeitsensitivität in der Nähe gleicher Nutzen repräsentieren. Die einzige (generisch formulierte) exogene Variable ist wieder die komplexe Reisezeit.

4 Modelle diskreter Entscheidungen

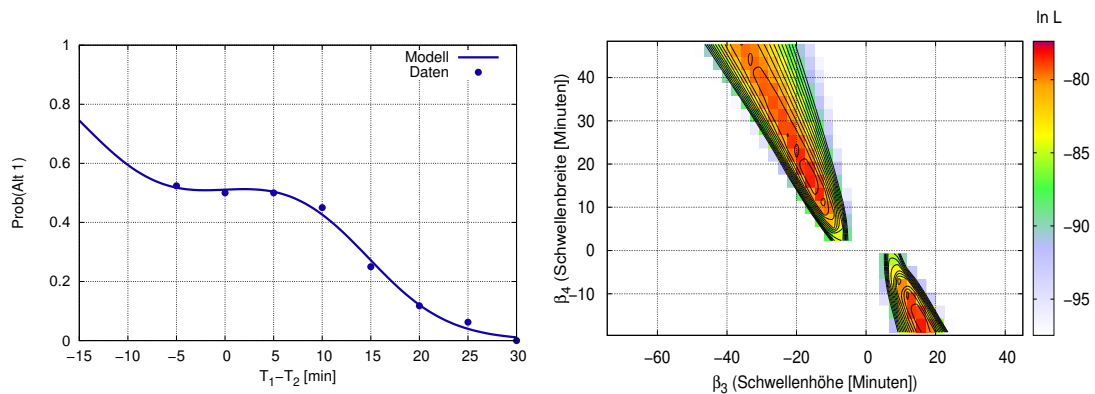


Abbildung 4.19: Parameterschätzung des binären Logitmodells mit nichtlinearen Nutzenfunktionen (4.101) an die Daten der Tabelle 4.5, welche eine Schwelle in der Nähe der Zeitgleichheit widerspiegelt.

welche sowohl die Breite als auch die Stärke (Sensitivitätsabnahme) von Schwellen berücksichtigt, kann folgendermaßen definiert sein:

$$V_{n1} - V_{n2} = \beta_1 + \beta_2 \left[\Delta T_n + \beta_3 \tanh \left(\frac{\Delta T_n}{\beta_4} \right) \right] \quad (4.101)$$

mit $\Delta T_n = T_{n1} - T_{n2}$. Hier bedeuten

- β_1 : alternativenspezifische Konstante (AC) der Alternative 1 gegenüber der Referenz 2,
- β_2 : lineare Zeitsensitivität,
- β_3 : Schwellenhöhe (bei sehr hohen Zeitdifferenzen) bzw. Erniedrigung der Zeitsensitivität (bei vergleichbaren Zeiten), falls β_3 und β_4 unterschiedliche Vorzeichen haben. Bei gleichen Vorzeichen beschreibt β_3 eine Erhöhung der Zeitsensitivität in der Nähe gleicher Zeiten im Sinne von *Kahnemann* und *Tversky*,
- β_4 : Das ungefähre Zeitdifferenz-Intervall, in welchem die Schwelle bzw. erhöhte Sensitivität wirkt. Da die Nichtlinearität unverändert ist, wenn sich sowohl bei β_3 als auch bei β_4 das Vorzeichen ändert, können wir ohne Einschränkung den Wertebereich von β_4 auf nichtnegative Werte beschränken, so dass $\beta_3 > 0 (< 0)$ eine Erhöhung (Erniedrigung) der Sensitivität in der Nähe des Bezugspunkts bedeutet. Schließlich repräsentiert $\beta_4 = 0$ eine Diskontinuität, während $\beta_3 = 0$ das Fehlen jeglicher Nichtlinearität bedeutet.

4 Modelle diskreter Entscheidungen

Das Ergebnis der Parameterschätzung (Schätzwert \pm Standardabweichung) ist

$$\begin{aligned}\hat{\beta}_1 &= 0.043 \pm 0.236, \\ \hat{\beta}_2 &= -0.29 \pm 0.38, \\ \hat{\beta}_3 &= -15 \pm 18, \\ \hat{\beta}_4 &= 14 \pm 21.\end{aligned}$$

Alle vier Parameter sind nicht signifikant von null verschieden! Dennoch zeigt Abbildung 4.19 links, dass die Fitgüte ganz ordentlich ist. Zusätzlich zeigt der Contourplot rechts, dass die Log-Likelihood am Punkt $\beta_3 = \beta_4 = 0$ deutlich negativer als am Schätzwert ist, zumindest, wenn man den Parameterraum in der Ebene der Schwellenparameter β_3 (Schwellenhöhe) und β_4 (Schwellenbreite) schneidet: Sowohl die Schwellenhöhe bzw. Sensitivitätserniedrigung als auch die Schwellenbreite sind deutlich von null verschieden (Die Sensitivitätserniedrigung überkompensiert sogar die lineare Sensitivität bei Zeitdifferenzen gegen null). Dies könnte aber daran liegen, dass die linearen Parameter so “verdreht” sind, dass sie weit vom linearen Optimum entfernt sind. Aber auch gegenüber dem rein linearen Modell (Abb. 4.20 rechts unten) ist die Fitgüte deutlich besser. Gründe für die mangelnde Signifikanz sind in diesem Beispiel:

- Zu wenig Daten: Zum Fitten von 4 Parametern sind nur 8 Datenpunkte verfügbar. Das erhöht die Fitgüte und verschlechtert – gemäß dem Spruch von den zu fittenden Elefanten – die Aussagekraft des Modells.
- Die Nichtlinearitäten bewirken, dass die Log-Likelihood im Parameterraum nicht mehr angenähert quadratisch (konkav) ist, so dass der übliche Ausdruck für die Kovarianzmatrix (die negative Hessematrix von $\tilde{L}(\beta)$) nur in unmittelbarer Nähe der maximalen Log-Likelihood gültig ist, nicht notwendigerweise aber mehr in einem Abstand von einer mit dieser Matrix berechneten theoretischen Standardabweichung, so dass diese Berechnung nicht mehr gültig ist: Für $\beta_3 = -30$ (statt -15) und $\beta_4 = 40$ (statt 14), erhält man nahezu die gleiche Fitgüte: Die Log-Likelihood ist nur um etwa eine Einheit niedriger als das Maximum (was etwa der Differenz bei einer Standardabweichung Abstand entspricht, konsistent mit den angegebenen Werten). Hingegen ist sie bei $\beta_3 = \beta_4 = 0$ um mehr als 20 Einheiten niedriger. Der im folgenden Abschnitt beschriebene *Likelihood-Ratio-Test* berücksichtigt dies und ist hier aussagekräftiger als einfache Konfidenzintervalle.⁴²

Schließlich sieht man, dass die Log-Likelihood nicht nur nicht konkav, sondern sogar bimodal ist. Ein zweites Maximum (von identischer Höhe) liegt bei $\hat{\beta}_3 = +15$ und $\hat{\beta}_4 = -14$. Dieses führte zu identischen Modellaussagen und es ist dementsprechend unproblematisch, wenn die Optimierungsroutine dieses Maximum “erwischt” hätte. Es sind jedoch auch böartige Nebenmaxima möglich. Deshalb ist es bei nichtlinearen Modellen sehr wichtig, die Software mit einem guten Anfangswert (*initial guess*) zu versorgen

⁴²In der Tat ergibt der Likelihood-Ratio Test, dass man die Nullhypothese “das zweiparametrische rein lineare Modell ist nicht schlechter” gerade noch (p -Wert 5%) ablehnen kann.

4 Modelle diskreter Entscheidungen

und die Wertebereiche der Parameter zu begrenzen. Insbesondere haben wir ja bereits die Schwellenbreite β_4 auf nichtnegative Werte begrenzt, so dass das zweite Optimum außerhalb des Gültigkeitsbereichs liegt.

Ein weiterer Grund für die geringe Signifikanz liegt in einer Eigenheit von Modellen der diskreten Wahltheorie: Der Verlauf von Nutzenfunktionen ist weitgehend unbeobachtbar, also nicht kalibrierbar, wenn diese um mehr als etwa 5-10 Nutzeneinheiten (NE) unterhalb der der besten Alternative liegen: Dann ist, wie direktes Einsetzen in die Logit-Formel unmittelbar zeigt, die modellierte Auswahlwahrscheinlichkeit für diese Alternativen unterhalb von e^{-5} bis e^{-10} , also *de-facto* null. Damit ist auch die beobachtete Häufigkeit $y_{ni} = 0$, denn andernfalls würde der entsprechende Summand $y_{ni} \ln P_{ni}$ der Log-Likelihood einen massiv negativen Beitrag liefern, sodass dieser β -Vektor sicherlich nicht geschätzt worden wäre. Als Ergebnis hängt der Summand $\sum_i y_{ni} \ln P_{ni}$ der Log-Likelihood für die entsprechende Person nahezu nicht davon ab, ob die Nutzenfunktion einer "schlechten" Alternative um 5-10 oder um 100 NE unterhalb der besten liegt.

Im Beispiel ist im Wesentlichen nur wichtig, dass der Anfangswert, die Anfangssteigung und die Anfangskrümmung der Nutzendifferenzen bezüglich ΔT stimmen. Diese unterscheiden sich bei z.B. $\beta_3 = 80$ und $\beta_4 = 100$ kaum von der Situation bei den geschätzten Werten. Dies legt nahe, dass man hier entweder die Schwellenbreite oder die Höhe auf einen sinnvollen Wert festlegen kann, oder gleich parameterlinear mit einem quadratischen Term modellieren kann.⁴³

Abbildung 4.20 zeigt das Ergebnis dieser drei Modellvarianten und als vierte Variante die rein lineare Modellierung.

- Fixiert man die Schwellenhöhe auf $\beta_3 = -10$, ergibt sich die Schwellenbreite zu $\hat{\beta}_4 = 9.0 \pm 2.6$. Der lineare Parameter ist ebenfalls signifikant.
- Fixiert man die Breite auf $\beta_4 = 10$, ist die geschätzte Schwellenhöhe $\hat{\beta}_3 = -11.0 \pm 2.0$ (ebenso wie der lineare Parameter) signifikant von Null verschieden.
- Beim quadratischen Ansatz ist immerhin noch der quadratische Vorfaktor $\hat{\beta}_3 = -0.0062 \pm 0.0028$ signifikant.

Schließlich sei noch angemerkt, dass, bei Beibehaltung der Bedingung nichtnegativer Schwellenbreiten, von den vier Modellen der Abb. 4.20 nur das Modell M_2 fähig ist, auch eine erhöhte Sensitivität in der Nähe von Zeitgleichheit zu beschreiben. Allerdings kann man bei Modell M_1 die Bedingung $\beta_4 \geq 0$ fallen lassen und so auch, wie im Modell M_2 und im vollen Modell (Abb. 4.21), Sensitivitätserhöhungen modellieren

Bei Modell M_3 könnte zwar $\beta_3 > 0$ eine erhöhte Sensitivität nahe der Zeitgleichheit beschreiben, führt allerdings auch zu den unsinnigen Ergebnis, dass besonders große Reisezeiten besonders attraktiv sind. Eine Modellierung von nichtlinearen Effekten, egal ob mit parameterlinearen oder streng nichtlinearen Nutzenfunktionen, erfordert also in jedem Fall großes Geschick.

⁴³Aus Konsistenzgründen benötigt man allerdings ungerade nichtlineare Anteile und muss deshalb statt $(\Delta T)^2$ die Zeitabhängigkeit durch $\Delta T|\Delta T|$ oder auch durch $(\Delta T)^3$ modellieren.

4 Modelle diskreter Entscheidungen

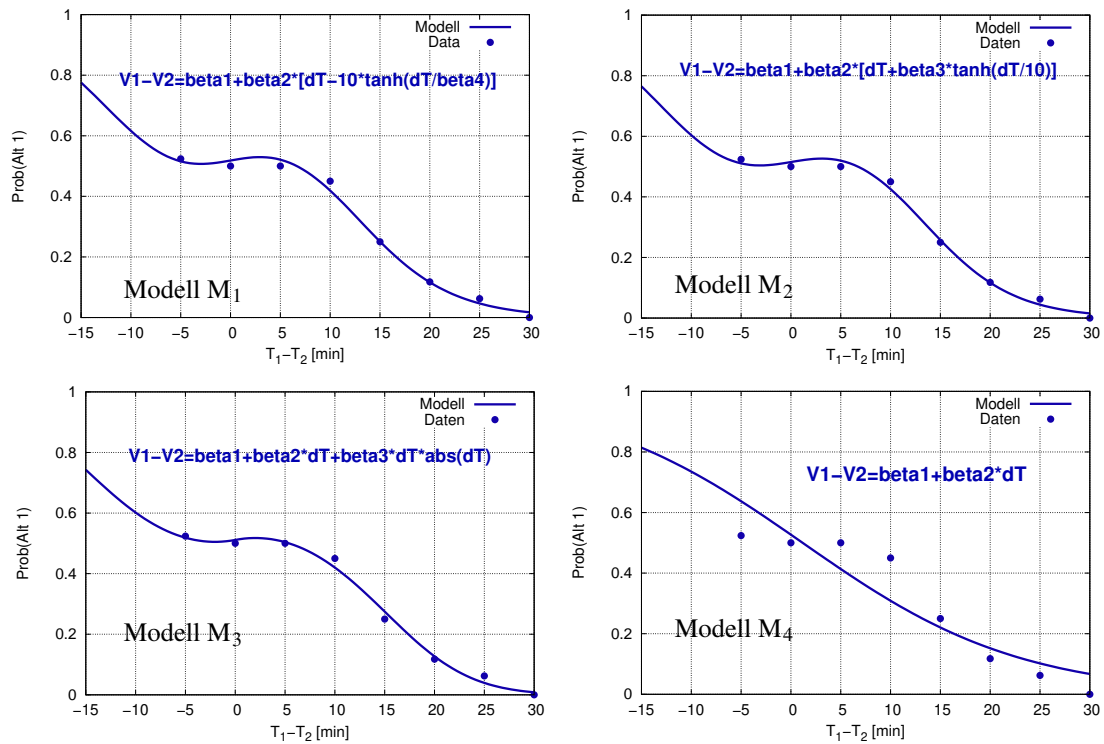


Abbildung 4.20: Fitgüte vierer Logit-Modelle mit verschiedenen Nutzenfunktionen.

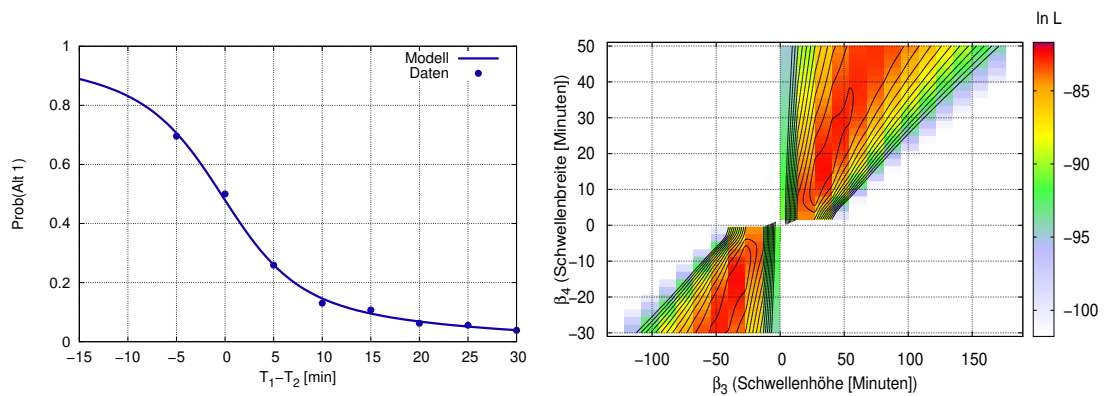


Abbildung 4.21: Parameterschätzung des binären Logitmodells mit nichtlinearen Nutzenfunktionen (4.101) an die Daten der Tabelle 4.6, welche eine erhöhte Zeitsensitivität in der Nähe der Zeitgleichheit widerspiegelt.

4 Modelle diskreter Entscheidungen

Verständnisaufgabe:

Kann man Modell M_2 (Abb. 4.20 rechts oben) auch parameterlinear formulieren?

Verständnisaufgabe:

Warum ist Modell M_3 für $\beta_3 > 0$ inkonsistent für sehr große und sehr kleine Zeitdifferenzen?

Verständnisaufgabe:

Warum ist es unsinnig, Modell M_3 einfach durch $V_1 - V_2 = \beta_1 + \beta_2\Delta T + \beta_3\Delta T^2$ zu beschreiben? Betrachten Sie hierfür sehr große positive und negative Zeitdifferenzen!

Verständnisaufgabe:

Warum führt beim Modell (4.101) eine gleichzeitige Vorzeichenumkehr von β_3 und β_4 zu denselben Modellaussagen?

4.10.2 Reales Beispiel: Wahl der Mensa bzw. Imbissbude

Nach der Vermutung des Autors ist im Falle einer *subjektiven Bewertung* objektiv unterscheidbarer Alternativen eine Erhöhung der Sensitivität in der Nähe des Referenzpunktes (im Sinne von Kahnemann und Tversky) wahrscheinlicher als eines Schwelle, welche durch eine reduzierte oder verschwindende Sensitivität ("Indifferenzbereich") charakterisiert wäre. Dies ist zu unterscheiden von der *objektiven Wahrnehmung* von Unterschieden, welcher auf Grund von physiologischen Schwellen und/oder endlichem Schätzvermögen Grenzen gesetzt sind. Auf einen konkreten Sachverhalt bezogen, kann dies beispielsweise durch folgende Hypothese formuliert werden:

Hypothese:

- Sind die Zeitdifferenzen bekannt, ist die Option, fünf mal eine Minute zu sparen, attraktiver als einmalig 5 Minuten.
- Muss man die Zeitdifferenzen selbst schätzen, erkennen die Probanden bei 10 gegenüber 11 Minuten keinen Unterschied, wohl aber bei 10 gegenüber 15 Minuten, so dass nun die Option, einmalig 5 Minuten zu sparen, attraktiver ist als fünf mal eine Minute.

4 Modelle diskreter Entscheidungen

Daraus folgt, dass bei Stated-Choice-Untersuchungen (übrigens auch die empirische Grundlage von Kahnemann & Tversky) eher eine Sensitivitätserhöhung zu erwarten ist und bei Revealed-Choice-Erhebungen eher eine Schwelle.

Diese Hypothesen empirisch zu testen ist aber komplizierter als obiges Einfach-Beispiel, da bei Stated-Choice-Bedingungen bei einem einzigen Kriterium trivialerweise immer eine Alternative **dominant** gegenüber den anderen ist: Beeinflusst nur die Reisezeit und nichts anderes die Güte der Alternativen, ist die Alternative mit der kürzesten Reisezeit die beste und wird wohl von allen vernünftig Räsonierenden gewählt. Man benötigt also mindestens ein weiteres Kriterium wie unterschiedliche Verkehrsmittel oder Kosten. Ferner ist es sinnvoll, die Choice-Sets möglichst gut zu "mischen", so dass die Probanden in der Befragungssituation keine Systematik erkennen können, die das Ergebnis verfälschen könnte. Konkret wurden unter Studenten folgende Stated-Choice Befragung durchgeführt:

Mensawahl

“Welche Mensa oder Imbissbude würden Sie für das Mittagessen wählen? Folgende Situationen unterscheiden sich nur durch die (immer zu Fuß durchgeführte) Zugangszeit und die Kosten der Mahlzeit. Die Qualität des Essens ist jeweils identisch. Ihre Freunde bzw. ‘Mitesser’ gehen gleich gerne zu beiden Zielen.”

Choice Set	Ziel 1	Ziel 2	Wahl Ziel 1	Wahl Ziel 2
1	1 Min., 3.10 €	2 Min., 3.00 €	1	28
2	5 Min., 2.70 €	3 Min., 2.80 €	7	20
3	2 Min., 3.70 €	9 Min., 3.20 €	8	21
4	5 Min., 3.60 €	9 Min., 3.50 €	28	1
5	10 Min., 3.00 €	1 Min., 3.50 €	5	24
6	6 Min., 2.70 €	1 Min., 2.20 €	0	29
7	1 Min., 2.00 €	3 Min., 1.50 €	0	29
8	2 Min., 2.90 €	5 Min., 2.80 €	24	5
9	3 Min., 3.20 €	4 Min., 3.10 €	6	23
10	10 Min., 2.60 €	6 Min., 3.10 €	19	10
11	7 Min., 2.85 €	4 Min., 3.35 €	26	3

Als Referenzmodell wird zunächst folgendes *lineare* Logit-Modell M_1 mit den generisch formulierten Charakteristika Kosten K_{ni} und Zeiten T_{ni} anhand dieser Stated-Choice-Daten geschätzt:

$$V_{ni} = \beta_1 \delta_{i1} + \beta_2 K_{ni} + \beta_3 T_{ni} \quad (4.102)$$

4 Modelle diskreter Entscheidungen

Die Kalibrierung ergab folgende Parameterschätzer und zugehörigen maximalen Log-Likelihood-Wert:

$$\hat{\beta}_1 = -0.84 \pm 0.18, \quad \hat{\beta}_2 = -11.6 \pm 1.3, \quad \hat{\beta}_3 = -0.86 \pm 0.10, \quad \tilde{L} = -127.2.$$

Alle Werte sind signifikant. Die Sensitivitäten sind plausibel (sie sind negativ, eine NE entspricht $-1/\hat{\beta}_3 = 1.2$ Minuten bzw. $-1/\hat{\beta}_2 = 9$ Euro-Cent, Zeitwert etwa $60\hat{\beta}_3/\hat{\beta}_2 = 5$ Euro/h), aber die Alternativenspezifische Konstante sollte eigentlich nichtsignifikant sein: Anscheinend gibt es einen Malus für Alternative 1 allein dadurch, dass sie an erster Stelle steht.⁴⁴

Die obigen Hypothesen werden nun mit zwei verallgemeinerten Modellen überprüft, einem nichtlinearen und einem parameterlinearen:

Nichtlineares Modell M_2 :

$$\begin{aligned} V_{n1} &= \beta_1 + \Delta V_{\text{in}} + \beta_4 \tanh\left(\frac{\Delta V_{\text{in}}}{\beta_5}\right), & V_{n2} &= 0, \\ \Delta V_{\text{in}} &= \beta_2 \Delta K_n + \beta_3 \Delta T_n, \end{aligned} \quad (4.103)$$

wobei

$$\Delta T_n = T_{n1} - T_{n2}, \quad \Delta K_n = K_{n1} - K_{n2}. \quad (4.104)$$

Parameterlineares Modell M_3 :

$$V_{n1} = \beta_1 + \beta_2 \Delta K_n + \beta_3 \Delta x_{3n} + \beta_4 \Delta x_{4n}, \quad V_{n2} = 0 \quad (4.105)$$

wobei (vgl. Abb. 4.24)

$$\Delta x_{3n} = \begin{cases} \Delta T_n & |\Delta T_n| \leq 5 \text{ Minuten} \\ \text{sign}(\Delta T_n) (5 \text{ Minuten}) & \text{sonst,} \end{cases} \quad (4.106)$$

$$\Delta x_{4n} = \begin{cases} 0 & |\Delta T_n| \leq 5 \text{ Minuten} \\ \text{sign}(\Delta T_n) (|\Delta T_n| - 5 \text{ Minuten}) & \text{sonst} \end{cases} \quad (4.107)$$

mit der Vorzeichenfunktion $\text{sign}(x) = -1$ falls $x < 0$ und $= 1$ falls $x \geq 0$. Diese Modelle haben zwei konzeptionelle Unterschiede:

- In Modell M_2 werden erst Zeit und Kosten zu einer linearen Nutzendifferenz ΔV_{in} "verrechnet", welche danach als Einheit nichtlinear bewertet wird. Hingegen wird im Modell M_3 Geld linear bewertet und die nichtlineare Bewertung bezieht sich rein auf die Zeit.
- Im nichtlinearen Modell ist auch die Breite der Nichtlinearität schätzbar, während im parameterlinearen Modell diese ad-hoc angenommen werden muss (hier wurde die Hälfte des maximalen Zeitunterschieds gewählt).⁴⁵

4 Modelle diskreter Entscheidungen

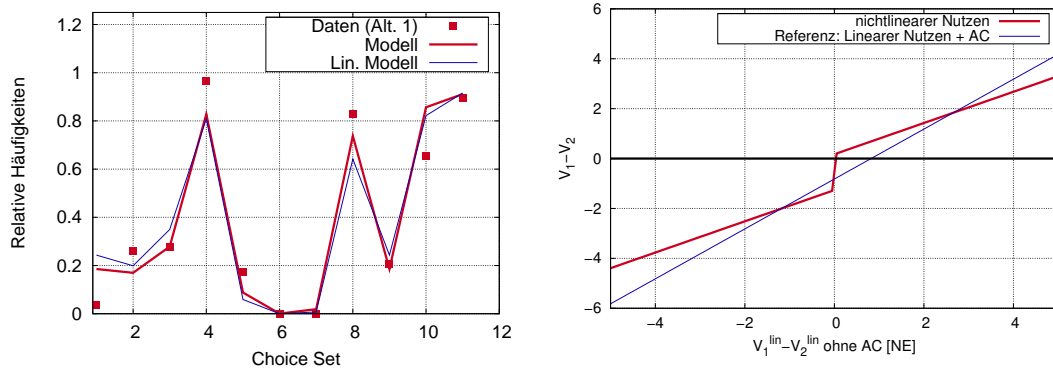


Abbildung 4.22: Parameterschätzung des nichtlinearen binären Logitmodells (4.103) an die Stated-Choice-Ergebnisse der Tabelle *Mensawahl*. Links: Modellierungsgüte; rechts: Nutzendifferenz in Abhängigkeit der verrechneten linearen Nutzendifferenz.

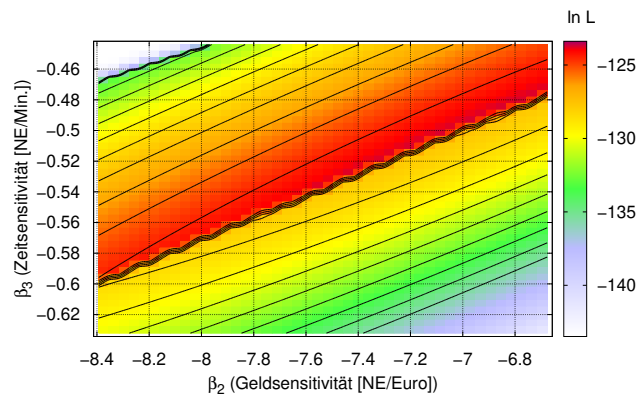


Abbildung 4.23: Ausschnitt der "Optimierungslandschaft", wenn man das nichtlineare Modell (4.103) anhand der Stated-Choice-Ergebnisse der Tabelle *Mensawahl* schätzt. Gezeigt sind die Dimensionen "lineare Zeit- und Geldsensitivitäten", während die anderen drei Parameter auf die geschätzten Werte gesetzt wurden.

4 Modelle diskreter Entscheidungen

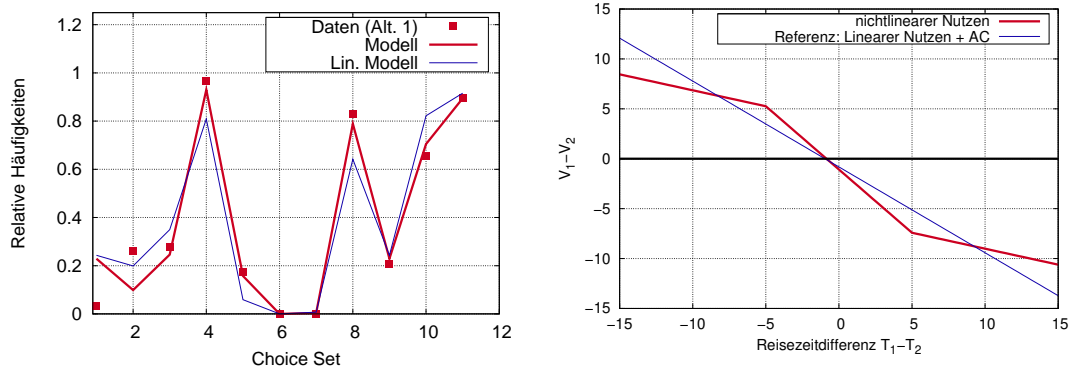


Abbildung 4.24: Parameterschätzung des parameterlinearen binären Logitmodells (4.105) an die Stated-Choice-Ergebnisse der Tabelle *Mensawahl*. Links: Modellierungsgüte; rechts: Zugangszeitabhängigkeit der Nutzendifferenz bei gleichen Kosten.

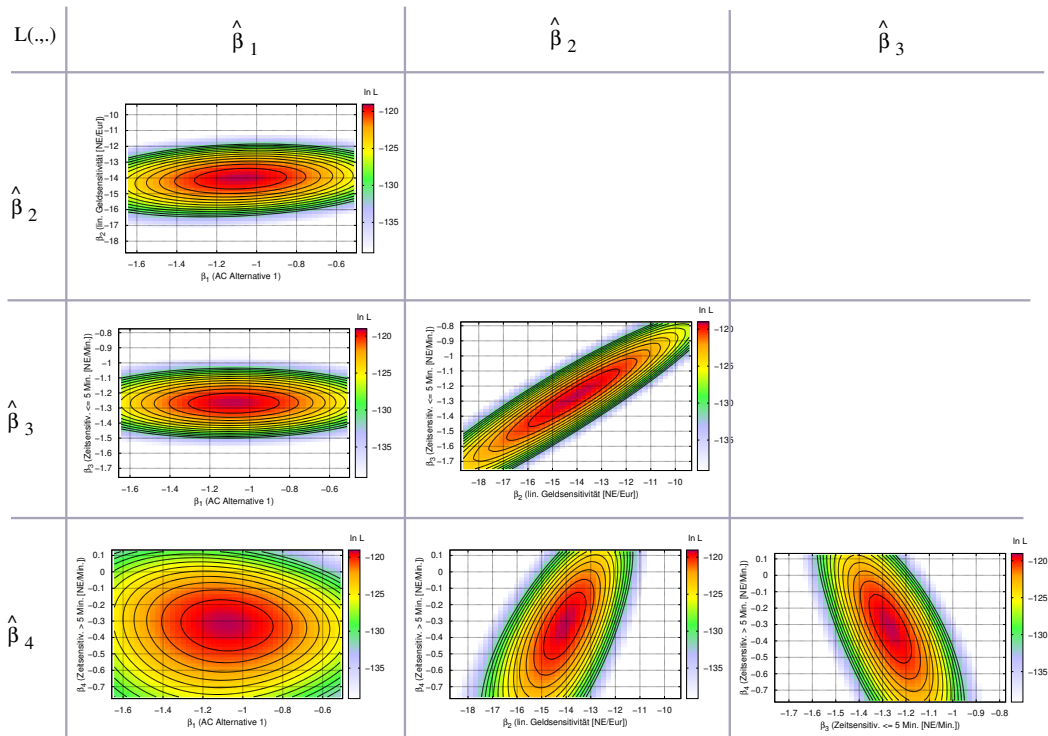


Abbildung 4.25: "Optimierungslandschaft" der Log-Likelihoodfunktion des parameterlinearen Logitmodells (4.105) bezüglich der Daten der Tabelle *Mensawahl*. Die jeweils nicht variierten Parameter wurden auf die Schätzwerte gesetzt.

4 Modelle diskreter Entscheidungen

Abbildung 4.22 zeigt ein Problem der parameter-nichtlinearen Modellierung. Das beste Ergebnis, d.h. die maximale Log-Likelihood, wird bei eigentlich unplausiblen Parameterkombinationen erreicht. Hier ergibt sich

$$M_2 : \begin{aligned} \hat{\beta}_1 &= -0.55 \pm 0.18, \\ \hat{\beta}_2 &= -7.5 \pm 0.3, \\ \hat{\beta}_3 &= -0.54 \pm 0.03, \\ \hat{\beta}_4 &= 0.7 \pm 0.2, \\ \hat{\beta}_5 &= 0.002 \pm 0.509, \\ \tilde{L} &= -123.5, \end{aligned}$$

d.h. die Breite β_5 der Nichtlinearität geht gegen null, während die Höhe $\beta_4 = 0.7$ NE endlich bleibt (vgl. rechte Grafik der Abbildung 4.22). Der Plot eines Ausschnitts der “Optimierungslandschaft” $\tilde{L}(\beta)$, beispielsweise bezüglich der linearen Geld- und Zeitsensitivität β_2 bzw. β_3 vor der nichtlinearen Bewertung (Abb. 4.23) zeigt zudem, dass das Optimum an einer Unstetigkeit der “Optimierungslandschaft” realisiert wird und damit den “verdächtig” niedrigen Standardabweichungen der zugehörigen Schätzer keine Bedeutung beigemessen werden darf.

Die Schätzung des parameterlinearen Modells M_3 , Gl. (4.105), an die Daten der Tabelle *Mensawahl* ergibt

$$M_3 : \begin{aligned} \hat{\beta}_1 &= -1.08 \pm 0.19, \\ \hat{\beta}_2 &= -14.0 \pm 1.6, \\ \hat{\beta}_3 &= -1.27 \pm 0.16, \\ \hat{\beta}_4 &= -0.32 \pm 0.15, \\ \tilde{L} &= -119.1. \end{aligned}$$

Dies sowie die Abbildungen 4.24 und 4.25 zeigen, dass die parameterlineare Modellierung in jeder Hinsicht zu bevorzugen ist:

- M_3 enthält nur vier Parameter, während M_2 fünf Parameter umfasst,
- Die “Fitgüte” ist besser, d.h. $\tilde{L}(\hat{\beta})$ größer,
- Die “Optimierungslandschaft” um $\hat{\beta}$ ist stetig und differenzierbar und es ist garantiert, dass man ein globales Maximum erreicht hat (Abb. 4.25).

Der Blick auf die rechte Grafik der Abb. 4.24 zeigt, dass, abgesehen von den ersten beiden Choice Sets, die relativen Häufigkeiten vom Modell nahezu 1:1 reproduziert werden.⁴⁶ Insbesondere erhält man signifikante Ergebnisse: Zwar ist die Signifikanz von $\hat{\beta}_4$ bezüglich

⁴⁴Es wurde häufig nur die erste Alternative abgefragt und die Wahlhäufigkeit der zweite Alternative durch die Zahl der verbleibenden Studierenden ergänzt. Jeder nicht aktive Student wird so Alternative 2 zugeschlagen, wodurch der Malus für Alternative 1 plausibel ist.

⁴⁵Natürlich könnte man auch im parameter-nichtlinearen Modell die Nichtlinearität rein auf die Zeitdifferenz beschränken und im parameterlinearen Modell die Bewertung wie in M_2 zweistufig durchführen, was hier aus Platzgründen unterlassen wird.

⁴⁶Im Choice Set 1 verhalten sich die Befragten aber nicht ganz konsistent: In diesem Set wählt nur einer Alternative 1 während im weitgehend identischen Set 9 sechs Befragte Alternative 1 wählen.

4 Modelle diskreter Entscheidungen

null nur marginal ($p = 1 - \Phi(0.32/0.15) = 4\%$). Aber, und das ist entscheidend, die Nullhypothese

$$H_0 : \beta_3 = \beta_4,$$

oder, gleichbedeutend, die Nullhypothese

$$H'_0 : \text{Modell } M_1 \text{ beschreibt die Daten gleich gut wie Modell } M_3$$

kann hochsignifikant ($p < 0.1\%$) widerlegt werden, wie wir mit den Mitteln des folgenden Abschnitts 4.11 zeigen können.

Damit zeigt die Tabelle *Mensawahl* in der Tat eine signifikante *Erhöhung* der Zeitsensitivität in der Nähe von $\Delta T = 0$. Der Vergleich der zwei Sensitivitäten $\hat{\beta}_3$ und $\hat{\beta}_4$ des Modells M_3 mit der geschätzten Sensitivität $\hat{\beta}_3^{M1}$ (Steigungen in Abb. 4.24 rechts) verdeutlicht diese Erhöhung der Sensitivität anschaulich.

Als Nachteil bleibt die Ad-Hoc-Annahme der Grenzen des Intervalls erhöhter Sensitivität. Eine Modellierung mit ± 3 oder ± 7 Minuten anstelle von ± 5 Minuten ergab aber ähnliche Ergebnisse. Wir kommen zum Ergebnis, dass, wann immer möglich, Nichtlinearitäten durch parameterlineare Modelle anstelle von irreduzibel nichtlinearen Modellen formuliert werden sollten.

Verständnisfrage:

Argumentieren Sie, dass im Modell M_3 eine Schwellenbreite von ± 10 Minuten dieses Modell auf M_1 reduzieren würden, wenn man es auf die Tabelle *Mensawahl* anwendete: Die Schätzer für β_1 bis β_3 wären dann gleich, während β_4 nicht mehr geschätzt werden kann, so dass das Modell bezüglich dieser Daten fehlspezifiziert wäre.

Verständnisfrage:

Gibt es in der Tabelle *Mensawahl* ein Choice Set mit einer dominanten Alternative? Wenn ja, welches?

4.11 Modellvergleich: Likelihood-Ratio-Test

Bei den Regressionsmodelle hatten wir den F-Test und den T^2 -Test zur Bestimmung der Relevanz von exogenen Variablen und zugehöriger Modellparameter kennengelernt. Die Frage war: Erklärt der bzw. die zusätzliche(n) Parameter des komplizierteren Modells den Sachverhalt signifikant besser als ein einfacheres Modell ohne diese Parameter, oder wird das Modell durch die zusätzlichen Parameter fehlspezifiziert? Bei Modellen der diskreten Wahltheorie übernimmt diese wichtige Aufgabe der **Modellspezifikation** der *Likelihood-Ratio-Test* (LR-Test).⁴⁷

⁴⁷Manchmal wird zwischen "Likelihood-Ratio-Test" und "Maximum Likelihood-Ratio-Test" unterschieden: Der Likelihood-Ratio-Test vergleicht das geschätzte volle Modell gegen das *nicht* geschätzte Tri-

4 Modelle diskreter Entscheidungen

Der LR-Test ist zum Vergleich zweier *verschachtelter* Modelle anwendbar, das heißt, das einfachere oder **restringierte** Modell $M^{(r)}$ mit M_r Parametern ist im komplexeren **unrestringierten** Modell M mit $M > M_r$ Parametern enthalten. Im letzten Abschnitt über nichtlineare Nutzenfunktionen sind beispielsweise die Modelle M_1 , M_2 und M_4 , nicht jedoch M_3 , restringierte Modelle bezüglich des durch (4.101) beschriebenen vollen nichtlinearen Modells.

- Typischerweise werden im restringierten Modell einige Parameter des unrestringierten Modells auf feste Werte, häufig auf null, gesetzt (daher der Name “restringiert” = eingeschränkt).
- Setzt man alle Parameter=0, erhält man das **Trivialmodell**: Bei quasilinearen stetigen Modellen (Regressionsmodellen) lautet dieses $\hat{y} = 0$ und bei diskreten Wahlmodellen mit I Alternativen $V_{ni} = 0$ bzw. $P_{ni} = 1/I$.
- Eine weitere Variante von Einfachstmodellen ist das **Konstantenmodell**: Bei den quasilinearen Regressionsmodellen lautet diese $\hat{y} = \beta_0$ mit dem LSE oder ML-Schätzer $\hat{\beta}_0 = \bar{y}$. Bei diskreten Wahlmodellen mit I Alternativen entspricht dies der Nutzenfunktion (siehe Abschnitt 4.7.7)

$$V_{ni} = \sum_{j=1}^{I-1} \beta_j \delta_{ji}, \quad (4.108)$$

die einen kompletten Satz von alternativenspezifischen Konstanten (ACs) repräsentiert, und das nach der ML-Methode geschätzte Modell führt sowohl bei Logit- als auch bei Probitmodellen zu den Auswahlwahrscheinlichkeiten $P_{ni} = f_i$ (Wahrscheinlichkeiten gleich den beobachteten relativen Häufigkeiten), unabhängig von der Wahlentscheidung n (siehe Abschnitt 4.7.7).

4.11.1 Vorgehen

Wie der Name schon sagt, wird beim LR-Test zum Bilden der Testvariable das Verhältnis der Likelihood-Werte der geschätzten Modelle, genauer der Logarithmus davon, verwendet: Das Vorgehen ist analog zu den sonstigen Tests:

1. **Nullhypothese H_0** : Die zusätzlichen Parameter des unrestringierten Modells sind irrelevant, tragen also nichts zur Erklärung des Sachverhalts bei
2. **Test-Statistik**: Die Test-Variable χ^{LR} (logarithmiertes Likelihood-Ratio)

$$\chi^{LR} = 2 \ln \left(\frac{L(\hat{\beta})}{\tilde{L}(\hat{\beta})} \right) = 2 [\tilde{L}(\hat{\beta}) - \tilde{L}(\check{\beta})] \sim \chi^2(M - M_r). \quad (4.109)$$

vialmodell (z.B. “alle Wahrscheinlichkeiten sind gleich”), während der Maximum Likelihood-Ratio-Test zwei geschätzte verschachtelter Modelle vergleicht. In dieser Diktion würde hier der “Maximum Likelihood-Ratio-Test” vorgestellt. Dies ist aber spitzfindig, da es beim Trivialmodell nichts zu schätzen gibt und es deshalb auch als “geschätztes Modell” durchgehen würde. Insbesondere ist die technische Durchführung des Tests immer gleich.

4 Modelle diskreter Entscheidungen

gehört bei Gültigkeit der Nullhypothese *asymptotisch* einer χ^2 -Verteilung mit $M - M_r$ Freiheitsgraden: Pro zusätzlichen Parameter (Parameterzahl M) des unrestringierten Modells ein Freiheitsgrad.

3. **Bestimmung der Realisierung $\lambda_{\text{data}}^{\text{LR}}$ von λ^{LR} :** Diese fällt als Nebenprodukt der ML-Schätzung automatisch mit ab. Hat das Modell M nur zwei Parameter, kann man $\lambda_{\text{data}}^{\text{LR}}$ auch direkt aus dem Plot der Log-Likelihood ablesen (vgl. das Beispiel auf S. 217 und die Aufgabe auf S. 222).

4. **Testergebnis:**

- Entweder klassisch mit einer vorgegebenen Fehlerwahrscheinlichkeit: H_0 wird bei einer Fehlerwahrscheinlichkeit α abgelehnt, d.h das unrestringierte Modell beschreibt die Daten signifikant besser als das restringierte, falls für die Realisierung gilt

$$\lambda_{\text{data}}^{\text{LR}} > \chi_{1-\alpha, M-M_r}^2, \quad (4.110)$$

- oder direkt mit dem p -Wert: Bei Zutreffen von H_0 erhält man nur mit der Wahrscheinlichkeit

$$p = 1 - F_{\chi^2}^{(M-M_r)}(\lambda_{\text{data}}^{\text{LR}}) \quad (4.111)$$

größere λ^{LR} -Werte als $\lambda_{\text{data}}^{\text{LR}}$. Wie bei den Regressionsmodellen stellt der p -Wert die minimale Fehlerwahrscheinlichkeit dar, bei der H_0 gerade noch abgelehnt werden kann.

Im Gegensatz zu den T^2 und F -Tests der Regressionsmodelle ist der LR-Test nur "asymptotisch" gültig, da nur dann λ^{LR} einer χ^2 -Verteilung gehorcht. "Asymptotisch" heißt, dass die zur exakten Reproduktion der Daten durch das Modell benötigte Information die Parameterzahl bei weitem übersteigt, so dass Fehler beim Schätzen der Varianz-Kovarianzmatrix vernachlässigbar werden. Dies entspricht in der Regressionsrechnung Tests bei bekannter Varianz: Beim T -Test geht die Student-t-Verteilung in eine Standardnormalverteilung über, beim T^2 -Test (als Spezialfall eines F -Tests für eine Parameterzahldifferenz von 1) in eine $\chi^2(1)$ -Verteilung. Konsistent dazu unterscheidet sich bei hinreichend hohem Datenumfang die Likelihoodfunktion an allen relevanten Stellen nur unwesentlich von der multivariaten Gaußverteilung, siehe Abschnitt 4.9.1, insbesondere Gl. (4.92).

Aufgabe T^2 Test einbauen!

Beispiel

Betrachtet wird eine auf zwei Alternativen aggregierte Revealed-Choice-Befragung wie in Abschnitt 4.7.3 und eine Modellierung durch das dort ausführlich diskutierte binomiale Logit-Modell mit den deterministischen Nutzenfunktion

$$V_{ni}(x) = (\beta_1 r_i + \beta_2) \delta_{i1}$$

4 Modelle diskreter Entscheidungen

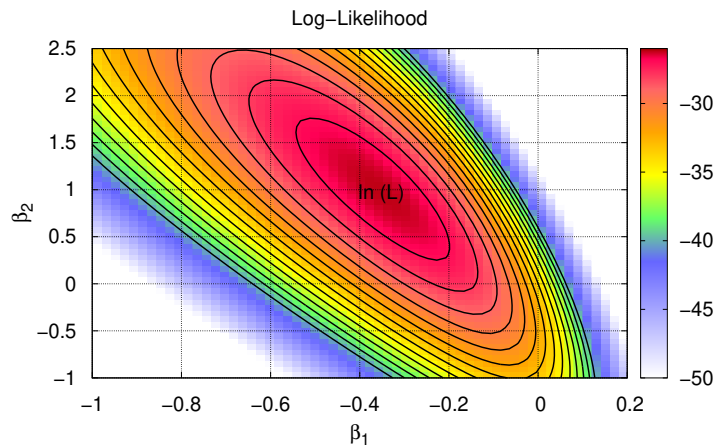


Abbildung 4.26: Log-Likelihood des gerechneten Beispiels zum LR-Test.

mit der Weglänge r als einziger exogener Faktor, nur dass hier die Daten nur einer Vorlesungsbefragung herangezogen werden. Aufgrund der kleineren Fallzahlen sind die Parameterschätzer nicht so hoch signifikant wie dort und die Auswertung interessanter.

Abbildung 4.26 zeigt die Log-Likelihood dafür, dass das Modell genau die Daten der Tabelle 4.3 liefert. LR-Tests lassen sich hier einfach graphisch durchführen (Höhenlinienabstand 1, maximale Log-Likelihood -26.2 , "höchste" Höhenlinie bei $\tilde{L} = \ln L = -27$)

Test 1: Ist die Entfernung ein signifikanter Einflussfaktor?

1. Nullhypothese H_0 : Die Entfernung ist nicht relevant, das restringierte Modell M^r : $V_{ni} = \beta_2 \delta_{i1}$ beschreibt die Daten nicht schlechter bzw. $\beta_1 = 0$.
2. Test-Statistik: Da das volle Modell M zwei Parameter und das restringierte einen Parameter enthält, ist

$$\chi^R = 2 \left[\tilde{L}(\hat{\beta}_1, \hat{\beta}_2) - \tilde{E}(\hat{\beta}_2) \right] \sim \chi^2(1).$$

3. Realisierung von χ^R aus den Daten bzw. der Log-Likelihood-Funktion:
 - Das volle Modell wird durch $\hat{\beta}_1 = -0.35$ und $\hat{\beta}_2 = 1$ geschätzt. Die zugehörige Log-Likelihood ist laut Angabe $\tilde{L}(\hat{\beta}_1, \hat{\beta}_2) = -26.2$.
 - Zur Schätzung des restringierten Modells wird die Log-Likelihood auf der Linie $\beta_1 = 0$ (der Restriktion) bezüglich β_2 maximiert. Dies ergibt $\hat{\beta}_2 = -0.5$ und die dazugehörige Log-Likelihood $\tilde{E}(\hat{\beta}_2)$ von -32 (einfach die Höhenlinien zählen)

4 Modelle diskreter Entscheidungen

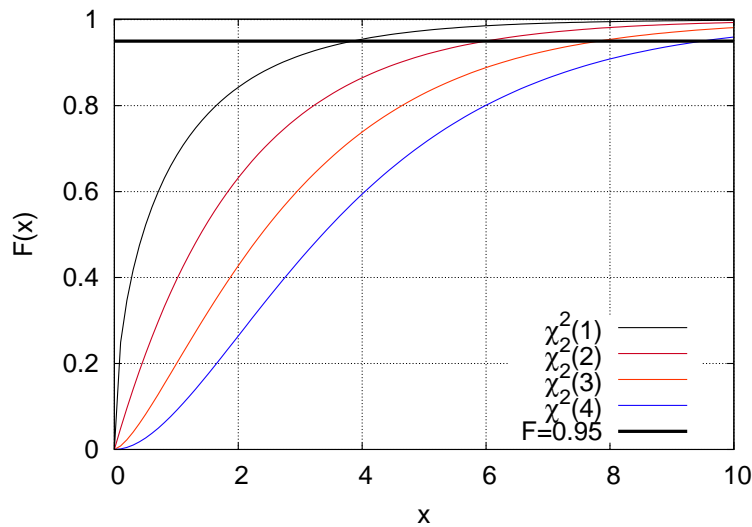


Abbildung 4.27: Verteilungsfunktion der $\chi^2(m)$ Verteilung für verschiedene Freiheitsgrad-Zahlen m .

Damit

$$\lambda^{\text{LR}} = 2(-26.2 + 32) = 11.6$$

- Entscheidung: Hier wird die Realisierung mit der Verteilungsfunktion der $\chi^2(1)$ -Verteilung verglichen (vgl. Abb. 4.27). Dabei liegt 11.6 schon außerhalb des Darstellungsbereichs der Grafik. Aber selbst bei $\lambda^{\text{LR}} = 10$ ist die Kurve für einen Freiheitsgrad schon oberhalb 0.99. Also

$$F_{\chi^2}^{(1)}(11.6) > 0.99 \Rightarrow p < 1\%$$

Der p -Wert ist also kleiner als 1% (er ist sogar kleiner als 1 Promille), also ist die Entfernung ein wichtiger Einflussfaktor und der zugehörige Parameter β_1 hochsignifikant von null verschieden. Der LR-Test geht also zugunsten des vollen Modells M aus.

Test 2: Ist die globale Bevorzugung (alternativenspezifische Konstante) ein signifikanter Einflussfaktor?

- Nullhypothese H_0 : Die Bevorzugung ist nicht relevant und das restringierte Modell $M^{(p)}$: $V_{ni}(x) = \beta_1 x \delta_{i1}$ beschreibt die Daten nicht schlechter als das volle Modell. Anders ausgedrückt: $\beta_2 = 0$.
- Test-Statistik: Wie beim Test 1
- Realisierung von λ^{LR} :

4 Modelle diskreter Entscheidungen

- Das volle Modell wurde bereits beim Test 1 geschätzt: $\hat{\beta}_1 = -0.35$, $\hat{\beta}_2 = 1$ und $\tilde{L}(\hat{\beta}_1, \hat{\beta}_2) = -26.2$.
- Zur Schätzung des restringierten Modells wird die Log-Likelihood auf der Linie $\beta_2 = 0$ bezüglich β_1 maximiert. Dies ergibt $\hat{\beta}_1 = -0.17$ und die dazugehörige Log-Likelihood $\tilde{L}(\hat{\beta}_2)$ von -27.7 (das Maximum ist deutlich näher an der Höhenlinie $\tilde{L} = -28$ als an der Linie $\tilde{L} = -27$).

Damit

$$\chi^{\text{LR}} = 2(-26.2 + 27.8) = 3.2$$

4. Entscheidung: Hier wird die Realisierung mit der Verteilungsfunktion der $\chi^2(1)$ -Verteilung verglichen (vgl. Abb. 4.27). Es gilt

$$F_{\chi^2}^{(1)}(3.2) \approx 0.92 \Rightarrow p = 8\%$$

Der p -Wert ist also 8%, was in der Regel nicht mehr als signifikant betrachtet wird. Lässt man z.B. eine Fehlerwahrscheinlichkeit $\alpha = 5\%$ zu und liest aus der Grafik das Quantil $\chi_{1-\alpha,1}^2 = 4$ ab, so ist die Realisierung χ^{LR} kleiner als das Quantil, der Test also nicht widerlegbar. Per Definition ist die geringste Fehlerwahrscheinlichkeit, bei der H_0 gerade noch widerlegbar ist, gleich $p = 8\%$.

Der LR-Test geht bei 10% Fehlerwahrscheinlichkeit zugunsten des Vollen Modells M , bei 5% Fehlerwahrscheinlichkeit zugunsten des restringierten Modells $M^{(r)}$. Der Einfluss der globalen Bevorzugung ist also grenzwertig. Generell sollte man beim LR-Test im Zweifelsfall das komplexere Modell beibehalten und erst bei p -Werten oberhalb von 10% das einfachere in Betracht ziehen.

Verständnisaufgabe:

Wie ist es zu erklären, dass im restringierten Modell $M^{(r)}$ (vgl. den Test 1) $\beta_2 < \beta_2$, der globale Bonus der Alternative 1 (Fuß/Rad) gegenüber dem vollen Modell wesentlich kleiner ist (insbesondere negativ statt positiv)? Warum ist beim restringierten Modell $M^{(r)}$ des Tests 2 $\beta_1 > \beta_1$, die Zeitbewertung der "langsameren" Verkehrsmittel also nicht ganz so negativ wie beim vollen Modell M

Da das (bezüglich der relevanten Faktoren) fehlspezifizierte Modell $M^{(r)}$ keine Zeitbewertung enthält, muss die offensichtliche Unattraktivität der langsameren Modi bei großen Entfernungen über den einzig verbleibenden Freiheitsgrad der globalen Bewertung modelliert werden. Diese wird dadurch negativ ($\beta_2 < 0$), obwohl eigentlich Fuß/Rad gegenüber ÖV/MIV bei *ceteris-paribus* Bedingungen einen Bonus $\beta_2 > 0$ besitzen. Analog wird im Modell \tilde{M} die fehlende Möglichkeit der Berücksichtigung einer globalen Bevorzugung der langsameren Verkehrsmittel über eine entsprechend weniger negative Entfernungsbewertung "verrechnet".

4.11.2 Test auf Relevanz der Einflussfaktoren mit dem Top-Down-Ansatz

Eine der wichtigsten Anwendungen ist die Eliminierung irrelevanter Parameter, da sonst das Modell nicht korrekt spezifiziert ist. Dabei kann man z.B. nach folgenden Top-Down-Ansatz vorgehen:

4 Modelle diskreter Entscheidungen

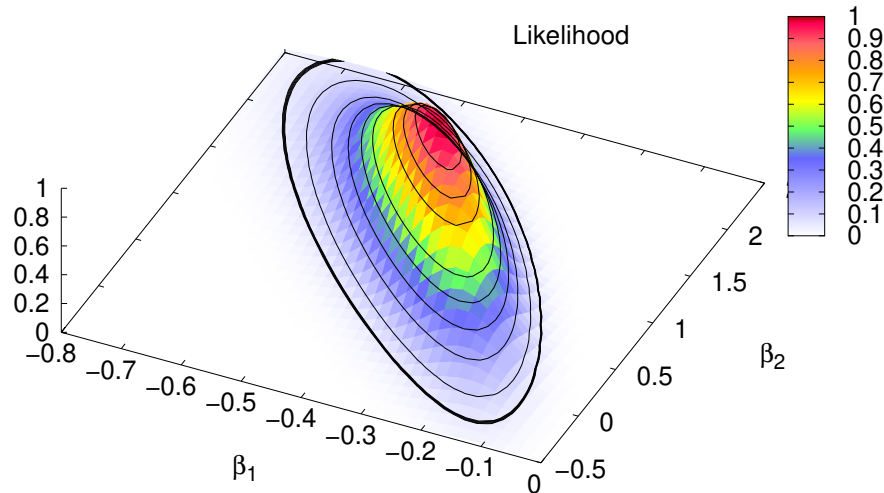


Abbildung 4.28: Der Quotient $L(\beta_1, \beta_2)/L_{\max}$, welcher direkt proportional zur zweidimensionalen Dichtefunktion der tatsächlichen Parameterwerte ist. Die Konfidenzregion des Likelihood-Ratio-Test bezüglich des Trivialmodells $V_{ni} = 0$ zur Fehlerwahrscheinlichkeit $\alpha = 5\%$ ist dick umrandet. Der Rand entspricht einer Log-Likelihood-Differenz von $\frac{1}{2} \chi_{2,1-\alpha}^2 \approx 3$, also $L(\beta_1, \beta_2)/L_{\max} = e^{-3}$.

- Man stellt ein komplexes Modell M auf, welches alle Einflussfaktoren enthält, welche aufgrund der Daten messbar sind und die man aufgrund theoretischer Überlegungen für sinnvoll hält.
- Dieses Modell wird nacheinander mit restringierten Modellen verglichen, bei denen jeweils ein Parameter =0 (oder auf einen aus theoretischen Überlegungen kommenden festen Wert) gesetzt wird.
- Der Parameter mit dem höchsten p -Wert wird gestrichen und das resultierende Modell wird das neue unrestringierte Modell.
- Das Verfahren wird so lange fortgesetzt, bis kein Parameter mehr p -Werte oberhalb von z.B. $\alpha = 5\%$ hat.

Dabei muss man allerdings auf Abhängigkeiten achten, da die Signifikanz eines Parameters durchaus von der Existenz anderer Parameter abhängt und im Extremfall eine Gruppe von Einflussfaktoren relevant sein kann, auch wenn es jeder einzelne nicht ist. Dies wird geprüft, indem man einen LR-Modellvergleich des durch die Top-Down-Methode resultierenden Modells mit dem ursprünglichen "komplexen" Ausgangsmodell durchführt.

4 Modelle diskreter Entscheidungen

Aufgabe: Top-Down-Ansatz

Testen Sie die Relevanz beider Parameter der Binomial-Logit und Probitmodelle mit deterministischen Nutzenfunktionen gemäß (4.64) bei der Beschreibung der in Tabelle 4.1 angegebenen Daten. Führen Sie den Likelihood-Ratio-Test graphisch anhand Abb. 4.12 durch.

Aufgabe: Trivialmodelle

Zeigen Sie, dass die Auswahlwahrscheinlichkeiten der Trivialmodelle der diskreten Wahlmodelle ($V_{ni} = 0$) durch $P_{ni} = 1/I$ gegeben sind.

Aufgabe: Konstantenmodelle

Zeigen Sie für das multinomiale Logit und Probit-Modell, dass die ML-Schätzung der Konstantenmodelle (4.74) der diskreten Wahltheorie zu Auswahlwahrscheinlichkeiten gleich den relativen Häufigkeiten führt.

Hinweis: Betrachten Sie die Wahrscheinlichkeiten direkt.

4.12 Modellqualität: Goodness-of-Fit Maße

Der LR-Test eignet sich gut zur Eliminierung überflüssiger Parameter und den Testen einzelner neuer Einflussfaktoren auf Signifikanz. Die z.B. nach einer Top-Down Elimination verbleibenden Variablen sind notwendig, aber ggf. nicht hinreichend für eine korrekte Modellspezifikation. Als Indiz, ob man noch wichtige Einflussfaktoren übersehen hat,⁴⁸ kann der Vergleich des Modells mit dem entsprechenden *Konstantenmodell* dienen.⁴⁹ Dabei sollte sich nicht nur Signifikanz beim LR-Test ergeben, sondern auch das Modell als solches eine hohe Güte haben.⁵⁰

Eine weitere Problemstellung, bei der der LR-Test nicht weiterhelfen kann, ist der Qualitätsvergleich zweier *nichtverschachtelter* Modelle: Welches ist besser und ist der Unterschied signifikant? Man benötigt also ein absolutes Maß für die Güte eines Modells. Mehrerer solcher Maße wurden vorgeschlagen und in der üblichen Statistik-Software implementiert (M bezeichnet die Zahl der zu schätzenden Parameter und n die Zahl der Beobachtungen, $\tilde{L} = \ln L$ die Log-Likelihood des kalibrierten Modells):

⁴⁸Faktoren aufzuspüren ist sehr viel schwieriger als welche zu eliminieren, da man nicht weiß, nach was man suchen soll!

⁴⁹Der Vergleich mit dem Trivialmodell kann bei Alternativen mit sehr geringen empirischen relativen Häufigkeiten irreführend sein, da dann das – eigentlich auch triviale – Konstantenmodell oft mehr erklärt als alle weiteren Parameter des vollen Modells.

⁵⁰Der LR-Test sagt nur aus, ob ein Modell signifikant besser als ein anderes ist, aber nichts über die absolute Güte: Besser ist nicht immer gut!

4 Modelle diskreter Entscheidungen

- **Akaike's information criterion:**

$$\text{AIC} = -2\tilde{L} + 2M \frac{n}{n - (M + 1)} \quad (4.112)$$

- **Bayesian information criterion:**

$$\text{BIC} = -2\tilde{L} + M \ln n \quad (4.113)$$

- **LR-Index** bzw. **McFaddens R2:**

$$\rho^2 = 1 - \frac{\tilde{L}}{\tilde{L}^0} \quad (4.114)$$

- **korrigierter LR-Index** bzw. **adjusted likelihood-ratio index** bzw. **McFaddens korrigiertes R2:**

$$\bar{\rho}^2 = 1 - \frac{\tilde{L} - M}{\tilde{L}^0} \quad (4.115)$$

Dabei ist \tilde{L}^0 die Log-Likelihood des Konstantenmodells oder, falls man das Gütemaß schwächer formulieren will, des Trivialmodells.⁵¹ Folgende Aussagen gelten bezüglich dieser Kenngrößen:

- Die beiden Informationskriterien geben die Information in Bit an, die man im Mittel *zusätzlich* benötigt, um von der Modellvoraussage zu den tatsächlichen Mikrodaten, d.h. den tatsächlich getroffenen Entscheidungen, zu kommen. Die beiden Kriterien haben geringfügig andere Annahmen bei der Herleitung, deshalb die unterschiedliche Form.
- Zur Modellauswahl kann man sowohl das AIC als auch das BIC heranziehen: Je niedriger, desto besser. Das BIC legt einen stärkeren Fokus auf parameter-sparsame (*parsimonious*) Modelle.
- Bei verschachtelten Modellen, welche der Nullhypothese des LR-Tests genügen, ist bei hinreichend viele Beobachtungen ($n \gg M$) der Erwartungswert des AIC gleich. Insbesondere entspricht das Ereignis "das erweiterte Modell hat ein kleineres AIC" einem p -Wert des LR-Tests < 0.5 . Insofern entspricht das AIC dem LR-Test, ist aber auch auf nichtverschachtelte Modelle anwendbar.
- Ein weiterer Test der Nullhypothese gleicher Güte zweier nichtverschachtelter Modelle ist der **Horowitz-Test**. Dieser wird hier aber nicht weiter behandelt.

⁵¹Dies ist analog wie bei der Gütemessung der Wettervorhersage: Das Trivialmodell wäre z.B. "Das Wetter wird schön" und das Konstantenmodell "Das Wetter ändert sich nicht".

4 Modelle diskreter Entscheidungen

- Der LR-Index bzw. der korrigierte LR-Index entspricht dem Bestimmtheitsmaß $R^2 = B = 1 - U$, Gl. (2.13), bzw. dem korrigierten Bestimmtheitsmaß \bar{R}^2 , Gl. (2.14), quasistetiger Regressionsmodelle, daher auch der Name “(korrigierter) McFadden R^2 ”: Während R^2 den erklärten Anteil der beobachteten Variation beschreibt, entspricht ρ^2 dem erklärten Anteil der zur exakten Reproduktion der Daten nötigen Log-Likelihood, wenn man vom Trivial- bzw. Konstantenmodell ausgeht. \tilde{L}/\tilde{L}_0 entspricht dann dem Anteil der nichterklärten Log-Likelihood. Im Sinne der Shannon’schen Definition entspricht ρ^2 und auch R^2 dem erklärten Anteil der in den Daten enthaltenen Informationsmenge.
- Die “korrigierten” Größen \bar{R}^2 und $\bar{\rho}^2$ berücksichtigen den “Verbrauch” von Daten bei der Fit-Prozedur und korrigieren die Tendenz zum Überfitten, falls n nicht $\gg M$ ist. Insbesondere liefert ein beliebig schlechtes Modell mit M Parametern, welches M Beobachtungen erklären soll, nach der Kalibrierung eine perfekte Übereinstimmung mit den Daten, also $R^2 = 1$ bzw. $\rho^2 = 1$, während \bar{R}^2 und $\bar{\rho}^2$ dann in der Nähe von Null oder undefiniert sind.
- Der korrigierte Likelihood-Ratio Index (korrigiertes McFadden R^2) $\bar{\rho}^2$ kann sowohl zur Modellauswahl (je höher, desto besser), also auch zur absoluten Gütebestimmung eines Modells angewandt werden. Die *Beurteilung* der *Güte* ist jedoch bei \bar{R}^2 und $\bar{\rho}^2$ konzeptionell unterschiedlich: Aufgrund der inhärenten Stochastizität hat selbst ein perfektes Auswahlmodell einen deutlich von 1 verschiedenen ρ^2 -Wert (siehe Beispiel weiter unten).
- Hat bei der Kalibrierung zweier verschachtelter Modelle mit vielen Daten ($n \gg M$) das erweiterte Modell ein “besseres” (höheres) ρ^2 , entspricht das $p < 0.5$ beim LR-Test. Dies macht die Eignung von ρ^2 zur Modellauswahl allgemeiner, auch nichtverschachtelter Modelle plausibel.

4 Modelle diskreter Entscheidungen

Beispiel: Zwei Alternativen, nur das Geschlecht ist relevant

Ist bei einer binomialen Wahlsituation für Frauen die beobachtete relative Häufigkeit, Alternative 1 zu wählen, $=0.1$, und für Männer $=0.9$, sind außerdem alle anderen Einflussfaktoren bei allen Personen gleich, so kann das beste Modell nicht besser sein, als für Frauen $P_{n1} = 0.1$ und $P_{n2} = 0.9$ vorauszusagen (und bei Männern umgekehrt). Dadurch erniedrigt sich die Log-Likelihood bei jeder Wahlentscheidung n (egal ob Mann oder Frau) *im Mittel* um

$$\Delta \tilde{L}_n = \sum_i y_{ni} \ln P_{ni} = 0.1 \ln 0.1 + 0.9 \ln 0.9 = -0.325$$

während im Nullmodell ($P_{ni} = 1/2$) die mittlere Änderung pro Entscheidung durch

$$\Delta \tilde{L}_n^{(0)} = \sum_i y_{ni} \ln P_{ni} = 0.1 \ln 0.5 + 0.9 \ln 0.5 = \ln 0.5 = -0.693$$

gegeben ist. Damit hat in dieser Situation selbst das beste Modell nur einen ρ^2 -Wert von

$$\rho^2 = 1 - \frac{\tilde{L} - M}{\tilde{Y}} = 0.531.$$

Aufgrund der Stochastizität der modellierten endogenen Variablen kann selbst das "beste" Modell die Mikro-Information nicht auch nur annähernd reproduzieren. Deshalb sind, je nach den beobachteten relativen Häufigkeiten, ρ^2 -Werte von 0.5 meist schon sehr gut, während $R^2 = 0.5$ ein ziemlich schlechtes Regressionsmodell charakterisiert.

Achtung: Als absolute Maße sind die Likelihood-Ratio-Indices durch *Quotienten* der Log-Likelihood formuliert, nicht durch *Differenzen* wie bei der Parameterschätzung oder der Testvariablen des LR-Tests. Fasst man $y_n > 1$ Beobachtungen zu Klassen n zusammen, müssen die sonst irrelevanten Multinomialkoeffizienten der Likelihoodfunktion berücksichtigt werden.

4.13 Generalized-Extreme-Value (GEV) Modelle

Die Modellklasse der Generalized-Extreme-Value Modelle (GEV Modelle) wurde von McFadden (1978) mit der Intention entwickelt, unter Beibehaltung der analytischen Berechenbarkeit der Auswahlwahrscheinlichkeiten möglichst reichhaltige Korrelationsstrukturen zwischen den Alternativen zuzulassen.

4.13.1 Motivation

Bei Entscheidungssituationen sind häufig mehrere Einzelentscheidungen so miteinander gekoppelt, dass sie sinnvollerweise nur gemeinsam modelliert werden können. Beispiele der Verkehrsökonomie sind

- *Tägliche Entscheidung von Einzelpersonen:* Zielwahl und Verkehrsmittelwahl, oder noch allgemeiner, Aktivitäten-, Ziel-, Verkehrsmittel- und Routenwahl. Insbesondere hängen die Ziel- und Verkehrsmittelwahl miteinander zusammen, da entferntere Ziele nur mit den "schnellen" Verkehrsmodi ÖPNV und MIV sinnvoll zu erreichen sind.
- *Langfristige Entscheidungen von Einzelpersonen oder Haushalten:* Wohnort- und Arbeitsplatzwahl: Es wäre unklug, den Wohnort ohne Berücksichtigung des zukünftigen Arbeitsplatzes oder den Arbeitsplatz ohne Berücksichtigung des Wohnortes auszuwählen.
- *Planungsentscheidung eines ÖPNV-Betriebs:* Auswahl einer Variante des zukünftigen Fahrplans und ggf. Kauf neuer Bahnen+Busse
- *Investitionsentscheidungen bei der Expansion einer Firma:* Neue Zweigstelle Ja / Nein? Ggf. Wahl der Region und des konkreten Ortes der neuen Zweigstelle.

Im einfachsten Fall fasst man alle Varianten der komplexen Entscheidung kombinatorisch zu einzelnen Alternativen zusammen und wendet die bisherigen Modelle, vorzugsweise das MNL-Modell, an. Bei einer Einkaufsentscheidung könnte man bei diesem Vorgehen beispielsweise eine Alternativenmenge der Form

$$A = \{(E, \text{ÖV}), (E, \text{MIV}), (D, \text{ÖV}), (D, \text{MIV})\} \quad (4.116)$$

(E für "Tante-Emma-Laden" und D für "Discounter") bekommen (vgl. Abb. 4.29 im Beispiel weiter unten).

Allerdings liegt es in der Natur geschachtelter Entscheidungen, dass die Zufallsnutzen, entgegen der bisherigen Modellspezifikation, nicht unabhängig voneinander sind. Beispielsweise haben die Alternativen $A_1 = (E, \text{ÖV})$ und $A_2 = (E, \text{MIV})$ den Ladentyp "Tante-Emma-Laden" als gemeinsames Merkmal und damit gemeinsame Zufallsnutzenkomponenten, nämlich alle dem Ladentyp zuzuordnenden Anteile des Zufallsnutzens. Damit sind für jede Person n die Zufallsnutzen $\epsilon_{n, \text{ÖV}}$ und $\epsilon_{n, \text{MIV}}$ positiv korreliert. Die Modellierung von Korrelationen ist daher häufig zwingend.⁵²

⁵²Am extremsten tritt dies beim schon erwähnten *Red-Bus, Blue-Bus Problem* zutage.

4 Modelle diskreter Entscheidungen

4.13.2 Allgemeine Formulierung der Modellklasse

Die GEV Modelle werden indirekt über eine beliebige “GEV-Funktion” $G(\mathbf{y}) = G(y_1, \dots, y_I)$ definiert, welche zunächst nur folgenden, rein formalen mathematischen Bedingungen genügen müssen:

$$G \text{ ist nicht-negativ: } G(\mathbf{y}) \geq 0 \text{ für alle } \mathbf{y}, \quad (4.117)$$

$$\text{Unendlich-Asymptotik von } G: G \rightarrow \infty \text{ falls ein } y_i \rightarrow \infty, \quad (4.118)$$

$$\text{Vorzeichen der Ableitungen: } G_i \equiv \frac{\partial G}{\partial y_i} \geq 0,$$

$$G_{ij} \equiv \frac{\partial^2 G}{\partial y_i \partial y_j} \leq 0 \text{ falls } i \neq j, \quad (4.119)$$

$$G_{ijk} \geq 0 \text{ usw.}$$

$$G \text{ ist homogene vom Grad 1:}^{53} G(\alpha \mathbf{y}) = G(\alpha y_1, \dots, \alpha y_I) = \alpha G(\mathbf{y}). \quad (4.120)$$

McFadden (1978) hat für beliebige Funktionen G , welche obigen Bedingungen genügen, folgendes gezeigt:

Aus G kann eine Verteilung erzeugt werden, und zwar die multivariate generalisierte Extremwertverteilung des Zufallsnutzenvektors ϵ mit der Verteilungsfunktion

$$F(\mathbf{e}) = \text{Prob}(\epsilon_1 \leq e_1, \dots, \epsilon_I \leq e_I) = \exp [-G(e^{-e_1}, \dots, e^{-e_I})], \quad (4.121)$$

dann folgen aus der allgemeinen Nutzenmaximierung (Homo Oeconomicus) des Gesamtnutzens $U_i = V_i + \epsilon_i$ die analytischen Auswahlwahrscheinlichkeiten

$$P_i = \frac{y_i G_i}{G} = \frac{e^{V_i} G_i(e^{V_1}, \dots, e^{V_I})}{G(e^{V_1}, \dots, e^{V_I})}, \quad (4.122)$$

wobei $G_i = \partial G / \partial y_i$ die partielle Ableitung von G nach dem i -ten Argument bezeichnet.

Man kann es auch kompakter formulieren:

Jede “GEV-Funktion” $G(\mathbf{y})$, welche den obigen vier Bedingungen genügt,

- erzeugt für eine vektorwertige Zufallsvariable ϵ eine generalisierte Extremwertverteilung mit der Verteilungsfunktion $F(\mathbf{e}) = e^{-G(\mathbf{y})}$, wobei $y_i = e^{-e_i}$ gilt,
- die Maximierung des Gesamtnutzens $V_i + \epsilon_i$ ergibt die Auswahlwahrscheinlichkeiten $P_i = \frac{y_i G_i(\mathbf{y})}{G(\mathbf{y})}$, wobei nun $y_i = e^{+V_i}$ gilt.

⁵³Allgemein muss nur ein Homogenitätsgrad > 0 gelten. Das verwirrt aber hier nur.

4 Modelle diskreter Entscheidungen

Begründung der Bedingungen an G

Nun werden auch die obigen Bedingungen an G klar:

- G muss ≥ 0 sein, da sonst $F(\boldsymbol{\epsilon}) > 1$ ist und damit keine Verteilungsfunktion sein kann.
- Falls ein $y_i \rightarrow \infty$, dann geht das entsprechende e^{-e_i} in (4.121) gegen unendlich und somit $e_i \rightarrow -\infty$. Dann ist aber $F = \text{Prob}(\epsilon_1 \leq e_1, \dots, \epsilon_I \leq e_I) = 0$, da die Wahrscheinlichkeit für das Ereignis $\epsilon_i < e_i$ gleich null ist und die verschiedenen Ereignisse in F logisch mit UND verknüpft sind. Also gilt $0 = \exp(-G)$ und somit $G \rightarrow \infty$.
- Damit $F(\boldsymbol{e})$ eine Verteilungsfunktion ist, muss sie nicht nur zwischen 0 und 1 liegen und =0 am "linken und unteren Rand" sein, sondern auch alle Dichtefunktionen, also die gemischten Ableitungen, nichtnegativ sein: $F_i = \partial F / \partial y_i \geq 0$, $F_{ij} = \partial^2 F / \partial y_i \partial y_j \geq 0$ für $i \neq j$ usw. Dies führt auf $\partial G / \partial y_i \equiv G_i \leq 0$, $G_{ij} \geq 0$ usw., also auf (4.119).
- Die Homogenitätsbedingung (4.120) hat nichts mit der Verteilungsfunktion zu tun, sondern mit der "Logit-Form" der analytischen Auswahlwahrscheinlichkeiten: Ist G homogen vom Grade 1, gilt $G(\boldsymbol{y}) = \sum_j y_j G_j$ (falls Zweifel, einfach die linke und rechte Seite von (4.120) nach α ableiten und danach $\alpha = 1$ setzen). Setzt man das in (4.122) ein, kann man die Auswahlwahrscheinlichkeiten (mit $y_i = e^{V_i}$) in "Logit-Form" schreiben:

$$P_i = \frac{y_i G_i}{G} = \frac{y_i G_i}{\sum_j y_j G_j} = \frac{e^{V_i} G_i}{\sum_j e^{V_j} G_j} = \frac{e^{V_i + \ln G_i}}{\sum_j e^{V_j + \ln G_j}} \quad (4.123)$$

Dies ist die MNL-Auswahlwahrscheinlichkeit, wenn man V_i durch den verallgemeinerten Nutzen $V_i + \ln G_i$ ersetzt.⁵⁴

Spezialfall Multinomial-Logit-Modell. Das einfachste GEV-Modell ist das Multinomial-Logit-Modell (MNL) selbst. Im Rahmen der GEV-Familie ist es definiert durch die GEV-Funktion

$$G(\boldsymbol{y})^{\text{MNL}} = \sum_{j=1}^I y_j. \quad (4.124)$$

Damit lautet nach (4.122) die Verteilungsfunktion der Zufallsnutzen

$$F(\boldsymbol{e}) = \exp[-G(e^{-e_1}, \dots)] = \exp\left(-\sum_j e^{-e_j}\right) = \prod_j \exp(-e^{-e_j}).$$

⁵⁴Dies ist kein Beweis für analytisch ausrechenbaren Wahrscheinlichkeiten, da man ja das Ergebnis (4.122) genutzt hat. Es zeigt aber sehr schön, dass die Auswahlwahrscheinlichkeiten des GEV immer logit-artig sind.

4 Modelle diskreter Entscheidungen

Dies ist ein Produkt von univariaten (0,1) Gumbelverteilungen, also sind die $\epsilon_i \sim$ i.i.d. Gumbel, was ja die definierende Eigenschaft des MNL ist. Die Auswahlwahrscheinlichkeiten ergeben sich mit (4.123) und $\ln G_i = \ln(\frac{\partial G}{\partial y_i}) = \ln 1 = 0$ zu

$$P_i^{\text{MNL}} = \frac{e^{V_i + \ln G_i}}{\sum_j e^{V_j + \ln G_j}} = \frac{e^{V_i}}{\sum_j e^{V_j}},$$

also dem bekannten Ausdruck.

4.13.3 Nested-Logit-Modell

Das bekannteste nichttriviale GEV-Modell ist das **Nested-Logit-Modell** (NL-Modell, von *nested*=geschachtelt). Im einfachsten zweistufigen Fall repräsentiert es eine hierarchisch gekoppelte komplexe Entscheidung für die Alternative

$$i = (l, m) : \quad (4.125)$$

- In der übergeordneten Entscheidung wird das **Nest** l gewählt, beispielsweise beim Einkaufen der Ladentyp l (z.B. Discounter oder "Tante Emma"),
- in der untergeordneten Entscheidung werden die Optionen innerhalb des Nests l gewählt, beispielsweise der Verkehrsmodus, mit welchem man den Laden l erreichen will.

Die GEV-Funktion des zweistufigen NL-Modells ist definiert durch

$$G^{\text{NL}}(\mathbf{y}) = \sum_{l=1}^L \left(\sum_{m=1}^{M_l} y_{lm}^{1/\lambda_l} \right)^{\lambda_l} \quad (4.126)$$

Jedes Nest l kann eine i.A. unterschiedliche Zahl $M_l \geq 1$ an untergeordneten Alternativen mit korrelierten Zufallsnutzensanteilen haben. Die Korrelation wird dabei durch $\lambda_l \in [0, 1]$ beschrieben.⁵⁵ Sind alle $\lambda_l = 1$, gibt es keine Korrelationen und das NL-Modell geht in das MNL-Modell über.

Verteilungsfunktion: Die Korrelationsstruktur des NL-Modells und insbesondere die Wirkung des Korrelationsparameters λ_l wird in der Verteilungsfunktion (4.121) deutlich:

$$F(\mathbf{e}) = \exp \left[- \sum_l \left(\sum_m e^{-e_{lm}/\lambda_l} \right)^{\lambda_l} \right] = \prod_l \exp \left[- \left(\sum_m e^{-e_{lm}/\lambda_l} \right)^{\lambda_l} \right] = \prod_l F_l(\mathbf{e}_l) \quad (4.127)$$

- Da die Gesamtverteilungsfunktion als Produkt der Nest-Verteilungsfunktionen $F_l(\mathbf{e}_l)$ geschrieben werden kann, sind die Zufallsnutzenanteile zwischen verschiedenen Nests l unabhängig.

⁵⁵Vom Sachverhalt her spricht zunächst nichts gegen $\lambda_l > 1$. Man kann aber zeigen, dass dann die GEV-Bedingung (4.119) nicht immer erfüllt ist. Außerdem versagt dann (4.129)

4 Modelle diskreter Entscheidungen

- Innerhalb eines Nests sind die Zufallsnutzenanteile in der Regel abhängig. Man kann zeigen, dass obige Verteilungsfunktion aus einer Summe zweier unabhängigen, Gumbel-verteilten Zufallsvariablen herrührt:

$$\begin{aligned} \epsilon_i = \epsilon_{lm} &= \sqrt{1 - \lambda_l^2} \epsilon_l + \lambda_l \epsilon_m, & \epsilon_l, \epsilon_m &\sim \text{Gumbel}, & (4.128) \\ \text{Corr}(\epsilon_l, \epsilon_{l'}) &= \delta_{ll'}, & \text{Corr}(\epsilon_m, \epsilon_{m'}) &= \delta_{mm'}, & \text{Corr}(\epsilon_l, \epsilon_m) &= 0. \end{aligned}$$

Mit einfachen Rechenregeln der Kovarianz und Korrelation kann man leicht zeigen (siehe Meier/Weiss), dass der gemeinsame übergeordnete Zufallsanteil ϵ_l zu einer positiven Korrelation der gesamt-Zufallsnutzen zweier Alternativen im selben Nest führt und durch

$$\text{Corr}(\epsilon_{lm} \epsilon_{lm'}) = 1 - \lambda_l^2 \quad (4.129)$$

gegeben ist ($m \neq m', 0 \leq \lambda_l \leq 1$). Je kleiner λ_l , desto stärker ist der Beitrag des übergeordneten Zufallsnutzens und desto größer die Korrelation innerhalb eines Nests.

- Im Grenzfall $\lambda_l \rightarrow 0$ gilt $F_l(e_l) = \exp[-e^{-e_l} \max_m (0 * e^{-e_m})] = \exp[-e^{-e_l}]$, also i.i.d. Gumbelverteilte Zufallsnutzen in den Alternativen der oberen Entscheidungsebene und keine Zufallsnutzen innerhalb der Nests. Dies ist konsistent mit der Beziehung (4.129).
- Im Grenzfall $\lambda_l \rightarrow 1$ zerfällt $F_l(e_l)$ in das Produkt $F_l(e_l) = \prod_m \exp(-e^{-e_{lm}})$, also in i.i.d. Gumbelverteilungen des Gesamt-Zufallsnutzens für alle Alternativenkombinationen (im Gegensatz zu nur einer Zufallsvariable pro Nest für $\lambda_l = 0$) und damit das Gesamtmodell in ein MNL.

Auswahlwahrscheinlichkeiten: Um (4.122) anzuwenden, benötigen wir zunächst die partielle Ableitung $G_i = \partial G / \partial y_i$ für $i = (l, m)$. Für ein bestimmtes Nest l fallen alle Summanden von $G = \sum_{l'} \left(\sum_{m'} y_{l'm'}^{1/\lambda_{l'}} \right)^{\lambda_{l'}}$ mit $l' \neq l$ heraus und man erhält für ein m in Nest l :

$$G_i = G_{m|l} = \lambda_l \left(\sum_{m'} y_{lm'}^{1/\lambda_l} \right)^{\lambda_l - 1} \frac{1}{\lambda_l} y_{lm}^{1/\lambda_l - 1} = y_{lm}^{1/\lambda_l - 1} \left(\sum_{m'} y_{lm'}^{1/\lambda_l} \right)^{\lambda_l - 1}$$

und damit die kombinierte Auswahlwahrscheinlichkeit

$$P_i = P_l P_{m|l} = \frac{y_{lm} G_{m|l}}{G} = \frac{y_{lm}^{1/\lambda_l} \left(\sum_{m'} y_{lm'}^{1/\lambda_l} \right)^{\lambda_l - 1}}{\sum_{l'} \left(\sum_{m'} y_{l'm'}^{1/\lambda_{l'}} \right)^{\lambda_{l'}}}$$

und nach Einsetzen von $y_i = e^{V_i}$ schließlich

$$P_i = P_l P_{m|l} = \frac{e^{V_{lm}/\lambda_l} \left(\sum_{m'} e^{V_{lm'}/\lambda_l} \right)^{\lambda_l - 1}}{\sum_{l'} \left(\sum_{m'} e^{V_{l'm'}/\lambda_{l'}} \right)^{\lambda_{l'}}} \quad (4.130)$$

4 Modelle diskreter Entscheidungen

Dies ist zwar ein analytischer, aber auch sehr unanschaulicher Ausdruck. Eine intuitivere Form bekommt man, wenn man den deterministischen Nutzen in einen übergeordneten gemeinsamen Nest-Anteil W_l und einen individuellen Anteil \tilde{V}_{lm} innerhalb eines Nests l aufspaltet:

$$V_i = V_{lm} = W_l + \tilde{V}_{lm}, \quad (4.131)$$

wobei \tilde{V}_{lm} i.A. natürlich vom Nest l abhängt, aber immer einen vom übergeordneten Zufallsnutzen ϵ_l separierten Anteil bezeichnet. Setzt man dies in (4.130) ein, ergibt sich nach einigen Rechenschritten das intuitive Ergebnis

$$P_i = P_l P_{m|l}, \quad P_l = \frac{e^{W_l + \lambda_l I_l}}{\sum_{l'} e^{W_{l'} + \lambda_{l'} I_{l'}}}, \quad P_{m|l} = \frac{e^{\tilde{V}_{lm}/\lambda_l}}{\sum_{m'} e^{\tilde{V}_{lm'}/\lambda_l}} \quad (4.132)$$

mit dem "Inklusionswert"

$$I_l = \ln \left(\sum_m e^{\tilde{V}_{lm}/\lambda_l} \right). \quad (4.133)$$

Dies bedeutet:

- Zunächst einmal die Baumstruktur der Entscheidung in eine unabhängige übergeordnete und eine bedingte untergeordnete Entscheidung, analog zum "Warscheinlichkeitsbaum" zur Definition der kombinierten und bedingten Wahrscheinlichkeiten der Ereignisse A und B : $P(A \cap B) = P(A)P(B|A)$.⁵⁶
- Die übergeordnete Entscheidung für ein Nest l erfolgt mit dem normalen MNL und den Nutzenfunktionen $W_l + \lambda_l I_l$. Hierbei beschreibt $\lambda_l I_l$ den vom Entscheider angenommenen "Pauschalnutzen" der individuellen Attribute der Alternativen im Nest l . Man kann zeigen (siehe z.B. Meier/Weiss), dass $\lambda_l I_l$ gleich dem Dichtemaximum des sich bei Nutzenmaximierung innerhalb des Nests ergebenden Nutzenanteils $U_m = \tilde{V}_{lm} + \epsilon_m$ ist.
- Die untergeordnete Entscheidung innerhalb des Nests geschieht wieder mit dem MNL unter Verwendung der Nutzen \tilde{V}_{lm}/λ_l oder, gleichbedeutend, mit den ursprünglichen deterministischen Individualnutzen \tilde{V}_{lm} , aber einer um den Faktor $\lambda_l \leq 1$ verringerten Standardabweichung der (Gumbelverteilten) Zufallsnutzen ϵ_m innerhalb des Nests, was konsistent mit (4.128) ist.
- Für $\lambda_l \rightarrow 0$ (**sequentielles Modell**) gilt $\lambda_l I_l = \max_{m'} \tilde{V}_{lm'}$. Bei der Nest-Wahl ist der angenommene Pauschalnutzen der Alternativen von Nest l also gleich dem maximalen deterministischen Nutzen der darin enthaltenen Alternativen. Konsequenterweise wird in der untergeordneten Entscheidung immer das Nest-Element mit maximalem deterministischen Nutzen gewählt (klar, Homo Oeconomicus zusammen mit einer verschwindenden Standardabweichung des Zufallsnutzens innerhalb des Nests). Beides gilt trivialerweise für allgemeine λ_l , falls das Nest l nur ein einziges Element enthält ($M_l = 1$).

⁵⁶Natürlich kann man auch rekursiv weitere Schachtelungen (=nests) hinzufügen.

4 Modelle diskreter Entscheidungen

- Falls $\lambda_l > 0$ und $M_l \geq 2$, ist der Pauschalnutzen der Elemente von Nest l größer als das Maximum der deterministischen Einzelnutzen. Das ist plausibel, denn dann gehen ja aus diesem Nest mehrere Alternativen “ins Rennen”.
- Falls alle $\lambda_l = 1$, erhält man aus (4.132) und (4.133) wieder die MNL-Auswahlwahrscheinlichkeiten

Aufgabe:

Zeigen Sie durch direktes Einsetzen der Ausdrücke (4.132) und (4.133), dass das Nested-Logit-Modell für $\lambda_l = 1$ in ein gewöhnliches Multinomial-Logit-Modell über alle Kombinationen der Teilalternativenmengen übergeht:

$$P_{(l,m)} = P(l)P(m|l) = \frac{\exp(W_l + \tilde{V}_{lm})}{\sum_{l'm'} \exp(W_{l'} + \tilde{V}_{l'm'})}$$

Zeigen Sie dies auch direkt mit der “unanschaulichen” Formulierung (4.130), welche diesen Sonderfall sogar einfacher herleiten lässt.

Modellkalibrierung: Bei der Kalibrierung ist es wichtig, zwischen von den Entscheidungen (Personen) n unabhängigen Parametern und den exogenen, i.A. von n abhängigen Faktoren zu unterscheiden, deshalb wird hier der Personenindex wieder mitgeführt. Mit $i = (l, m)$ sind die deterministischen Nutzenfunktionen also durch

$$V_{ni} = V_{nlm} = W_{nl} + \tilde{V}_{nlm}$$

gegeben. Die Kalibrierung basiert auf der Aufspaltung der Auswahlwahrscheinlichkeiten (4.132) in zwei geschachtelte Logit-Auswahlwahrscheinlichkeiten, man geht also zweistufig vor:

- Zuerst werden die Parameter der untergeordneten Entscheidungen kalibriert, welche nach (4.132) einem MNL für die deterministischen Nutzenfunktionen $\tilde{V}_{nlm}/\lambda_l$ entsprechen. Da man λ_l noch nicht kennt, ist es günstig, die mit λ_l skalierten Individualnutzen der Alternativen m innerhalb eines Nests l direkt als lineare Funktion der zugehörigen exogenen Faktoren \mathbf{X}_{nlm} (einschließlich ACs) für Person n in der Alternative m des Nests l anzusetzen:

$$\tilde{V}_{nlm}/\lambda_l = \beta_l' \mathbf{X}_{nlm}. \quad (4.134)$$

Mit diesem Ansatz kalibriert man, für jedes Nest unabhängig, den zum Nest l gehörigen Parametervektor β_l wie beim normalen MNL.

- Nun werden nach (4.133) die Inklusionswerte I_{nl} bestimmt:

$$I_{nl} = \ln \left(\sum_m e^{\beta_l' \mathbf{X}_{nlm}} \right). \quad (4.135)$$

4 Modelle diskreter Entscheidungen

Die Inklusionswerte stellen den von außen sichtbaren mittleren Nutzen \tilde{V}_{nl}/λ_l des Nests l dar. Da sie deterministisch angesetzt werden, aber die individuellen Maximierung des Gesamtnutzens der Nestalternativen darstellen, ist der Inklusionswert der Erwartungswert des maximierten Gesamtnutzens. Als solcher ist er bei $M_l \geq 2$ Nest-Alternativen größer als der größte deterministische Nutzen im Nest, speziell wenn der zweitgrößte Nutzen kaum kleiner ist.

- Nach dem Obigen werden die Nutzen \tilde{V}_{nlm} der Nest-Alternativen pauschal als $\lambda_l I_{nl}$ betrachtet und man erhält für die übergeordnete Entscheidung

$$W_{nl} + \tilde{V}_{nlm} = W_{nl} + \lambda_l I_{nl} = \beta'_{\text{ext}} \mathbf{X}_{nl} + \lambda_l I_{nl}. \quad (4.136)$$

Dies ist eine parameterlineare Funktion der im übergeordneten Nutzenanteil W_{nl} enthaltenen exogenen Faktoren \mathbf{X}_{nl} . Die Inklusionswerte I_{nl} spielen hier die Rolle weiterer exogener Variablen⁵⁷ und die Korrelationsparameter λ_l sind Teil des zu kalibrierenden Parametervektors $(\beta'_{\text{ext}}, \lambda')$.

Mit den geschätzten Parametervektoren β'_{ext} , λ und β_l kann man nun mit Hilfe von (4.132) die modellierten Nest-Auswahrscheinlichkeiten P_{nl} (Nutzenfunktionen $W_{nl} + \lambda_l I_{nl}$) bestimmen, und auch die Auswahlwahrscheinlichkeiten P_{nml} innerhalb eines Nests (Nutzenfunktionen $\tilde{V}_{nlm}/\lambda_l = \beta'_l \mathbf{X}_{nlm}$).

Beispiel: Entscheidungen beim Einkauf

Zur Untersuchung der Auswirkung großer Discounter auf Tante-Emma-Läden sowie auf die Verkehrsmittelwahl wurde von 10 Personen die Ladenkategorie ($l = 1$: Tante Emma, $l = 2$: Discounter) sowie das Verkehrsmittel ($m = 1$: ÖV, $m = 2$: MIV) der letzten Einkaufstouren befragt. Exogene Einflussfaktoren sind dabei einerseits die (Haustür-Haustür) Fahrtzeit $T_i = T_{lm}$ zur Tante Emma oder dem Discounter mit ÖV oder MIV und andererseits der "Füllstand" des Kühlschranks von $F = 0$ (total leer) bis $F = 0.9$ (fast voll). In Tabelle 4.7 ist dies zusammengefasst.⁵⁸

Die Entscheidung wird als zweistufiger Prozess gemäß Abb. 4.29 aufgefasst und mit dem NL-Modell analysiert:

- Die Wahl des Ladentyps l wird dabei als übergeordnete (Top-Level) Entscheidung aufgefasst,
- und die Verkehrsmittelwahl m als davon abhängige Entscheidung.⁵⁹

⁵⁷Da diese aber selbst geschätzt und nicht fest sind, ist der resultierende Gesamtschätzer nicht effizient, der Unterschied zum effizienten Schätzer durch direkte Maximierung der Gesamt-Log-Likelihood ist aber meist gering.

⁵⁸Nur zur Einfachheit; in Wirklichkeit betrachtet man jede Entscheidung einzeln!

⁵⁹Dies ist rein formaler Natur und entspricht dem klassischen Vorgehen bei der Verkehrsplanung, man könnte es auch anders herum modellieren. Insbesondere läuft die Entscheidung *im Menschen* wohl simultan ab.

4 Modelle diskreter Entscheidungen

Per- sonen gruppe	T [min] Emma, ÖV	T [min] Emma, MIV	T [min] Disc, ÖV	T [min] Disc, MIV	Füll- stand F	y_{11}	y_{12}	y_{21}	y_{22}
1	25	15	25	20	0.9	1	2	0	0
2	25	30	40	30	0.8	3	0	0	1
3	20	20	30	30	0.7	2	1	1	1
4	25	10	25	10	0.6	0	3	0	2
5	15	5	30	20	0.5	1	2	0	2
6	15	15	25	20	0.4	1	1	0	1
7	15	20	45	45	0.3	3	1	0	1
8	15	15	15	15	0.2	1	0	2	3
9	25	15	40	30	0.1	1	1	0	1
10	25	10	25	20	0.0	0	1	1	3

Tabelle 4.7: Ein einfaches Beispiel einer geschachtelten *Revealed Choice*-Befragung der kombinierten Zielwahl ($l = 1$: Tante Emma, $l = 2$: Discounter) und Verkehrsmittelwahl ($m = 1$: ÖV, $m = 2$: MIV). Die endogenen Variablen y_{lm} geben die Zahl der Entscheidungen der jeweiligen Personengruppe für die Alternative (l, m) an.

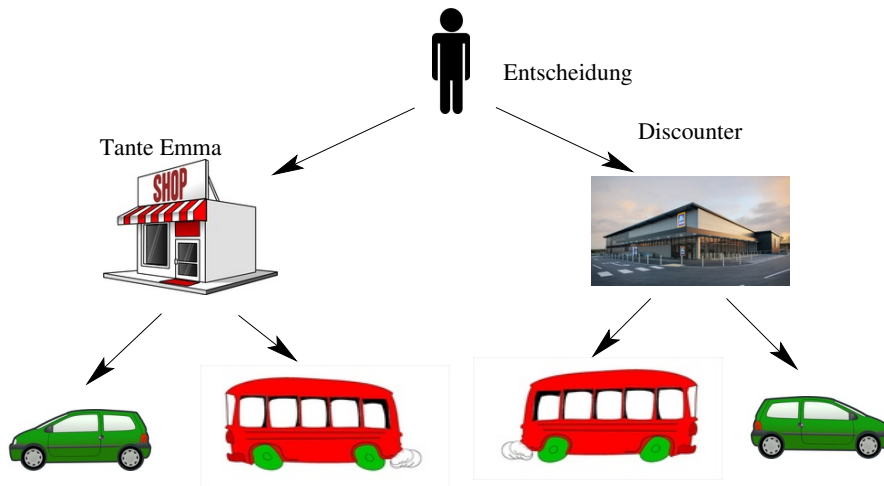


Abbildung 4.29: Beispiel eines Entscheidungsbaums bei der kombinierten Aktivitäten- und Verkehrsmittel

4 Modelle diskreter Entscheidungen

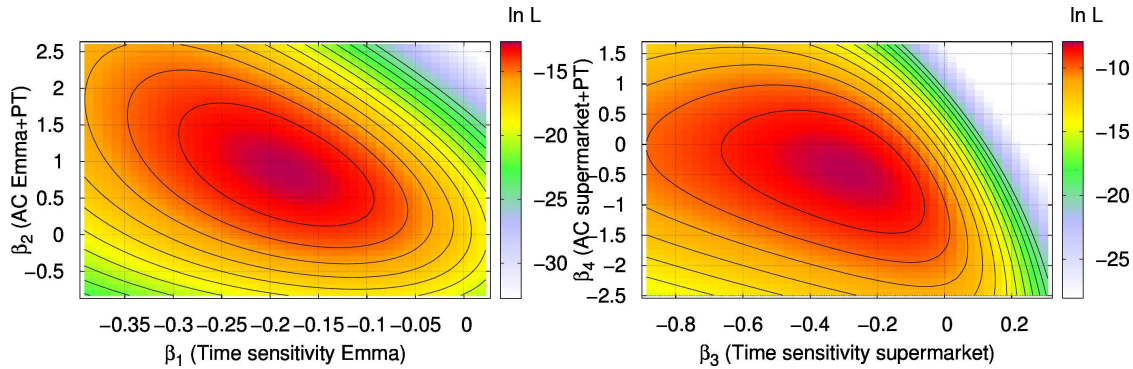


Abbildung 4.30: Log-Likelihood der Nest-Nutzenfunktionen (4.137) und (4.138) zur Tabelle 4.7. Links: Log-Likelihood der bedingten Verkehrsmittelwahl für Ziel ‘Tante Emma’, rechts für den Discounter. Der Abstand zwischen zwei Höhenlinien entspricht einer Log-Likelihood-Differenz von 1

Die deterministischen Nutzenfunktionen der Nester $l = 1$: Tante Emma, $l = 2$: Discounter, enthalten nur die Einflussfaktoren bei gegebener Ladentypwahl, also die Zeiten, während der Füllstand F des Kühlschranks wohl eher die Top-Level-Entscheidung (‘Ladenwahl’) beeinflusst. Der Einfachheit halber behandeln wir die Zeiten in jedem Nest generisch durch eine einzige Zeitsensitivität β_1 , die aber, da die Entscheidungen innerhalb der Nester unabhängig sind, für beide Nester i.A. unterschiedliche Werte aufweisen. Zusammen mit der jeweiligen AC werden die Entscheidungen innerhalb der Nester durch formal identische Nutzenfunktionen modelliert:

$$\tilde{V}_{n1m}/\lambda_1 = \beta_1 T_{n1m} + \beta_2 \delta_{m1}, \quad (4.137)$$

$$\tilde{V}_{n2m}/\lambda_2 = \beta_3 T_{n2m} + \beta_4 \delta_{m1}. \quad (4.138)$$

Die Kalibrierung durch Maximierung der Log-Likelihood kann jeweils direkt grafisch durch Contourplots der Log-Likelihood erfolgen. Anhand der Abbildung 4.30 ließt man direkt ab

$$\hat{\beta}_1 = -0.18, \quad \hat{\beta}_2 = +0.88, \quad \hat{\beta}_3 = -0.29, \quad \hat{\beta}_4 = -0.42.$$

- Die Zeitsensitivitäten haben das korrekte (negative) Vorzeichen und die Sensitivität beim ‘Tante-Emma’-Einkauf ist betragsmäßig etwas niedriger, allerdings nicht signifikant (weniger als zwei Log-Likelihood-Einheiten Unterschied in Abb fig:nestedLogit-logL bei den jeweils anderen Werten von β_1 bzw. β_3).
- Die AC ist beim ‘Tante-Emma’-Einkauf positiv (ÖV gegenüber dem Auto bevorzugt), beim Discounter dagegen nicht: Dies ist plausibel, denn beim Discounter kauft man i.A. viel ein und dann ist ein Auto einfach praktischer.

Mit den kalibrierten Nest-Parametern lassen sich für alle $n = 10$ Personengruppen die

4 Modelle diskreter Entscheidungen

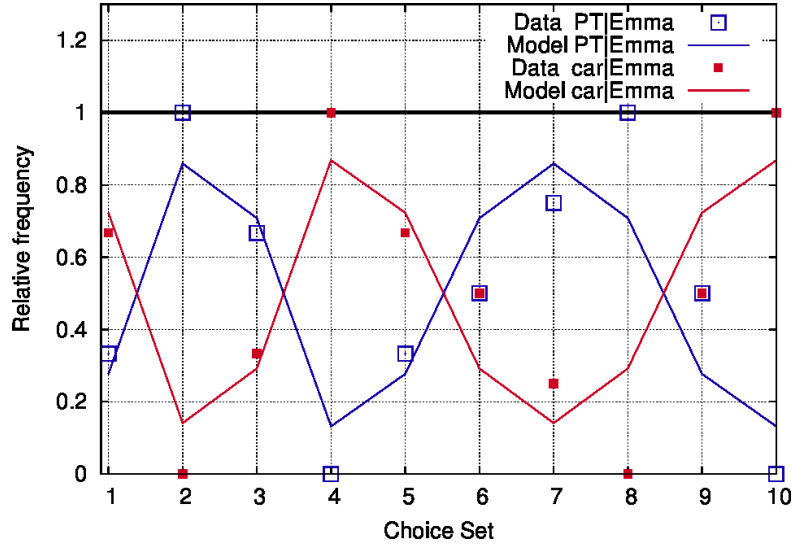


Abbildung 4.31: Verkehrsmittelwahl bei der Fahrt zu “Tante Emma”: Beobachtete relative Häufigkeiten (Symbole) und durch das NLM modellierte Wahrscheinlichkeiten $P_{nm|l=1}$.

Inklusionsvariablen beider Nests berechnen:

$$I_{n1} = \ln \left[\sum_m \exp \frac{\tilde{V}_{n1m}}{\lambda_1} \right] = \ln \left[\sum_m \exp \left(\hat{\beta}_1 T_{n1m} + \hat{\beta}_2 \delta_{m1} \right) \right],$$

$$I_{n2} = \ln \left[\sum_m \exp \frac{\tilde{V}_{n2m}}{\lambda_1} \right] = \ln \left[\sum_m \exp \left(\hat{\beta}_3 T_{n2m} + \hat{\beta}_4 \delta_{m1} \right) \right].$$

Beispielsweise ergeben sich für die erste Gruppe $n = 1$ und Nest $l = 1$ die Nutzenfunktionen $\tilde{V}_{111}/\lambda_1 = -3.72$ und $\tilde{V}_{112}/\lambda_1 = -2.77$ und daraus die Inklusionsvariable $I_{11} = -2.44$. Wie zu erwarten, ist die Inklusionsvariable I_{11} geringfügig größer (hier weniger negativ) als das Maximum der beiden reduzierten Nutzenfunktionen $\tilde{V}_{111}/\lambda_1$ und $\tilde{V}_{112}/\lambda_1$.

Mit bekannten Inklusionsvariablen hat man nun ein parameterlineares MNL-Modell für die Top-Level-Entscheidung der binären Ladenwahl mit dem deterministischen Nutzen

$$W_{nl} = \beta_5 F_n \delta_{l1} + \beta_6 \delta_{l1} + \lambda_1 I_{nl} \delta_{l1} + \lambda_2 I_{nl} \delta_{l2}. \quad (4.139)$$

Neben der obligaten alternativenspezifischen Konstanten δ_{l1} und dem zugehörigem Parameter β_6 (globaler Vorteil für Tante Emma gegenüber dem Discounter bei leerem Kühlschrank und verschwindenden Zeiten) muss der Füllstand als sozioökonomische Variable ebenfalls alternativenspezifisch formuliert werden. Die Inklusionsvariablen stellen Charakteristika der Top-Level-Alternativen dar und können als solche generisch (ein λ)

4 Modelle diskreter Entscheidungen

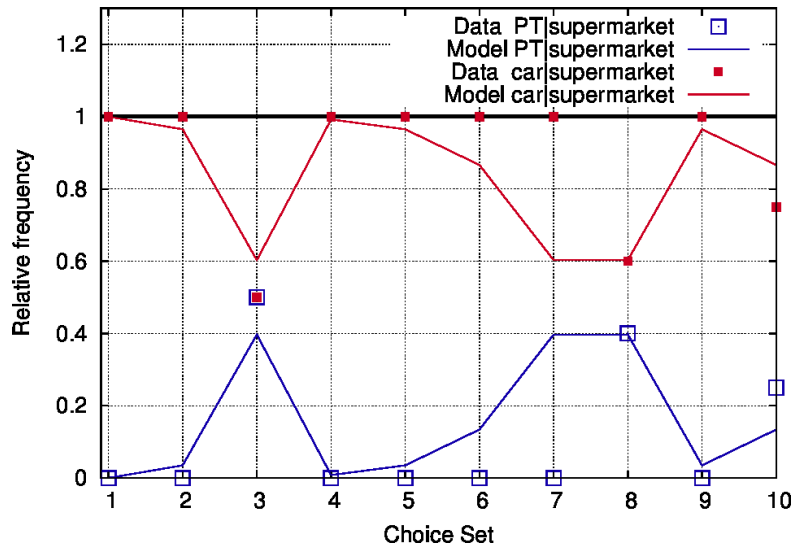


Abbildung 4.32: Verkehrsmittelwahl bei der Fahrt zum Discounter: Beobachtete relative Häufigkeiten (Symbole) und durch das NLM modellierte Wahrscheinlichkeiten $P_{nm|l=2}$.

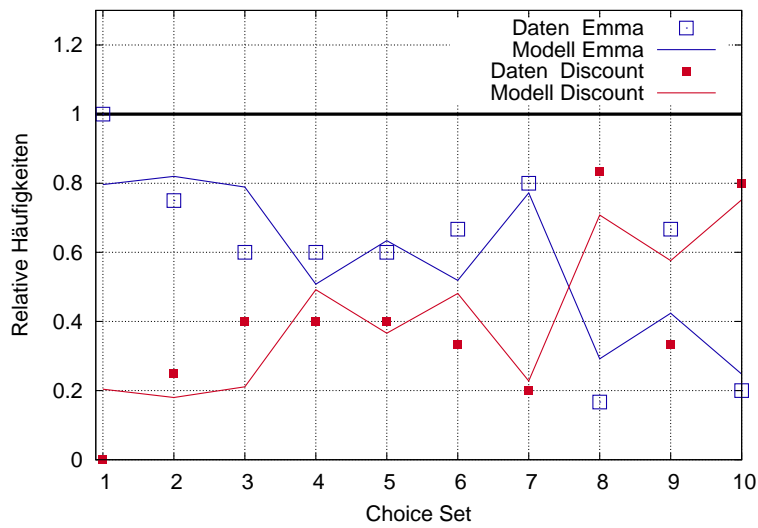


Abbildung 4.33: Wahl des Ladentyps: Beobachtete relative Häufigkeiten (Symbole) und durch das NLM modellierte Wahrscheinlichkeiten P_{nl} .

4 Modelle diskreter Entscheidungen

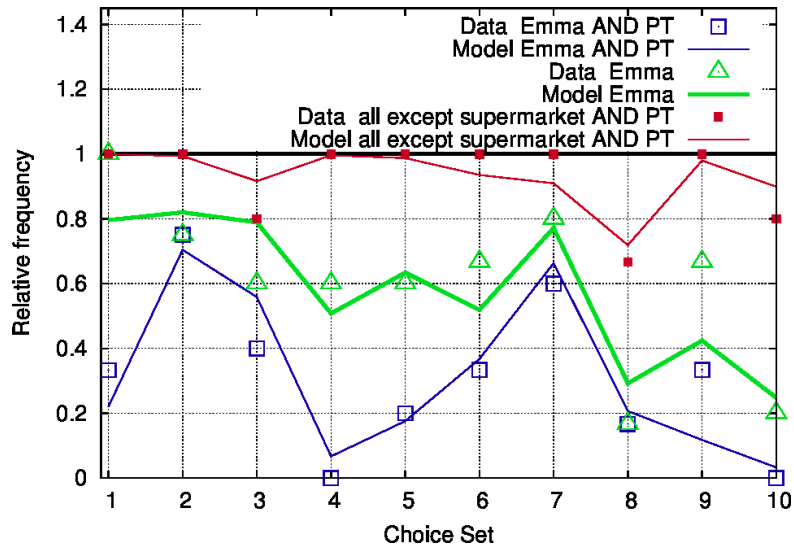


Abbildung 4.34: Kombinierte Entscheidung Ladentyp und Verkehrsmittel: Beobachtete relative (kumulierte) Häufigkeiten (Symbole) und durch das NLM modellierte Wahrscheinlichkeiten $P_{nl}P_{nm|l}$.

oder alternativenspezifisch (λ_1 und λ_2 mit entsprechenden Selektoren) formuliert werden. Hier wurde letzteres angenommen. Die ML-Parameterschätzung ergibt

$$\hat{\beta}_5 = 2.9, \quad \hat{\beta}_6 = -2.0, \quad \hat{\lambda}_1 = 0.17, \quad \hat{\lambda}_2 = 0.21.$$

Die Werte sind plausibel:

- $\hat{\beta}_5$ signifikant > 0 : Der relative Nutzen des “Tante-Emma”-Ladens gegenüber dem Discounter steigt mit dem Füllstand des Kühlschranks: Ist dieser noch fast voll, ist es nicht sinnvoll, beim Discounter größere Mengen einzukaufen.⁶⁰
- $\hat{\beta}_6$ signifikant < 0 : Bei leerem Kühlschrank ist hingegen der Discounter “Tante Emma” eindeutig vorzuziehen
- $\hat{\lambda}_1$ und $\hat{\lambda}_2$ liegen beide im geforderten Bereich $[0,1]$,⁶¹ aber näher an 0: Der Zufallsnutzen der Top-Level-Nutzenfunktion (unbeobachtete Faktoren bei der Ladenwahl) ist deutlich höher als der der Verkehrsmittelwahl; der Gesamt-Zufallsnutzen zwischen ÖV und MIV bei gegebenem Ladentyp hat eine hohe positive Korrelation.

Mit dem nun vollständig geschätzten Modell kann man schließlich noch die Fitgüte durch Vergleich der bei den Gruppen vorgefundenen relativen Häufigkeiten mit den modellierten Wahrscheinlichkeiten abschätzen (Abb. 4.31, 4.32, 4.33 und 4.34).

⁶⁰Man könnte noch ein drittes triviales Nest (“gar nicht einkaufen”) zur Top-Level-Entscheidung hinzufügen.

⁶¹Bei der Schätzung können aber durchaus auch “verbotene” Werte vorkommen, diese sollten allerdings innerhalb der Schätzunsicherheit von den Grenzen abweichen.

4 Modelle diskreter Entscheidungen

Aufgabe:

Modellieren Sie dieses Beispiel mit einem normalen MNL und definieren Sie dabei die Nutzenfunktion so, dass die Parameter β_1 bis β_6 dieselbe Bedeutung haben.

4.13.4 Weitere GEV-Modelle

Zwei weitere häufig benutzte GEV-Varianten sind die **Paired Combinatorial Logit (PCL)** und **Generalized Nest Logit (GNL)** Modelle. Beide Modelle haben eine Nest-Struktur, jedoch kann eine Alternative Mitglied mehrerer Nests sein. Dies ist vorteilhaft, wenn eine Abhängigkeitsstruktur nicht streng hierarchisch ist.

Beispiel: Bei einer Verkehrsmittelwahl mit den Alternativen Fuss, Rad, Bus, Bahn, MIV-Fahrer und MIV-Mitfahrer würde man im NL-Modell ein Nest "MIV" für MIV-Fahrer und -Mitfahrer (gemeinsame unbeobachtete "Auto-Merkmale"), ein Nest "ÖV" für Bus und Bahn (gemeinsame unbeobachtete ÖV-Merkmale) sowie zwei triviale Nests für Fuss und Rad ansetzen. Man könnte auch Fuß und Rad zu einem Nest "sportlicher Modus" zusammenfassen, da es gemeinsame unbeobachtete "Schlechtwetter" und "körperliche Anstrengungsmerkmale" gibt. Bei diesem Ansatz missachtet man jedoch, dass die Alternative "MIV-Mitfahrer" viele mit dem ÖV gemeinsame Merkmale hat, insbesondere den Mangel an Privatspäre und an Flexibilität (der "Lift" fährt zu einer festen Zeit los). Also wäre es günstig, wenn die Alternative "MIV-Mitfahrer" gleichzeitig Mitglied der Nester "MIV" und "ÖV" sein könnte.

Paired Combinatorial Logit

Hier werden die I Alternativen zu $I(I-1)/2$ Paaren zusammengefasst, so dass jede Alternative paarweise mit jeder anderen korreliert ist. Das PCL wird durch

$$G^{\text{PCL}}(\mathbf{y}) = \sum_{i=1}^{I-1} \sum_{j=i+1}^I \left(y_i^{1/\lambda_{ij}} + y_j^{1/\lambda_{ij}} \right)^{\lambda_{ij}} \quad (4.140)$$

definiert. Formal hat jeder Summand die Struktur eines NL-Nests mit zwei Alternativen ("Paar"). Allerdings ist hier jede Alternative gleichzeitig Mitglied von $I-1$ Paaren anstelle von genau einem Nest im NL-Modell.

Generalized Nest Logit

Das NL-Modell wird im Generalized Nest Logit (GNL) in zweierlei Hinsicht verallgemeinert:

- Jede Alternative kann Mitglied mehrerer Nests sein
- Der "Grad der Mitgliedschaft" (von "bloßer Sympathisant" bis "eingetragenes Mitglied") kann stufenlos gewählt werden.

4 Modelle diskreter Entscheidungen

Die GEV-Funktion des GNL lautet

$$G^{\text{GNL}}(\mathbf{y}) = \sum_{l=1}^L \left(\sum_{i \in B_l} (\alpha_{li} y_i)^{1/\lambda_l} \right)^{\lambda_l} \quad (4.141)$$

Hierbei bezeichnet B_l die Menge der Mitglieder in Nest l und α_{li} den Zugehörigkeitsgrad der Alternative i zu Nest l . Jede Alternative muss Mitglied in mindestens einem Nest sein.

- Gibt es $I(I - 1)/2$ Nests mit je zwei Mitgliedern unterschiedlicher Paarungen und sind alle $\alpha_{li} = 1$, so wird das GNL zum PCL,
- sind die B_l exklusiv und vollständig (jedes i ist Mitglied in genau einem Nest) und ist $\alpha_{li} = 1$, so wird das GNL zum NL,
- hat darüberhinaus jedes Nest nur ein Mitglied oder sind alle $\lambda_l = 1$, so degeneriert das GNL weiter zum MNL

4.14 Mixed-Logit Modell



Die Klasse der **Mixed-Logit-Modelle** enthält stochastische Modellparameter mit vorgegebener Verteilung der Dichte $f(\boldsymbol{\beta})$. Die Auswahlwahrscheinlichkeit

$$P_{ni} = \int d\boldsymbol{\beta} f(\boldsymbol{\beta}) P_{ni}^{\text{MNL}}(\boldsymbol{\beta}) \quad (4.142)$$

ist ein gewichtetes Mittel der normalen Logit-Auswahlwahrscheinlichkeiten für bestimmte $\boldsymbol{\beta}$, gewichtet über die vorgegebene Verteilung der $\boldsymbol{\beta}$ mit Dichte $f(\boldsymbol{\beta})$.

Mixed-Logit-Modelle entsprechen einer ‘‘Mischung’’ von normalen, mit verschiedenen $\boldsymbol{\beta}$ parametrisierten Logit-Modellen, wobei die Wichtung der verschiedenen Parametervektorwerte durch die Dichtefunktion bzw. **mixing function** $f(\boldsymbol{\beta})$ gegeben ist.

In der Praxis gibt man meist die Form der Verteilung vor (z.B. $\beta_j \sim N(\mu_j, \sigma_j^2)$) und kalibriert die Parameter der *mixing function*, also ermittelt z.B. die Schätzer $\hat{\mu}_j$ und $\hat{\sigma}_j^2$ für alle j . Es gibt zwei Interpretationsmöglichkeiten desselben Modells:

- I **Taste-Variation**, also unterschiedliche Vorlieben für jede Person (*Kenneth E. Train*, Kap. 6.2). Dieser Interpretation liegt auch der Name dieser Modellklasse zugrunde,

4 Modelle diskreter Entscheidungen

II **verallgemeinerte Zufallsnutzen**, welche nicht iid Gumbel sind (Kap. 6.3). Dabei trennt man einfach von den stochastischen Parametern β_j den Erwartungswert μ_j ab und schlägt den Streuanteil dem Zufallsnutzen zu. Damit hat man ein gewöhnliches Logit mit festen Parametern $\boldsymbol{\mu}$, aber zusätzlich zu den i.i.d. gumbelverteilten Zufallsnutzen des Logitmodells einen weiteren nicht-iid, nicht-gumbelverteilten Zufallsnutzenanteil, den man beliebig wählen kann.

Des Weiteren hat diese Modellklasse folgende Eigenschaften:

- Im Sonderfall, dass nur die alternativenspezifischen Konstanten (ACs) stochastisch modelliert werden, modelliert man in der Interpretation als *Taste-Variation* unterschiedliche globale Bevorzugungen verschiedener Personen, in der Interpretation als verallgemeinerte Zufallsnutzen zusätzliche nicht-Gumbel, nicht-i.i.d. Zufallsnutzen, welche zwischen den Personen, nicht aber mit den exogenen Variablen korrelieren können.
- Für das Mixed-Logit-Modell gilt i.A. nicht mehr die IIA-Auswahl-Verschiebungsregel (*substitution pattern*), Genaueres in *Train*, Kap. 6.4.
- Durch einen “Skalentrick” kann man jedes beliebige **Random-Utility Model** (RUM), z.B. das Probit-Modell, als Mixed-Logit-Modell schreiben (siehe *Train*, Kap. 6.5): Dazu macht man in der obigen Interpretation II die verallgemeinerte Zufallsnutzen sehr viel größer als die normalen i.i.d. Gumbel-Zufallsnutzen ϵ , lässt also den Gumbel-Skalenparameter $\lambda \rightarrow \infty$ gehen anstatt den Standardwert $\lambda = 1$ anzusetzen.

4.14.1 Paneldaten

Der begriff **Panel** bezieht sich, wie üblich, auf die mehrfache Befragung derselben Personen. In diesem Sinne sind Stated-Choice-Untersuchungen mit mehr als einem Choice Set pro Person (also der Art, wie ich sie in der Vorlesung machte) stets Panel-Befragungen. Ein konkret von einer Person beantwortetes Choice Set hat daher statt n zwei Indices:

- den Personenindex n ,
- zusätzlich den Choice-Set-Index t .⁶²

Im normalen Logit-Modell kann man die mit ein und derselben Person einhergehenden Korrelationsanteile nicht unterbringen: Jedes konkret beantwortete Choice Set (n, t) ist unabhängig und wird so behandelt, als ob jede Person nur ein einziges Choice Set beantworten würde.

Im Mixed-Logit-Modell kann man die Korrelationen ausnutzen, indem man gemäß Interpretation I (*Taste-Variation*) den totalen Nutzen ansetzt als

$$U_{nti} = \beta'_n \mathbf{x}_{nti} + \epsilon_{nti}. \quad (4.143)$$

⁶²Die Bezeichnung t kommt von der ursprünglichen Form der Panelbefragung, also Befragung derselben Person an mehreren Zeitpunkten t .

4 Modelle diskreter Entscheidungen

Der Parametervektor β hängt also von der Person n ab. Gemäß Interpretation II (Nicht-Gumbel-Anteile der Zufallsnutzen) setzt man

$$U_{nti} = \beta' x_{nti} + \sigma' \eta_n x_{nti} + \epsilon_{nti}, \quad (4.144)$$

wobei σ den Standardabweichungsvektor und η_n den für Person n realisierten Zufallsvektor aus einer Einheitsverteilung (z.B. Standardnormal) kennzeichnet.⁶³ Zu schätzen gibt es also nicht nur die über alle Personen gemittelten Parameter β (bzw. den Erwartungswert μ der personenabhängigen Parametervektoren), sondern auch die Varianzen $\hat{\sigma}_j^2$ der Schwankungen $\beta_n - \beta$ bzw. $\beta_n - \mu$ von Person zu Person.

4.14.2 Parameterschätzung des Mixed-Logit-Modells

Wie bei allen Modellen der diskreten Wahltheorie wird die Likelihoodfunktion minimiert, also nach (4.62) mit (4.142):

$$L(\vartheta) = \prod_n P_{ni_n}(\vartheta) = \prod_n \int d\beta f(\beta|\vartheta) P_{ni_n}^{\text{MNL}}(\beta) \quad (4.145)$$

wobei i_n die bei Person n beobachtete Wahlentscheidung ist. Hier wurde Unabhängigkeit bezüglich der Personen vorausgesetzt. Man beachte, dass die Likelihoodfunktion nun in Abhängigkeit der Parameter β der *Dichtefunktion* $f(\beta)$ des ursprünglichen Parametervektors gegeben ist, nicht mehr als Funktion des Parametervektors β selbst. Wie oben bereits erwähnt, enthält der Vektor ϑ in den meisten Modellierungen die Erwartungswerte und Varianzen der ursprünglichen Parameter.

Generell gibt es zwei Ansätze, $L(\vartheta)$ bzw. $\tilde{L}(\vartheta) = \ln L(\vartheta)$ zu berechnen (und zu maximieren):

- Direkte numerische Integration, z.B. durch Trapezregel. (Geht in `Python-Biogeme` durch fertige Routinen). Dies ist meist nur für einen stochastischen Parameter sinnvoll,
- Monte-Carlo-Simulation bei ≥ 2 stochastischen Parametern:

$$P_{ni} = \int d\beta f(\beta) P_{ni}^{\text{MNL}}(\beta) \approx \frac{1}{R} \sum_{r=1}^R P_{ni}^{\text{MNL}}(\beta_r) \equiv P_{ni}^{\text{um}}, \quad (4.146)$$

wobei β_r eine Realisierung (“Auswürfeln”) der mehrdimensionalen Zufallsvariable β mit Dichte $f(\beta)$ darstellt. Auch dafür liegen in `Biogeme` fertige Routinen vor.⁶⁴

Bei der eigentlichen Optimierung ist wichtig, dass das “Auswürfeln” der β_r nur bei der ersten Berechnung der Likelihood (mit den Startwerten von ϑ) geschieht, z.B.

⁶³Hier wurden keine Kovarianzen zwischen Parametern zugelassen, sonst muss man eine ganze Kovarianzmatrix schätzen, was prinzipiell (jedoch selten in der Praxis) auch geht.

⁶⁴Die Zahl R der nötigen *draufs* ist in `Biogeme` nicht dokumentiert. $R \geq 1000$ ist plausibel, für vieldimensionale Probleme auch deutlich mehr.

4 Modelle diskreter Entscheidungen

indem man im Ansatz $\beta(\vartheta) = \mu + \sigma\eta$ nur Realisierungen η_r der Standardverteilung (z.B. Standardnormalverteilung) "auswürfelt" und im Verlauf der eigentlichen Maximierung bezüglich ϑ (siehe auch (4.144)) die Werte $\beta_r = \mu + \sigma\eta_r$ durch Verschiebung (Änderung der Komponenten μ_j von ϑ) bzw. Skalierung (Änderung der Komponenten σ_j^2 von ϑ) ermittelt: Andernfalls ist $L(\vartheta)$ nicht bezüglich ϑ differenzierbar, was die Optimierung erschwert.

4.14.3 Spezialfall: Schätzung bei einer Panel-Struktur

In der allgemeinen Likelihood (4.145) ist zwar prinzipiell auch eine Panelstruktur einbaubar, indem man die Parameter-Verteilungen $f(\beta|\vartheta)$ von der Person n abhängen lässt, also $f_n(\beta|\vartheta)$ (z.B. für jede Person einen eigenen deterministischen Wert), aber das vervielfacht die Zahl der Modellparameter.⁶⁵

Gebräuchlicher ist deshalb das Vorgehen nach Abschnitt (6.7), v.A. Gl (6.2) und (6.3), des Lehrbuchs von *Train*: Eine 100%ige Korrelation von β bezüglich der Choice Sets t einer Person n wird erreicht, indem das Integral über $f(\beta)$ *simultan* für alle Choice Sets einer Person durchgeführt wird, also

$$L^{\text{Panel}}(\vartheta) = \prod_n \int d\beta f(\beta|\vartheta) \prod_t P_{nti_{nt}}^{\text{MNL}}(\beta) \quad (4.147)$$

Hingegen erhält man bei Annahme der Unabhängigkeit zwischen den Choice Sets (und Personen)

$$L^{\text{unabh}}(\vartheta) = \prod_n \prod_t \left(\int d\beta f(\beta|\vartheta) P_{nti_{nt}}^{\text{MNL}}(\beta) \right). \quad (4.148)$$

Ohne stochastische Modellparameter, also ohne Integration, gibt es keinen Unterschied. Mit Integration ergibt sich jedoch ein Unterschied, da das Produkt über Summen (bzw. Integrale) nach (4.148) i.A. nicht gleich der Summe (bzw. Integral) über Produkte nach (4.147) ist.

Veranschaulichung des Unterschiedes

Im einfachsten Fall, der auch gleichzeitig die Monte-Carlo-Simulation veranschaulicht, werden pro Person n zwei Choice Sets $t = 1$ und 2 betrachtet und das Integral durch nur zwei gleichgewichtete Summanden (Wichtung $1/2$) mit den (von ϑ abhängigen) Parametervektoren β_1 und β_2 angenähert. Nach (4.147) ergibt sich

$$L^{\text{Panel}} = \prod_n P_n^{\text{Panel}} = \prod_n \frac{P_{n11}P_{n21} + P_{n12}P_{n22}}{2},$$

nach (4.148) hingegen

$$L^{\text{unabh}} = \prod_n P_n^{\text{unabh}} = \prod_n \left(\frac{P_{n11} + P_{n12}}{2} \right) \left(\frac{P_{n21} + P_{n22}}{2} \right),$$

⁶⁵Im Extremfall läuft es auf die separate Schätzung des Modells für jede Person n heraus.

4 Modelle diskreter Entscheidungen

wobei die Abkürzungen

$$P_{n11} = P_{n1i_{n1}}^{\text{MNL}}(\beta_1), \quad P_{n12} = P_{n1i_{n1}}^{\text{MNL}}(\beta_2), \quad P_{n21} = P_{n2i_{n2}}^{\text{MNL}}(\beta_1), \quad P_{n22} = P_{n2i_{n2}}^{\text{MNL}}(\beta_2)$$

eingesetzt wurden.⁶⁶

Obige Formel beschreibt auch die tatsächliche Struktur der numerischen Annäherung durch “Simulation”, nur dass die Anzahl R der *draws* mindestens 1000 und nicht 2 betragen sollte.

4.15 Charakterisierung der Zuverlässigkeit



In Abschnitt 4.2.2 haben wir neben den “Klassikern” der Charakteristika von verkehrsmittelbezogenen Alternativen, nämlich totale Reisezeit und Ad-Hoc-Kosten, auch die “Zuverlässigkeit” durch die Standardabweichung der Reisezeit quantifiziert. Ist dies die beste Spezifikation der Zuverlässigkeit oder ist vielleicht ein Abstand des Medians der Reisezeit zu einem bestimmten Quantil, die Varianz oder etwas ganz anderes besser?

4.15.1 Annahmen und Modell

Da die Zuverlässigkeit Teil des deterministischen Nutzens V sein soll, berechnen wir den **erwarteten Nutzen** (*expected utility*) bzw. die *expected disutility* $W = -V$ eines unzuverlässigen Verkehrsmittels unter folgenden Annahmen (die Bezeichnungen sind an [M. Fosgerau, A Karlström “The value of Reliability”, Transportation Research Part B 44 (2010) 38–49] angelehnt):

- Die optimale Ankunftszeit sei durch $t = 0$ definiert. Bei späterer Ankunft, $t_a > 0$, fallen “Strafkosten” proportional zur Verspätungszeit an:

$$W_{\text{delay}}(t_a) = \nu t_a \vartheta(t_a), \quad \vartheta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (4.149)$$

mit der Verspätungs-Zeitsensitivität $\nu > 0$ [NE/min] und der schon bekannten Sprungfunktion $\vartheta(x)$. Kommt man zu früh an ($t_a < 0$), gibt es keine Strafe, aber auch keinen Nutzen.

- Verpasst man durch Zuspätkommen Verbindungen, insbesondere Flüge, ist es plausibler, anstattdessen einen diskreten Nachteil zu formulieren:

$$W_{\text{miss}}(t_a) = M \vartheta(t_a). \quad (4.150)$$

Kommt man auch nur eine Sekunde zu spät, hat man den Zug/Flug verpasst und verbucht einen Schaden von M Nutzeinheiten (die Parameterbezeichnung M kommt von “missed”).

⁶⁶Leider hat Train in seinem Buch diesen Unterschied, der m.M nach vieles erklärt, nicht explizit gemacht.

4 Modelle diskreter Entscheidungen

- Zu spät zu kommen oder etwas zu verpassen, kann man durch einen früheren Reisestart (departure time $t_d = -D < 0$) vermeiden, muss dann aber früher aufstehen bzw. eine andere Tätigkeit eher beenden und verbucht einen Nachteil von

$$W_{\text{depart}} = \lambda D \quad (4.151)$$

Nutzeneinheiten mit der Zeitsensitivität $\lambda > 0$.⁶⁷

- Die komplexe Reisezeit T ist eine Zufallsgröße, deren Verteilung *nicht* von der Startzeit abhängt:

$$T = \mu + \sigma X, \quad E(X) = 0, \quad \text{Var}(X) = 1 \quad (4.152)$$

Die Zeit T hat also einen Erwartungswert μ und eine Varianz σ^2 . Ferner seien die Verteilungs- und Dichtefunktionen der standardisierten Reisezeit X durch $F(x)$ bzw. $f(x)$ gegeben. Des Weiteren sind die Verteilungs- und Dichtefunktionen der nicht-standardisierten Reisezeit durch

$$F_T(T) = F\left(\frac{T - \mu}{\sigma}\right), \quad f_T(T) = \frac{1}{\sigma} f\left(\frac{T - \mu}{\sigma}\right) \quad (4.153)$$

gegeben.

Die Reisezeit T ist eine Zufallsgröße, so dass auch die sich aus den obigen Annahmen ergebende Disutility $-U_T$ der Reisezeit eine Zufallsgröße darstellt. Mit der Ankunftszeit $t_a = T - D$ gilt

$$-U_T = \lambda D + \nu(T - D)\vartheta(T - D) + M\vartheta(T - D). \quad (4.154)$$

Die *erwartete* Disutility W lautet damit

$$W(D) = E(-U_T) = \lambda D + \nu \int_{T=D}^{\infty} (T - D) f_T(T) \, dT + M \mathfrak{R}_{\text{delay}}(D) \quad (4.155)$$

mit der Verspätungswahrscheinlichkeit (vgl. (4.153))

$$\mathfrak{R}_{\text{delay}}(D) = \int_{T=D}^{\infty} f_T(T) \, dT = 1 - F_T(D) = 1 - F\left(\frac{D - \mu}{\sigma}\right). \quad (4.156)$$

Mit den normalisierten Größen $T = \mu + \sigma X$, $dT = \sigma \, dx$ und (4.153) vereinfacht sich das erste Integral von (4.155) mit der Bedeutung einer *mittleren Verspätungsdauer* $E(t^+)$ zu

$$E(t^+) = \int_{T=D}^{\infty} (T - D) f_T(T) \, dT = \int_{x=(D-\mu)/\sigma}^{\infty} (x\sigma + \mu - D) f(x) \, dx$$

⁶⁷Man könnte auch noch eine separate Disutility ωT für die Reisezeit T als solche definieren (siehe Fosgerau), diese würde aber bei der Bestimmung des optimalen Startzeitpunktes wieder herausfallen und bei der resultierenden minimierten Disutility nur dann separat erscheinen, wenn sie ungleich λ ist: Die Reisezeit ωT ist ja bereits indirekt in der Start-Disutility λD enthalten, da D die komplette Zeitspanne bis $t = 0$ einschließt. Wir nehmen hier $\omega = \lambda$ an.

4 Modelle diskreter Entscheidungen

Damit wird die erwartete Disutility

$$\begin{aligned} W(D) &= \lambda D + \nu E(t^+) + M \mathbb{P}_{\text{delay}} \\ &= \lambda D + \nu \int_{x=(D-\mu)/\sigma}^{\infty} (x\sigma + \mu - D)f(x) \, dx + M \left[1 - F\left(\frac{D-\mu}{\sigma}\right) \right] \end{aligned} \quad (4.157)$$

Nun *optimieren* wir den Startzeitpunkt $-D$ durch eine Maximierung des erwarteten Nutzens bzw. eine Minimierung der erwarteten Disutility. Eine notwendige Bedingung dafür ist, dass die Ableitung $=0$ ist:⁶⁸

$$\begin{aligned} 0 &\stackrel{!}{=} W'(D) \\ &= \lambda + \nu \left[-\frac{1}{\sigma} * 0 * f\left(\frac{D-\mu}{\sigma}\right) + \int_{x=(D-\mu)/\sigma}^{\infty} (-f(x) \, dx) \right] - \frac{M}{\sigma} f\left(\frac{D-\mu}{\sigma}\right) \end{aligned}$$

und damit

$$0 = \lambda + \nu \left[F\left(\frac{D-\mu}{\sigma}\right) - 1 \right] - \frac{M}{\sigma} f\left(\frac{D-\mu}{\sigma}\right) \quad (4.158)$$

4.15.2 Zwei Spezialfälle

Sinnvolle Ergebnisse erhält man für zwei Fälle:

Fall 1: keine diskreten Verspätungskosten, $M = 0$:

Dann ergibt sich aus (4.158) mit (4.156) direkt die optimierte Verspätungswahrscheinlichkeit

$$\frac{\lambda}{\nu} = 1 - F((D-\mu)/\sigma) = \mathbb{P}_{\text{delay}} \quad (4.159)$$

und durch Auflösen nach dem Startzeitpunkt D unter Nutzung der Inversen der Verteilungsfunktion, der Quantilsfunktion $F^{-1}(p)$ erhält man schließlich den optimalen Startzeitpunkt:

$$D^* = \mu + \sigma F^{-1}\left(1 - \frac{\lambda}{\nu}\right). \quad (4.160)$$

Dies gilt, solange die Verspätungskosten ν pro Zeiteinheit größer als die Kosten λ pro Zeiteinheit eines früheren Startes sind: Ansonsten ist die Inverse nicht definiert (Quantil einer negativen Wahrscheinlichkeit) und es ist am besten, gar nicht aufzubrechen (Optimum bei $D \rightarrow -\infty$). Durch Einsetzen dieses Ergebnisses in (4.157) ergibt sich schließlich die minimierte erwartete Disutility⁶⁹

$$w^* = \lambda\mu + \nu E(t^+) = \lambda\mu + \nu\sigma H \quad (4.161)$$

⁶⁸Man beachte, dass beim Ableiten des Integrals die Variable D sowohl in den unteren Integrationsgrenzen zu finden ist (Ableitung ergibt $-1/\sigma$ mal Integrand) als auch im Integranden selbst (unabhängig vom Integral wird dann einfach der Integrand abgeleitet). Ferner wird $xf(x) \rightarrow 0$ für $x \rightarrow \infty$ angenommen.

⁶⁹Dies wurde in Fosgerau umständlicher direkt berechnet.

4 Modelle diskreter Entscheidungen

mit der in Standardabweichungen gemessenen mittleren Verspätung (nicht des Verkehrsmittels sondern bezüglich des Termins)⁷⁰

$$H = \int_{F^{-1}(1-\lambda/\nu)}^{\infty} x f(x) dx$$

welche nicht von σ abhängt. Damit ist die gesamte erwartete Utility unzuverlässiger Verkehrsmittel von der Form

$$V = V^* = \beta_1 D + \beta_2 \sigma \tag{4.162}$$

mit $\beta_1 = -\lambda$ der Zeitsensitivität bezüglich des Erwartungswertes und $\beta_2 = -\nu H$ die Sensitivität bezüglich der Standardabweichung

Fall 2: keine stetigen Verspätungskosten, $\nu = 0$:

Anstelle von (4.160) erhalten wir nun die optimale Verspätungswahrscheinlichkeit bzw. den optimalen Startzeitpunkt als

$$D^* = \mu + \sigma f^{-1} \left(\frac{\lambda \sigma}{M} \right) \tag{4.163}$$

sowie

$$p_{\text{delay}}^* = 1 - F \left(\frac{D^* - \mu}{\sigma} \right) = 1 - F \left(f^{-1} \left(\frac{\lambda \sigma}{M} \right) \right) \tag{4.164}$$

Dies ist anschaulich: Schiebe ich meinen Startzeitpunkt nach vorn, $dD > 0$, erhalte ich einerseits einen Nachteil λdD durch früheres Aufstehen, aber auch einen Vorteil $-M dp_{\text{delay}} = M/\sigma f((D - \mu)/\sigma) dD$. Im Optimum ist der Vorteil gleich dem nachteil, woraus man D isolieren kann. Der optimale Startzeitpunkt ist aber nur definiert, falls die Inverse der Dichtefunktion $f(x)$ für den Wert $\lambda\sigma/M$ existiert. Da aufgrund der Normalisierung das Maximum der Dichtefunktion in der Größenordnung 1 liegt, ist dies für $M \gg \lambda\sigma$ erfüllt: Der Schaden durch einen verpassten Anschluss muss größer sein als der Aufwand, eine Standardabweichung früher zu starten. Dies ist offensichtlich plausibel (andernfalls lohnt es sich, überhaupt nicht zu starten).

Wenn f^{-1} existiert, ist es aber nicht eindeutig. Generell nimmt man den größten Wert, der einer lokalen Nutzenmaximierung entspricht (während die anderen auch eine Nutzenminimierung darstellen können). Aufgrund des bei steigender Verspätung nicht weiter anwachsenden Schadens bei steigenden Nutzen (ich kann länger und länger auschlafen bzw. was anderes tun) ist bei einer rein diskreten Verspätungsstrafe jedoch der optimale Startpunkt immer nur lokal optimal: Am besten wäre es, gar nicht zu starten. Dies kann man durch Restriktionen wie $D > \mu$ aber verhindern (und damit auch meist f^{-1} eindeutig machen).

In jedem Fall hängt die resultierende minimierte Disutility nicht linear von der Standardabweichung ab und deshalb ist eine einfache Aussage wie bei kontinuierlich ansteigender Verspätungsstrafe nicht möglich.

⁷⁰Das von Fosgerau entnommene H ist aber irgendwie wegen des Faktors x statt $x - (D - \mu)/\sigma$ nicht genau die mittlere Termin-Verspätung, obwohl sie so bezeichnet wird ...

4 Modelle diskreter Entscheidungen

4.15.3 Beispiel: exponentialverteilte Verspätungen

Unter den einfachen Verteilungsfunktionen sind exponentialverteilte Verspätungen mit am plausibelsten:

$$T = \mu + \sigma X, \quad X + 1 \sim E(1)$$

mit der Dichte und Verteilungsfunktion

$$f(x) = e^{-(x+1)}\vartheta(x+1), \quad F(x) = 1 - e^{-(x+1)}\vartheta(x+1) = 1 - f(x). \quad (4.165)$$

Ferner gilt folgende Quantilsfunktion sowie Umkerfunktion der Dichte:

$$F^{-1}(p) = 1 - \ln(1 - p), \quad f^{-1}(d) = -1 - \ln d. \quad (4.166)$$

Die planmäßige Reisezeit $T_{\min} = \mu - \sigma$ wird nie unterschritten,⁷¹ aber durch exponentialverteilte Verspätungen überlagert.

Für rein lineare Verspätungskosten (Fall 1) ergeben die obigen Verteilungsgrößen (4.165) und (4.166) einen optimalen Abstand zwischen Start und Termin von

$$D^*|_{M=0} = \mu + \sigma F^{-1}\left(1 - \frac{\lambda}{\nu}\right) = \mu + \sigma \left(1 + \ln\left(\frac{\nu}{\lambda}\right)\right)$$

während sich für einen rein diskreten (Ja-Nein) Verspätungsschaden der Höhe M der optimale Start-Abstand von

$$D^*|_{\nu=0} = \mu + \sigma f^{-1}\left(\frac{\lambda\sigma}{M}\right) = T_{\min} + \sigma \ln\left(\frac{M}{\lambda\sigma}\right),$$

sowie eine optimierte negative Nutzenfunktion von

$$W^*|_{\nu=0} = \lambda \left(\mu + \sigma \ln\left(\frac{M}{\lambda\sigma}\right) \right).$$

ergeben. Der optimale Abstand des Startzeitpunkts vom Termin ist also der Erwartungswert zuzüglich der gewichteten Standardabweichung, wobei die Wichtungsfaktor gleich dem Logarithmus der diskreten Verspätungskosten in Einheiten des Zeitwertes einer Standardabweichung beträgt. Dann ergibt sich die minimierte Disutility ebenfalls aus den Zeitkosten für den Erwartungswert zuzüglich der mit demselben Wichtungsfaktor gewichteten Standardabweichung. Sind die Verspätungskosten beispielsweise 100 mal so hoch, wird die Standardabweichung mit dem Faktor $\ln 100 \approx 4.5$ bewertet: Man sollte 4.5 Standardabweichungen Zeitpuffer einplanen und hat dann durch die Verspätungen im Mittel einen Nachteil, der ebenfalls dem Zeitwert von 4.5 Standardabweichungen entspricht.

Für exponentialverteilte Verspätungen ergibt also auch die diskrete Verspätungsstrafe einen anschaulichen Sinn und der Gewichtungsfaktor der Standardabweichung hat für lineare und diskrete Verspätungskosten nahezu dieselbe mathematische Form, wobei im Logarithmus-Argument ν/λ (linear) durch $M/(\lambda\sigma)$ (diskret) ersetzt wird.

⁷¹Da die exponentialverteilte Verspätung $Y \sim E(\sigma)$ in den nichtskalierten Variablen einen Erwartungswert gleich der Standardabweichung von σ hat, ist die planmäßige Reisezeit $= \mu - \sigma$ und nicht $= \mu$ und die auf Erwartungswert=0 und Varianz=1 normierte Zufallsgröße X gegeben durch $X = (Y - 1)/\sigma$.

4.16 Einige Herleitungen



Veranschaulichung bzw. direkte Herleitung von GI (4.21)

Für die Wahrscheinlichkeit P_1 , von zwei Alternative Alternative 1 auszuwählen, gilt

$$\begin{aligned} P_1 &= P(U_1 > U_2) \\ &= P(V_1 + \epsilon_1 > V_2 + \epsilon_2) \\ &= P(\epsilon_2 - \epsilon_1 \leq V_1 - V_2) \\ &= \underline{\underline{F_{\epsilon_2 - \epsilon_1}(V_1 - V_2)}}. \end{aligned}$$

Drückt man nun mit Hilfe der Faltungssätze (4.20) die Verteilungsfunktion der Differenz $\epsilon_2 - \epsilon_1$ durch die Verteilungsfunktionen bzw. Dichten der beiden Zufallsnutzen aus, erhält man den Ausdruck von (4.21) für zwei Alternativen, $I = 2$.

Herleitung der Gumbelverteilung als Grenzverteilung

Zunächst wird ein Spezialfall von Verteilungen mit exponentiellem Abfall betrachtet: Rein exponentialverteilte Zufallsgrößen $X_i \sim E(\lambda)$ mit der Verteilungsfunktion $F_E(x) = 1 - e^{-\lambda x}$. Wir betrachten die Verteilungsfunktion folgender Maximumsfunktion:

$$Y = \max_{i=1, \dots, K} X_i, \quad X_i \sim E(\lambda) \text{ mit } F_E(x) = 1 - e^{-\lambda x}.$$

Dann gilt

$$F_y(y) = P(Y \leq y) = P\left(\max_i X_i \leq y\right) = P\left((X_1 \leq y) \cap \dots \cap (X_I \leq y)\right).$$

Sind die X_i unabhängig, spalten sich die Wahrscheinlichkeit der kombinierte Bedingung in multiplikativ verknüpfte Wahrscheinlichkeiten für die Einzelbedingungen auf. Damit ergibt sich

$$\begin{aligned} F_y(y) &= P(X_1 \leq y)P(X_2 \leq y) \dots P(X_I \leq y) \\ &= \prod_i F_i(y) \\ &= \left(1 - e^{-\lambda y}\right)^I. \end{aligned}$$

Ist nun y hinreichend groß, so dass

$$\alpha := e^{-\lambda y} \ll 1, \tag{4.167}$$

4 Modelle diskreter Entscheidungen

so kann man folgende Näherung anwenden:

$$(1 + \alpha) \approx e^\alpha,$$

also $1 - e^{-\lambda y} \approx \exp(-e^{-\lambda y})$ und damit

$$\begin{aligned} F_y(y) &= \left(1 - e^{-\lambda y}\right)^I \\ &= \exp\left(-Ie^{-\lambda y}\right) \\ &= \exp\left(-e^{-\lambda y + \ln I}\right) \\ &= \exp\left[-e^{-\lambda\left(y - \frac{\ln I}{\lambda}\right)}\right]. \end{aligned}$$

Bei Gültigkeit der Näherung gilt also $Y \sim \text{Gu}(\ln I/\lambda, \lambda)$. Das Maximum ist also gumbelverteilt mit demselben Skalierungsparameter wie der der ursprünglichen Exponentialverteilung und einem Lageparameter, der logarithmisch mit der Zahl der zu maximierenden Variablen ansteigt. Nun muss man noch zeigen

1. dass bzw wann die Näherung (4.167) gültig ist,
2. dass das Ergebnis auch für beliebige Verteilungen mit *exponential tails* gültig ist.

Zu 1.: Für interessante Werte von y in der Nähe des Modalwertes $\eta_y = \ln I/\lambda$ (vgl. Gl. (4.26)) gilt $\alpha = e^{-\lambda y} = 1/I$, also ist Bedingung (4.167) für $I \gg 1$ erfüllt.

Zu 2.: Dies gilt auch für beliebige exponentiell abfallende Verteilungen, so lange der *exponential tail* spätestens beim $1 - 1/I$ -Quantil beginnt, also die Dichte exponentiell abfällt für alle x oberhalb des Wertes, für den $F(x) = 1 - 1/I$ (vgl. Abb. 4.9).

Herleitung der Identität (4.28)

Sei

$$Y = \max_{i=1,\dots,I} X_i, \quad X_i \sim \text{Gu}(\eta_i, \lambda).$$

Dann gilt, ganz analog wie bei der obigen Herleitung der Grenzverteilungseigenschaft

$$F_y(y) = P(Y \leq y) = \prod_i F_i(y),$$

4 Modelle diskreter Entscheidungen

nur dass die F_i diesmal die Verteilungsfunktionen von (η_i, λ) -Gumbelverteilungen darstellen. Mit (4.24) ergibt sich

$$\begin{aligned}
 F_y(y) &= \prod_i F_i(y) \\
 &= \prod_i \exp \left[-e^{-\lambda(y-\eta_i)} \right] \\
 &= \exp \left[-\sum_i e^{-\lambda(y-\eta_i)} \right] \\
 &= \exp \left[-e^{-\lambda y} \left(\sum_i e^{\lambda \eta_i} \right) \right] \\
 &= \exp \left[-e^{-\lambda \left[y - \frac{1}{\lambda} \ln \left(\sum_i e^{\lambda \eta_i} \right) \right]} \right]
 \end{aligned}$$

Vergleich mit (4.24) zeigt, dass $Y = \max_i X_i$ gumbelverteilt ist mit unverändertem Skalierungsparameter λ und neuem Lageparameter

$$\eta_y = \frac{1}{\lambda} \ln \left(\sum_i e^{\lambda \eta_i} \right).$$

Diese Identität ist auch konsistent mit der Tatsache, dass die Gumbelverteilung die Grenzverteilung des Maximums vieler unabhängiger Zufallsvariablen mit *exponential tails* ist: Das Maximum zweier gumbelverteilten Zufallsgrößen entspricht also dem Maximum zweier Maxima, also dem Maximum aller durch die beiden Gumbelverteilungen repräsentierten exponentiell abfallenden Zufallsvariablen, für die der Grenzwertsatz natürlich auch gilt.

Herleitung der Identität (4.30)

Seien zunächst zwei identisch gumbelverteilte Zufallsvariablen $X_1 \sim \text{Gu}(0, \lambda)$ und $X_2 \sim \text{Gu}(0, \lambda)$ gegeben. Dann gilt für die Verteilungsfunktion der Differenz $Y = X_2 - X_1$ unter Verwendung der Faltungssätze (4.20):

$$\begin{aligned}
 P(Y < y) = F_y(y) &= \int_{-\infty}^{\infty} dx f_1(x) F_2(y+x) \\
 &= \int_{-\infty}^{\infty} dx \lambda e^{-\lambda x} \exp[-e^{-\lambda x}] \exp[-e^{-\lambda(y+x)}] \\
 &= \int_{-\infty}^{\infty} dx \lambda e^{-\lambda x} \exp \left[- \left(e^{-\lambda x} + e^{-\lambda(y+x)} \right) \right] \\
 &= \int_{-\infty}^{\infty} dx \lambda e^{-\lambda x} \exp \left[-e^{-\lambda x} \left(1 + e^{-\lambda y} \right) \right]
 \end{aligned}$$

4 Modelle diskreter Entscheidungen

Im Integranden steht nun $1/(1+e^{-\lambda y})$ multipliziert mit der Ableitung von $\exp[-e^{-\lambda x}(1+e^{-\lambda y})]$ nach x , also

$$F_y(y) = \frac{1}{1+e^{-\lambda y}} \left[\exp[-e^{-\lambda x}(1+e^{-\lambda y})] \right]_{x=-\infty}^{x=\infty} = \frac{1}{1+e^{-\lambda y}}.$$

Verallgemeinert man das Ergebnis nun nachträglich auf nichtverschwindende Lageparameter η_1 und η_2 für die Zufallsvariablen X_1 bzw. X_2 , erhält man mit dem durch (4.27) gegebenen Verschiebungssatzes für den Lageparameter direkt die Identität (4.30): $\tilde{X}_1 = X_1 + \eta_1$ ist gumbelverteilt mit dem Lageparameter $\eta = \eta_1$ und $\tilde{X}_2 = X_2 + \eta_2$ gumbelverteilt mit $\eta = \eta_2$. Damit ist $Y + \eta_2 - \eta_1 = \tilde{X}_2 - \tilde{X}_1$ einerseits die gesuchte Differenz der Gumbelverteilungen mit den Lageparametern η_2 und η_1 . Andererseits gilt allgemein

$$X \text{ hat Verteilungsfunktion } F(x) \Leftrightarrow X + a \text{ hat Verteilungsfunktion } F(x - a)$$

Also hier

$$Y \text{ hat Verteilungsfunktion } \frac{1}{1+e^{-\lambda y}} \Leftrightarrow Y + \eta_2 - \eta_1 \text{ hat Verteilungsfunktion } \frac{1}{1+e^{-\lambda(y-\eta_2+\eta_1)}},$$

was der Aussage (4.30) entspricht.

Herleitung der Multinomial-Logit-Wahrscheinlichkeiten (4.34) unter Berücksichtigung der Gumbel-Eigenschaften

Hier werden die eben hergeleitete Identitäten (4.28) und (4.30) ausgenutzt. Für die Auswahlwahrscheinlichkeit P_1 gilt

$$\begin{aligned} P_1 &= P \left(V_1 + \epsilon_1 > \max_{i=2,\dots,I} (V_i + \epsilon_i) \right) \\ &= P \left(\epsilon_1 > \max_{i=2,\dots,I} (\epsilon_i + V_i - V_1) \right). \end{aligned}$$

Im Logit-Modell sind alle Zufallsnutzen gumbelverteilt: $\epsilon_i \sim \text{Gu}(0, \lambda)$.⁷² Definiert man nun unter Verwendung von (4.26) und (4.28):

$$\begin{aligned} X_1 &= \epsilon_1 \sim \text{Gu}(0, \lambda), \\ X_2 &= \max_{i=2,\dots,I} (\epsilon_i + V_i - V_1) \sim \text{Gu} \left(\frac{1}{\lambda} \sum_{i=2}^I e^{\lambda(V_i - V_1)}, \lambda \right) \end{aligned}$$

so gilt

$$P_1 = P(X_1 \geq X_2) = P(X_2 - X_1 \leq 0) = F_{x_2 - x_1}(0)$$

⁷²Den Lageparameter kann man ohne Einschränkung =0 setzen, da es nur auf Differenzen ankommt.

4 Modelle diskreter Entscheidungen

Nun wendet man Beziehung (4.30) für die Verteilungsfunktion der Differenz $Y = X_2 - X_1$ zweier gumbelverteilten Zufallsvariablen an:

$$\begin{aligned}
 P_1 &= F_{x_2-x_1}(0) \\
 &= \frac{1}{1 + e^{-\lambda \left(0 - \frac{1}{\lambda} \sum_{i=2}^I e^{\lambda(V_i - V_1)}\right)}} \\
 &= \frac{1}{1 + \sum_{i=2}^I e^{\lambda(V_i - V_1)}} \\
 &= \frac{e^{\lambda V_1}}{e^{\lambda V_1} + \sum_{i=2}^I e^{\lambda V_i}} \\
 &= \frac{e^{\lambda V_1}}{\sum_{i=1}^I e^{\lambda V_i}}
 \end{aligned}$$

Die Herleitung lässt sich analog für P_2, P_3 etc machen, so dass die Auswahlwahrscheinlichkeiten P_i durch (4.34) gegeben sind.

Direkte Herleitung der Multinomial-Logit-Wahrscheinlichkeiten (4.34)

Diese elegante Herleitung stammt von Kenneth E. Train⁷³ Setzt man in den Integralausdruck (4.19) der Auswahlwahrscheinlichkeit P_1 (o.E.d.A) für allgemeine unabhängige Zufallsnutzen die Gumbelverteilung ein, ergibt sich zunächst das Integral

$$\begin{aligned}
 P_1 &= \int_{-\infty}^{\infty} \prod_{i=2}^I F_i(\epsilon_1 + V_1 - V_i) f_1(\epsilon_1) d\epsilon_1 \\
 &= \int_{-\infty}^{\infty} \prod_{i=2}^I \exp\left(-e^{-(\epsilon_1 + V_1 - V_i)}\right) e^{-\epsilon_1} \exp(-e^{-\epsilon_1}) d\epsilon_1.
 \end{aligned}$$

Man sieht, dass der Faktor $\exp(-e^{-\epsilon_1})$ gerade einem zusätzlichen Faktor des Produkts $\prod_{i=2}^I(\cdot)$ für $i = 1$ bedeutet:

$$P_1 = \int_{-\infty}^{\infty} \prod_{i=1}^I \exp\left(-e^{-(\epsilon_1 + V_1 - V_i)}\right) e^{-\epsilon_1} d\epsilon_1.$$

⁷³Kenneth E. Train, *Discrete Choice Methods with Simulation*, frei verfügbar unter <http://elsa.berkeley.edu/books/choice2.html>. Kann aber auch als schönes Buch z.B. [hier](#) gekauft werden.

4 Modelle diskreter Entscheidungen

Nun schreiben wir das Produkt von e-Funktionen als Summe im Exponenten:

$$\begin{aligned} P_1 &= \int_{\epsilon_1=-\infty}^{\infty} \exp\left(-\sum_{i=1}^I e^{-(\epsilon_1+V_1-V_i)}\right) e^{-\epsilon_1} d\epsilon_1 \\ &= \int_{\epsilon_1=-\infty}^{\infty} \exp\left(-e^{-\epsilon_1} \sum_{i=1}^I e^{-(V_1-V_i)}\right) e^{-\epsilon_1} d\epsilon_1. \end{aligned}$$

Nun kommt der Trick, um dieses Integral auszuführen: Substituiere $t = e^{-\epsilon_1}$, $dt = -e^{-\epsilon_1} d\epsilon_1$. Die Integrationsgrenzen werden damit zu

$$\begin{aligned} \epsilon_1 = -\infty &\Rightarrow t = \infty, \\ \epsilon_1 = +\infty &\Rightarrow t = 0. \end{aligned}$$

also

$$\begin{aligned} P_1 &= \int_{t=\infty}^0 \exp\left(-t \sum_{i=1}^I e^{-(V_1-V_i)}\right) (-dt) \\ &= \int_{t=0}^{\infty} \exp\left(-t \sum_{i=1}^I e^{-(V_1-V_i)}\right) dt \\ &= \left[\frac{-1}{\sum_{i=1}^I e^{-(V_1-V_i)}} \exp\left(-t \sum_{i=1}^I e^{-(V_1-V_i)}\right) \right]_0^{\infty} \\ &= \frac{1}{\sum_{i=1}^I e^{-(V_1-V_i)}} \\ &= \frac{e^{V_1}}{\sum_{i=1}^I e^{V_i}}, \end{aligned}$$

q.e.d. Für P_2 usw. geht die Rechnung ganz analog.

Herleitung der allgemeinen Kalibrierungsbedingungen (4.71) für das MNL-Modell

Aus (4.70) ergibt sich durch Ableiten

$$\begin{aligned} \frac{d\tilde{L}(\beta)}{d\beta_m} &= \sum_{n,i} y_{ni} \left(\frac{dV_{ni}}{d\beta_m} - \frac{1}{\sum_{i'} e^{V_{i'i}}} \sum_{i''} e^{V_{ni''}} \frac{dV_{ni''}}{d\beta_m} \right) \\ &= \sum_{n,i} y_{ni} \frac{dV_{ni}}{d\beta_m} - \sum_n \left(\sum_i y_{ni} \right) \left(\frac{\sum_{i''} e^{V_{ni''}} \frac{dV_{ni''}}{d\beta_m}}{\sum_{i'} e^{V_{i'i}}} \right) \end{aligned}$$

4 Modelle diskreter Entscheidungen

In der zweiten Zeile haben wir ausgenutzt, dass der zweite Summand der runden Klammer der ersten Zeile gar nicht von i abhängt. Insbesondere können wir nun nach (4.37) die Summe über i mit $\sum_i y_{ni} = y_n$, also mit der Gesamtzahl der Entscheidungen in Personengruppe n , identifizieren. Damit können wir die Summe über die Laufvariable i'' wieder als eine über i schreiben:

$$\frac{d\tilde{L}(\boldsymbol{\beta})}{d\beta_m} = \sum_{n,i} y_{ni} \frac{dV_{ni}}{d\beta_m} - \sum_{n,i} y_n \sum_i \left(\frac{e^{V_{ni}}}{\sum_{i'} e^{V_{ni'}}} \right) \frac{dV_{ni}}{d\beta_m}.$$

Nun kann man mit Hilfe von (4.34) den Faktor in der runden Klammer im zweiten Summanden als die MNL-Wahrscheinlichkeit P_{ni} identifizieren. Damit erhält man

$$\begin{aligned} \frac{d\tilde{L}(\boldsymbol{\beta})}{d\beta_m} &= \sum_{n,i} y_{ni} \frac{dV_{ni}}{d\beta_m} - \sum_{n,i} y_{ni} P_{ni} \frac{dV_{ni}}{d\beta_m} \\ &= \sum_{n,i} (y_{ni} - y_n P_{ni}) \frac{dV_{ni}}{d\beta_m}, \end{aligned}$$

also die Beziehung (4.71).

Herleitung der Beziehung (4.80) für die Elemente der Hessematrix

Zunächst erhält man durch Ableiten nach der Kettenregel:

$$\begin{aligned} H_{jm} &= \frac{\partial g_j}{\partial \beta_m} \\ &= \frac{d}{d\beta_m} \left(\sum_i x_{nmi} y_i P_{ni}(\boldsymbol{\beta}) - \bar{x}_j \right) \\ &= \sum_i x_{nmi} y_i \frac{dP_{ni}}{d\beta_m}. \end{aligned}$$

Die Ableitungen der MNL-Wahrscheinlichkeiten nach den Parametern ergeben sich zu

4 Modelle diskreter Entscheidungen

$$\begin{aligned}
\frac{dP_{ni}}{d\beta_m} &= \frac{d}{d\beta_m} \left(\frac{1}{\sum_{i'} e^{V_{i'} - V_{ni}}} \right) \\
&= -P_{ni}^2 \frac{d}{d\beta_m} \left(\sum_{i'} e^{V_{i'} - V_{ni}} \right) \\
&= -P_{ni}^2 \sum_{i'} e^{V_{i'} - V_{ni}} \frac{d}{d\beta_m} (V_{i'} - V_{ni}) \\
&= -P_{ni}^2 \sum_{i'} e^{V_{i'} - V_{ni}} \frac{d}{d\beta_m} \sum_j \beta_m (x_{ji'} - x_{nmi}) \\
&= -P_{ni}^2 \sum_{i'} e^{V_{i'} - V_{ni}} (x_{mi'} - x_{nmi}) \\
&= P_{ni}^2 \left(x_{nmi} \sum_{i'} e^{V_{i'} - V_{ni}} - \sum_{i'} x_{mi'} e^{V_{i'} - V_{ni}} \right) \\
&= P_{ni} \left(x_{nmi} - \sum_{i'} x_{mi'} P_{i'} \right).
\end{aligned}$$

Hierbei wurde in der letzten Zeile wieder Gl. (4.34) für die MNL-Wahrscheinlichkeiten mehrfach ausgenutzt: $\sum_{i'} e^{V_{i'} - V_{ni}} = 1/P_{ni}$ und $e^{V_{i'} - V_{ni}} = P_{i'}/P_{ni}$. Setzt man die letzte Zeile in den Ausdruck für H_{jm} ein, ergibt sich

$$H_{jm} = \sum_{n,i} x_{nmi} y_i P_{ni} \left(x_{nmi} - \sum_{i'} x_{mi'} P_{i'} \right),$$

also die Beziehungen (4.80) des Haupttextes.

Herleitung der Hessematrixelemente (4.81) für den binomialen Spezialfall

!!! Wie im Haupttext begründet, kann die gesamte Nutzenfunktion auf Alternative 1 verlagert werden:

$$V_{j1i} = \sum_j \beta_m x_{j1i}, \quad V_{2li} = 0$$

Damit ist auch $x_{2li} = 0$ und die Summen \sum_i und $\sum_{i'}$ in Gl. (4.80) beschränken sich auf jeweils einen Summanden, $i = 1$. Damit wird (4.80) zu

$$\begin{aligned}
H_{jm} &= \sum_i x_{j1i} y_i P_{1i} (x_{m1i} - x_{m1i} P_{1i}) \\
&= \sum_i x_{j1i} y_i P_{1i} x_{m1i} (1 - P_{1i})
\end{aligned}$$

und mit $1 - P_{n1} = P_{n2}$ schließlich Beziehung (4.81).

4 Modelle diskreter Entscheidungen

4 Modelle diskreter Entscheidungen

4.17 Verwendete Symbole

$n = 1, \dots, N$	Personenindex
$i = 1, \dots, I$	Index der einzelnen Alternativen
$A_n = \{a_{ni}\}$	Alternativenmenge für Person n
$j = 1, \dots, J_c$ bzw. J_S	Index der J_c Charakteristika (generische exogene Variable) bzw. der J_S sozioökonomischen exogenen Variablen
$C_{ni} = \{c_{jni}\}$	Satz der "Charakteristika" für Person n bei Alternative i
$S_n = \{s_{jn}\}$	Satz von sozioökonomischen Variablen für Person n
$m = 1, \dots, M$	Index der exogenen parameterlinearen Einflussfaktoren
$g_{mi}(\mathbf{C}, \mathbf{S})$	Kriteriums-Funktion zum Bilden des m -ten parameterlinearen exogenen Einflussfaktors für Alternative i ; i.A. nichtlineare Funktionen der Charakteristika und der sozioökonomischen Variablen
x_{nmi} $= g_{mi}(\mathbf{C}_{ni}, \mathbf{S}_i)$	m -ter Einflussfaktor für Person n bei Alternative i (exogene Variable der parameterlinearen deterministischen Nutzenfunktionen)
\mathbf{x}_{ni}	Vektor aller Einflussfaktoren
y_{ni}	Beobachtete Wahlentscheidung: $y_{ni} = 1$, wenn Person n Alternative i gewählt hat, sonst $y_{ni} = 0$ (endogene Variable des vollen stochastischen Modells). $y_{ni} > 1$ möglich bei Zusammenfassung zu Personengruppen.
$y_n = \sum_i y_{ni}$	Zahl an Personen bzw. Entscheidungen bei aggregierter Darstellung, bei der der Index n nicht einzelne Personen sondern Klassen bezeichnet (sonst $y_n = 1$).
P_{ni}	Wahrscheinlichkeit dafür, dass Person n Alternative i wählt (endogene Variable des deterministischen Teilmodells)
$N_i = \sum_n y_{ni}$	Gesamtzahl der Entscheidungen für Alternative i
$U_{ni} = V_{ni} + \epsilon_{ni}$	Gesamtnutzen als Summe von deterministischen Nutzen und Zufallsnutzen
$V_{ni} = \sum_m \beta_m x_{nmi}$ $= \boldsymbol{\beta}' \mathbf{x}_{ni}$	Deterministischer Nutzen für Person n bei Alternative i in parameterlineare Formulierung
δ_{ij}	"Selektor"-Dummyvariable. $\delta_{ij} = 1$ falls $i = j$ und $=0$ sonst.
$\boldsymbol{\beta} = \{\beta_m\}, \hat{\boldsymbol{\beta}}$	Satz von Modellparametern, dazugehörige Schätzer

4 Modelle diskreter Entscheidungen

η, λ	Lageparameter (Modalwert) und Skalierungsparameter (proportional zur inversen Standardabweichung) der Gumbel-Verteilung
σ_ϵ	Standardabweichung des Zufallsnutzens
$\Phi(x)$	(kumulierte) Verteilungsfunktion der Standardnormalverteilung
$\epsilon_{nii}^{(\text{mic}, x_m)}$ $= \frac{x_{nmi}}{P_i} \frac{\partial P_i}{\partial x_{nmi}}$	Mikroskopische Eigenelastizitäten bezüglich des Faktors x_m für Person n und Alternative i
$\epsilon_{nij}^{(\text{mic}, x_m)}$ $= \frac{x_{nmj}}{P_i} \frac{\partial P_i}{\partial x_{nmj}}$	Mikroskopische Kreuzelastizitäten bezüglich des Faktors x_m für Person n für Alternative i aufgrund Attributänderung der Alternative j
$\epsilon_{ii}^{(\text{mac}, X_m)}$ $= \frac{\bar{x}_{mi}}{N_i} \frac{\partial N_i}{\partial \bar{x}_{mi}}$	Makroskopische Eigenelastizität bezüglich des Faktors x_m (\bar{x}_{mi} ist ein geeignetes gewichtetes Mittel über die Einzelattribute der Personen n).
$L(\beta)$	Likelihoodfunktion
$\tilde{L}(\beta) = \ln L(\beta)$	Log-Likelihood
H	Hesse-Matrix bzw. Informationsmatrix mit den Komponenten $H_{mm'} = - \frac{\partial^2 \tilde{L}(\hat{\beta})}{\partial \beta_m \partial \beta_{m'}}$
V	Varianz-Kovarianzmatrix der Parameterschätzer
$V(\hat{\beta}_m) = V_{mm}$	Varianz von $\hat{\beta}_m$
$\rho_{mm'}$	Korrelationskoeffizient zwischen den Parameterschätzern
z_q	q -Quantil (Inverses der Verteilungsfunktion) der Standardnormalverteilung.
$\chi_{q,n}^2$	q -Quantil (Inverses der Verteilungsfunktion) der χ^2 -Verteilung mit n Freiheitsgraden.
α	Fehlerwahrscheinlichkeit
χ^{LR}	Test-Statistik des Likelihood-Ratio-Tests (doppelte Differenz der Log-Likelihoods des vollen und des restringierten Modells).
$\rho^2, \bar{\rho}^2$	Likelihood-Ratio (LR) Index bzw. korrigierter LR-Index
I_l, λ_l	Nested-Logit-Modell: Inklusivwert und Logit-Skalenparameter ("Inklusivwertparameter") des Nests (Gruppe von untergeordneten Entscheidungen) l

5 Ökonometrische Verflechtungsmodelle

5.1 Allgemeines

Eine wichtige Aufgabenstellung der Ökonometrie und insbesondere der Verkehrsökonomie ist die quantitative Beschreibung der Ströme von Waren, Produkten und sonstiger wirtschaftlicher Leistungen vom Anbieter (Output) zum Empfänger (Input). Dies ist nichttrivial, da Produkte und Dienstleistungen nicht nur vom Endverbraucher konsumiert werden, sondern auch von den Herstellern/Anbietern dieser oder anderer Produkte und Dienstleistungen.

Ein aktuelles Beispiel ist die Frage, wie viel CO_2 die Atomkraftwerke (AKW) effektiv zur Herstellung einer kWh an Strom erzeugen (Abb. 5.1 und 5.2). Die Frage ist nun: Wie hoch ist die tatsächliche Kohlendioxidbilanz (g CO_2 pro kWh), wenn man nicht nur den direkten Output, sondern auch die unendliche Kette all dieser indirekten Emissionen berücksichtigt? Wie hoch ist diese Bilanz im Vergleich der Gesamtbilanz für Kohle- oder Gaskraftwerken oder bei der Erzeugung erneuerbarer Energie?¹

Generell bewirken diese Kopplungen, dass Nachfrageschwankungen bezüglich eines Wirtschaftsbereiches auch die anderen Sektoren beeinflussen. Beispielsweise bewirkt eine steigende Nachfrage nach Kraftfahrzeugen auch eine Konjunktur bei den Herstellern von Stahl, beim Maschinenbau und bei der Kfz-Elektronik. Dadurch ergibt sich folgende Fragestellung:

Problemformulierung: Wie ändern sich die Nachfragen nach Produkten oder Dienstleistungen der verschiedenen Wirtschaftszweige, wenn sich die Nachfrage der Endverbraucher an Leistungen eines bestimmten Wirtschaftszweiges um einen bestimmten Betrag ändert?

5.2 Formulierung des Verflechtungsmodells

Bemerkenswerterweise lassen sich im linearen statischen Grenzfall all diese Verflechtungen und Kopplungen sehr kompakt durch das ökonometrische **Verflechtungsmodell**, auch **Input-Output-Modell (IOM)** oder nach seinem Begründer Wassily Leontief **Leontief-Modell** genannt, beschreiben. Im Rahmen dieses Modells gliedert man die gesamte Volkswirtschaft in verschiedene Sektoren $i = 1, \dots, n$ auf und modelliert die

¹Die Durchführung der Rechnung ergibt für AKWs etwa 40 g CO_2/kWh , für Kohlekraftwerke etwa 800 g CO_2/kWh und für alternative Energiequellen 50-100 g CO_2/kWh .

5 Ökonometrische Verflechtungsmodelle

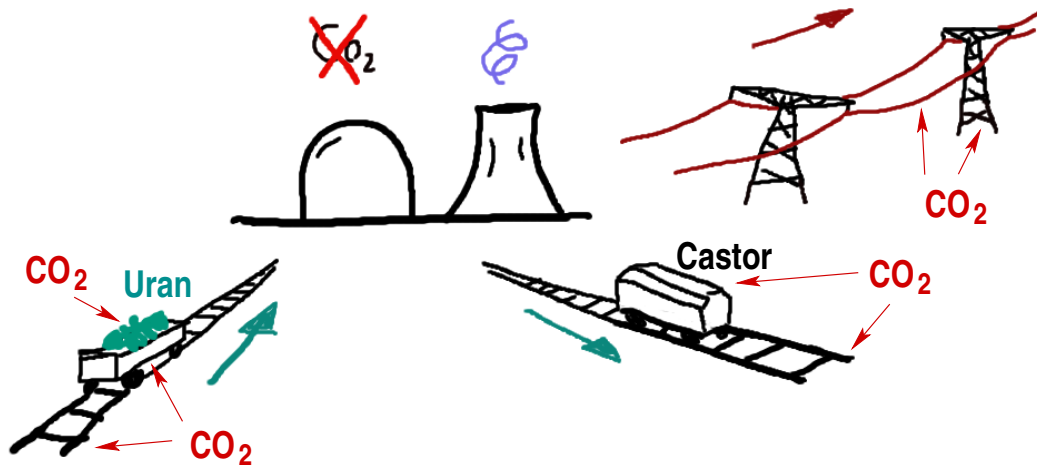


Abbildung 5.1: Atomkraftwerke (AKW) emittieren bei der eigentlichen Stromerzeugung durch Kernspaltung keinerlei CO_2 . Aber zum Betrieb eines AKW sind andere Produkte und Dienstleistungen nötig, deren Bereitstellung sehr wohl mit der Emission von CO_2 verbunden ist: Rohstoffgewinnung, die Aufbereitung des Rohstoffes zum Brennstoff, Entsorgung, der Transport von Roh- Brenn- und Abfallstoffen sowie der elektrischen Energie, Bereitstellung und Betrieb der Infrastruktur zum Transport von Warenströmen und elektrischer Energie (Hochspannungsleitung), aber auch anteilig der Bau sowie der Betrieb (u.a. Heizung) der eigentlichen Bauwerke der Anlage usw. Diese Kette muss man, streng genommen, ad infinitum fortsetzen: Unter anderem sollten auch die Emissionen durch die Heizung der Verwaltungsgebäude der Firmen, welche die Anlage herstellen oder den Stahl für die Strommasten liefern, anteilig enthalten sein.

5 Ökonometrische Verflechtungsmodelle

Waren- und Dienstleistungsströme x_{ij} vom Sektor i nach j sowie die Ströme y_i vom Sektor i zum Endverbraucher² durch eine *lineare, zeitunabhängige, deterministische* Kopplung:

$$x_i = y_i + \sum_{j=1}^n x_{ij} = y_i + \sum_{j=1}^n A_{ij}x_j. \quad (5.1)$$

Hierbei bedeuten (vgl. auch Abb. 5.2):

- x_i der Gesamtausstoß an Produkten oder Leistungen im Sektor i in mit den anderen Sektoren kommensurablen Einheiten (im Allgemeinen Geldeinheiten),
- x_{ij} die Menge an Produkten oder Dienstleistungen des Sektors i , welche für Sektor j benötigt wird,
- y_i die direkt von Sektor i an den Endverbraucher bzw. an nicht direkt berücksichtigte Sektoren gehenden Produkte oder Leistungen,
- A_{ij} die Elemente der **Koeffizientenmatrix des direkten Aufwandes**, also die Zahl der Einheiten des Produkts/der Dienstleistung i , welche zur Herstellung/Erbringung einer Einheit des Produkts/der Dienstleistung des Sektors j benötigt wird.

Die verschiedenen Sektoren werden beispielsweise durch die Input-Output-Rechnung des Bundes definiert. In der aktuellsten Ausgabe 2009 werden 73 Sektoren unterschieden³. Das IOM ist durch folgende Annahmen charakterisiert:

- Alle Güter- und Dienstleistungsströme müssen sich *genau einem Sektor* zuordnen lassen, z.B. einem der 73 Sektoren der Input-Output-Rechnung des Bundes.
- Alle Ströme werden in *Geldeinheiten* pro Zeiteinheit gemessen.
- Es wird *Stationarität* angenommen, d.h. alle Ströme werden aus der laufenden Produktion gespeist. Lagerbestände sowie ggf. die Altersstruktur der Güter (z.B. beim Fahrzeugsektor) bleiben unverändert.
- Es wird *Linearität* angenommen. Skaleneffekte, beispielsweise Effizienzsteigerungen durch Massenproduktion, bleiben unberücksichtigt.

Im Rahmen der allgemeinen Klassifikation ökonometrischer Modelle stellt das Verflechtungsmodell ein lineares, statisches, deterministisches Modell mit Kopplung dar:

²Der Begriff des "Endverbrauchers" kann auch allgemeiner gefasst werden und neben dem eigentlichen Konsumenten auch alle externen Sektoren umfassen, die man in einer gegebenen Untersuchung nicht separat erfassen will.

³Statistisches Bundesamt (2013): Volkswirtschaftliche Gesamtrechnungen – Input- Output-Rechnung 2009, <https://www.destatis.de/> (Zugriff 06.08.2013).

5 Ökonometrische Verflechtungsmodelle

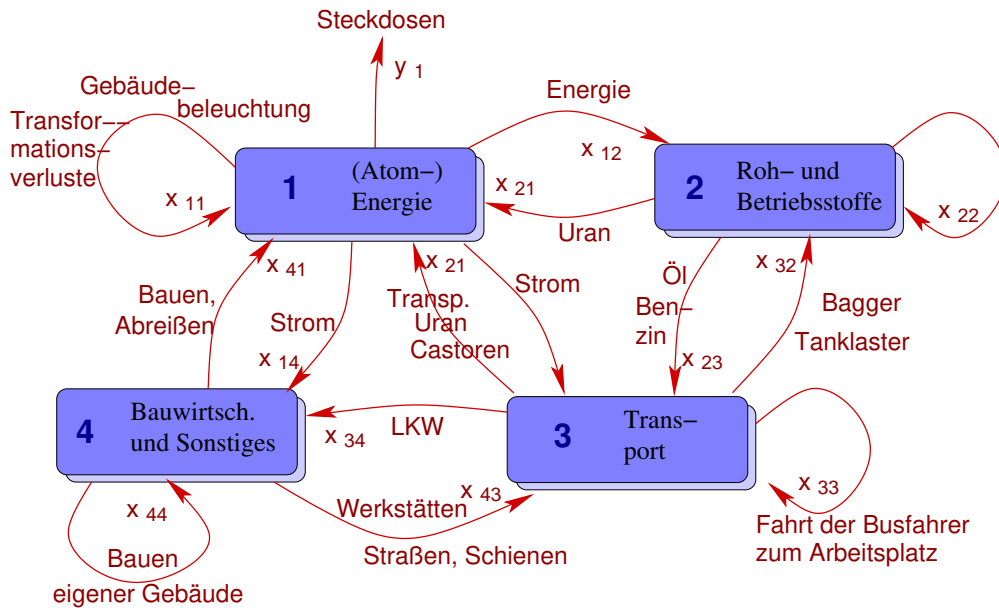


Abbildung 5.2: Auswahl an Güterströmen im AKW-Beispiel.

- Die endogenen Variablen⁴ sind durch die Gesamtproduktionen x_i der Sektoren gegeben. Die verschiedenen x_i sind gekoppelt, da sie auch auf der rechten Seite der Gleichungen vorkommen.
- Die exogenen Variablen sind durch die externen Nachfragen y_i gegeben.
- Die Modellparameter sind durch die Elemente A_{ij} gegeben.

Je nach Sachverhalt ist es sinnvoll, die x_i und y_i entweder als **Bestandsmassen** oder **Bewegungsmassen** zu definieren.

- Bewegungsmassen sind nur über *Zeitintervalle* definiert und nehmen im **Fließgleichgewicht** linear mit der Länge des Intervalls zu. Es handelt sich um Gütermengen, welche z.B. durch den Wert in € angegeben werden.
- Bestandsmassen sind für *Zeitpunkte* definiert. Es handelt sich um Güterströme, welche z.B. durch die “Stromstärke” in €/h gemessen werden. Im Fließgleichgewicht sind sie konstant.

⁴Um Kompatibilität mit den sonstigen Darstellungen des IOM zu bewahren, bedeuten hier ausnahmsweise y_i die exogenen und x_i die endogenen Variablen. Ohnehin ist die Interpretation “exogen” bzw. “endogen” hier nicht eindeutig zu treffen. Handelt es sich um eine rein nachfragegetriebene (Keynesianische) Wirtschaft, ist der Nachfragevektor \mathbf{y} exogen (die Nachfrage bestimmt das Angebot \mathbf{x}). Bei angebotsgetriebenen Märkten (“der Kunde muss essen, was auf den Tisch kommt”) ist es umgekehrt.

5 Ökonometrische Verflechtungsmodelle

Häufig ist es am anschaulichsten, alles als Bestandsmassen bei festen Betrachtungszeitintervall anzusehen, vgl. die Beispiele.

Die Modellparameter, also die **Koeffizienten A_{ij} des direkten Aufwandes**, werden an den ökonomischen Sachverhalt angepasst, indem die tatsächlichen Verhältnisse in die Definitionsgleichung

$$A_{ij} = \frac{x_{ij}}{x_j} \quad (5.2)$$

eingesetzt werden bzw. das Modell bezüglich dieser Parameter kalibriert wird. In Worten bedeutet Gl. (5.2):

Der direkte Aufwandskoeffizient A_{ij} ist die Zahl der Einheiten des Wirtschaftsgut i , welche direkt zur Herstellung einer Einheit des Produktes j bzw. Erbringung einer Einheit der Dienstleistung j benötigt wird.

Verständnisfrage:

Machen Sie sich klar, dass im Falle der Verwendung kommensurabler Einheiten (z.B. der Waren- oder Dienstleistungswert in €) die Koeffizient des direkten Aufwandes nur im Bereich $0 \leq A_{ij} < 1$ liegen dürfen.

Antwort: Da sonst nach Gl. (5.1) ein Teil des Güterstroms größer ist als der gesamte Strom.

5.2.1 Kompakte Formulierung in Vektor-Matrix-Notation

Das Verflechtungsmodell 5.1 stellt ein lineares inhomogenes Gleichungssystem für die unbekanntes Gesamtproduktionen x_i bei gegebenen externen Nachfragen y_i dar. Zur Darstellung der formalen (und auch der expliziten) Lösung dieses Systems werden die Produktionen und externen Nachfragen durch Spaltenvektoren dargestellt:

$$\begin{aligned} \mathbf{x}^T &= (x_1, x_2, \dots, x_n), \\ \mathbf{y}^T &= (y_1, y_2, \dots, y_n). \end{aligned}$$

Hierbei bedeutet, wie im Kapitel 2, das Superskript T, "transponiert". Damit kann man (5.1) schreiben als

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad (5.3)$$

oder

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{x} + \mathbf{y}. \quad (5.4)$$

5.3 Lösung des Verflechtungsmodells

Mit der Einheitsmatrix

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \dots & \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ & \dots & 0 & 1 \end{pmatrix}$$

und

$$\mathbf{M} = \mathbf{1} - \mathbf{A}$$

kann man Gl. (5.4) schreiben als

$$\mathbf{M} \cdot \mathbf{x} = \mathbf{y}.$$

Ist \mathbf{M} invertierbar, gibt es also eine Inverse \mathbf{M}^{-1} mit $\mathbf{M}^{-1} \cdot \mathbf{M} = \mathbf{M} \cdot \mathbf{M}^{-1} = \mathbf{1}$, so kann man diese Gleichung durch Links-Multiplikation von \mathbf{M}^{-1} "lösen":

$$\begin{aligned} \mathbf{M}^{-1} \cdot \mathbf{M} \cdot \mathbf{x} &= \mathbf{M}^{-1} \cdot \mathbf{y}, \\ \mathbf{x} &= \mathbf{M}^{-1} \cdot \mathbf{y}. \end{aligned}$$

Dies führt nach Einsetzen der Definition von \mathbf{M} zur formalen Lösung des Verflechtungsmodells:

$$\mathbf{x} = (\mathbf{1} - \mathbf{A})^{-1} \cdot \mathbf{y} \equiv \mathbf{B} \cdot \mathbf{y}. \quad (5.5)$$

Hierbei wird die neu definierte Matrix $\mathbf{B} = (\mathbf{1} - \mathbf{A})^{-1}$ als die **Koeffizientenmatrix des vollen Aufwandes** bezeichnet. Die Bedeutung ihrer Elemente ist wie folgt:

B_{ij} ist die Menge an Wirtschaftsgut i , welche zur Herstellung einer Einheit des Wirtschaftsgutes j vom Sektor j selbst *und von allen Zuliefersektoren k des Sektors j* benötigt wird. Also die Menge an Wirtschaftsgut i , welche zur Bereitstellung einer Einheit von j *an den Verbraucher* benötigt wird.

5.3.1 Veranschaulichung der Koeffizientenmatrix des vollen Aufwandes

Ist eine Funktion durch eine Reihenentwicklung definiert, wie $e^x = 1 + x + \frac{x^2}{2} + \dots$ oder auch z.B. $\cos(x)$, so sind als Argumente dieser Funktion nicht nur Zahlenwerte möglich, sondern alle Objekte, welche sich addieren, multiplizieren und invertieren lassen, wobei das Ergebnis dieser Operationen in der Objektmenge bleiben muss. Dies ist insbesondere für die Menge der quadratische Matrizen erfüllt.⁵ Insbesondere gilt dies für die "geometrische Reihe"

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots = \sum_{j=0}^{\infty} x^j.$$

⁵Genauer muss die sog. "Gruppeneigenschaft" bezüglich Addition und Multiplikation erfüllt sein, das Ergebnis dieser Operationen muss also wieder ein Element aus der betrachteten Objektmenge sein; dies ist z.B. bei Vektoren mit dem Skalarprodukt als Multiplikation nicht erfüllt.

5 Ökonometrische Verflechtungsmodelle

Schreibt man die Inverse einer Matrix \mathbf{M} als $\mathbf{M}^{-1} = "1/\mathbf{M}"$ und $\mathbf{M} \cdot \mathbf{M} = \mathbf{M}^2$, $\mathbf{M} \cdot \mathbf{M} \cdot \mathbf{M} = \mathbf{M}^3$ usw., sowie $\mathbf{M}^0 = \mathbf{1}$, sieht man, dass die Lösung des Input-Output-Modells nichts anderes ist als die geometrische Reihe, angewandt auf die Matrix \mathbf{A} :

$$\mathbf{B} = (\mathbf{1} - \mathbf{A})^{-1} = " \frac{1}{\mathbf{1} - \mathbf{A}} " = \mathbf{1} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots = \sum_{j=0}^{\infty} \mathbf{A}^j. \quad (5.6)$$

Im Folgenden wird anhand von zwei Sektoren gezeigt, wie diese unendliche Reihe die unendliche Sequenz der Kopplungen repräsentiert.

Beispiel: Zwei-Sektoren-Modell

Es werden die beiden Sektoren

- Sektor 1: Transport und Verkehr,
- Sektor 2: Maschinenbau

betrachtet. Ferner werden alle anderen Sektoren zu den externen Verbrauchern, also zu den Konsumenten y_i von Verkehr und Maschinenbau, gezählt. Es soll gelten

$$\mathbf{A} = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Die Sektoren verbrauchen also je 25% ihrer Produkte für den Eigenbedarf und benötigen außerdem Produkte des jeweils anderen Sektors in einer Menge, welche ein Viertel des Wertes der gesamten eigenen Produktion entspricht. Fortgesetzte Matrixmultiplikation ergibt für $n \geq 1$

$$\mathbf{A}^n = \frac{1}{2^{n+1}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

und damit z.B.

$$B_{11} = \left(\sum_{j=0}^{\infty} \mathbf{A}^j \right)_{11} = 1 + \frac{1}{4} \sum_{j=0}^{\infty} \left(\frac{1}{2} \right)^j = 1 + \frac{1}{4 \left(1 - \frac{1}{2} \right)} = \frac{3}{2},$$

und

$$B_{12} = \frac{1}{4} \sum_{j=0}^{\infty} \left(\frac{1}{2} \right)^j = \frac{1}{2}.$$

Damit kann man die Kopplungen der Wirtschaftssektoren Verkehr und Maschinenbau veranschaulichen: Der Koeffizient B_{11} hat die Anteile

$$B_{11} = (\mathbf{1} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots)_{11} = 1 + A_{11} + \sum_{k=1}^2 A_{1k} A_{k1} + \sum_{k=1}^2 \sum_{l=1}^2 A_{1k} A_{kl} A_{l1} + \dots$$

mit den Bedeutungen

5 Ökonometrische Verflechtungsmodelle

- Beitrag 1: Eine Einheit an Verkehrsleistung an die Endverbraucher.
- Beitrag $A_{11} = \frac{1}{4}$: Für eine Einheit an den Verbraucher benötigen die Verkehrsbetriebe $1/4$ an Verkehrsleistungen für den Eigenbedarf (Arbeitswege des Personals, Überführung von Fahrzeugen, Wartungsfahrten, Leerfahrten zur ersten Einsatzstelle etc).
- Beitrag A_{11}^2 : Der Eigenbedarf erhöht die Verkehrsnachfrage, welche wiederum *zusätzlichen* Eigenbedarf hervorruft. Ein Straßenbahnfahrer fährt z.B. mit ÖV zum Startpunkt seiner Schicht; aber auch die Fahrer dieser Verkehrsmittel nehmen Verkehrsleistung zum Antritt und am Ende *ihrer* Schicht in Anspruch.
- Beitrag $A_{12}A_{21}$: Die Verkehrsmittel altern und müssen kontinuierlich durch neue ersetzt werden (Anteilige Aufträge A_{21} an den Maschinenbau). Der Maschinenbau benötigt für seinen Betrieb (Mitarbeiterwege, Transport von Teilen etc) für eine Einheit eine Verkehrsleistung von A_{12} Einheiten, also für den an die Verkehrsbetriebe gelieferten Bruchteil A_{21} anteilig die Verkehrsleistung $A_{12}A_{21}$.
- Die Ströme höherer Ordnung werden schnell unübersehbar. Beispielsweise bedeutet der Beitrag $A_{12}A_{22}A_{21}$ die Verkehrsleistung A_{12} an den Maschinenbau, damit dieser seinen eigenen Maschinenpark up-to-date hält (A_{22}), um schließlich neue Fahrzeuge an den Verkehrssektor liefern zu können (A_{21}).

Aufgabe

Verdeutlichen Sie sich die Anteile von B_{12} .

5.3.2 Explizite Lösungen für zwei und drei Sektoren

Glücklicherweise nimmt man die Inversion einer einzigen Matrix das Verfolgen der einzelnen Verflechtungen zwischen den Sektoren nach der Art des letzten Unterabschnitts ab. "Per Hand" kann man die Inverse einer $n \times n$ Matrix bis maximal für $n = 5$ sinnvoll mit dem **Gauß'schen Eliminationsverfahren** lösen, numerisch ist die Lösung auch für Hunderte von Sektoren problemlos (der Aufwand steigt mit n^3). Für 2×2 und 3×3 - Matrizen gibt es einfache explizite Formeln, vgl. Abschnitt 2.11.

Zwei Sektoren

Sei \mathbf{M} eine allgemeine invertierbare 2×2 Matrix mit den Elementen M_{ij} . Dann gilt nach (2.126)

$$\mathbf{M}^{-1} = \frac{1}{\det(\mathbf{M})} \begin{pmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{pmatrix}$$

5 Ökonometrische Verflechtungsmodelle

und damit

$$\mathbf{B} = (\mathbf{1} - \mathbf{A})^{-1} = \frac{1}{\det(\mathbf{1} - \mathbf{A})} \begin{pmatrix} 1 - A_{22} & A_{12} \\ A_{21} & 1 - A_{11} \end{pmatrix}. \quad (5.7)$$

Das obige Beispiel für $A_{11} = A_{12} = A_{21} = A_{22} = \frac{1}{4}$ liefert

$$\mathbf{1} - \mathbf{A} = \frac{1}{4} \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}, \quad \det(\mathbf{1} - \mathbf{A}) = \frac{1}{16}(9 - 1) = \frac{1}{2}, \quad \mathbf{B} = \frac{1}{2} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

also insbesondere die oben mittels der geometrischen Reihe berechneten Koeffizienten des vollen Aufwands $B_{11} = \frac{3}{2}$ und $B_{12} = \frac{1}{2}$.

Drei Sektoren

Hier benutzt man Gleichung (2.127). Alternativ kann man auch die Eigenschaften von Skalar- und Kreuzprodukten von Vektoren nutzen, vgl. (2.128). Schreibt man die drei Zeilen der Inversen einer invertierbaren 3×3 Matrix mit Hilfe von drei Zeilenvektoren,

$$\mathbf{M}^{-1} = \begin{pmatrix} -z_1^T & - & - \\ -z_2^T & - & - \\ -z_3^T & - & - \end{pmatrix},$$

und die Spalten der Matrix selbst mit Spaltenvektoren,

$$\mathbf{M} = \begin{pmatrix} | & | & | \\ m_1 & m_2 & m_3 \\ | & | & | \end{pmatrix}$$

dann gilt

$$z_1 = \frac{m_2 \times m_3}{\det(\mathbf{M})}, \quad z_2 = \frac{m_3 \times m_1}{\det(\mathbf{M})}, \quad z_3 = \frac{m_1 \times m_2}{\det(\mathbf{M})}. \quad (5.8)$$

sowie

$$\det(\mathbf{M}) = m_1 \cdot (m_2 \times m_3). \quad (5.9)$$

Natürlich muss man zur Bestimmung der Matrix \mathbf{B} für \mathbf{M} wieder die Differenzmatrix $\mathbf{1} - \mathbf{A}$ einsetzen.

Aufgabe: Zwei-Sektor-Modell

Im obigen Beispiel der zwei Sektoren Verkehr und Maschinenbau mit den direkten Aufwandskoeffizienten $A_{11} = A_{12} = A_{21} = A_{22} = \frac{1}{4}$ ist bisher die externe Nachfrage nach Transport-Dienstleistungen doppelt so hoch wie die nach Produkten des Maschinenbaus. Nun erhöht sich die Nachfrage nach Verkehr um 10%, während die nach Produkten des Maschinenbaus unverändert bleibt. Um wieviel Prozent müssen die Gesamtleistungen der beiden Sektoren steigen, um die neue Nachfrage zu befriedigen?

5.4 Lebenszyklusanalyse und Einbindung des Input-Output-Modells

Ausgehend von einer aktuellen Fragestellung könnte dieser Abschnitt auch lauten: *Sind E-Fahrzeuge wirklich umweltfreundlicher?* Die Lebenszyklusanalyse wird deshalb anhand der CO₂-Emissionen von Fahrzeugen erläutert, sie kann aber für beliebige, Ressourcenverbrauch oder Emissionen betreffende Produkten oder Prozessen eingesetzt werden wie die Emissionen bei Produktion eines Kilowatts elektrischer Energie aus Kohle- oder Atomkraftwerken, Windkraft, Wasserkraft, Biomasse, Solarzellen oder Solarthermie.

Hintergrund ist, dass Elektrofahrzeuge im Betrieb *lokal* weder CO₂ noch andere Schadstoffe (außer Feinstaub) emittieren, weshalb sie von einigen Stadtverwaltern sehr geliebt werden, um lokale Umweltprobleme zu mildern. Aber geht man hier nicht lediglich nach dem St-Florians-Prinzip vor? Dies ist insbesondere für nichtflüchtige Emissionen, allen voran CO₂, relevant. Früher oder später verteilen sie diese weltweit, egal ob sie im Betrieb lokal (Verbrennungsmotoren), nichtlokal (Produktion der Kraftstoffe bei Verbrennungsmotoren, Stromerzeugung bei E-Fahrzeugen, außerdem Verschleißteile und Verbrauchsmaterialien), oder in vor- oder nachgelagerten Prozessen (Herstellung, Entsorgung) anfallen. Dies gilt im abgemilderten Schärfe auch für konventionelle Kfz mit Verbrennungsmotor, bei der neben den direkten CO₂ Emissionen beim Betrieb ebenfalls weitere Emissionen anfallen.

Man benötigt also eine Analyseverfahren, welche *alle* Emissionen erfasst, welche irgendwo oder irgendwann unmittelbar oder mittelbar durch die Nutzung eines Fahrzeugs (oder allgemein bei der Nutzung eines ressourcenverbrauchenden und die Umwelt beeinflussenden Produkts oder Prozesses) anfallen. Dazu werden zwei Methoden kombiniert:

- Die **Lebenszyklusanalyse** bzw. **Life-Cycle-Assessment (LCA)**, auch **Ökobilanz** genannt, umfasst welche alle direkt mit der Herstellung, dem Betrieb und der Entsorgung anfallenden Emissionen umfasst,⁶
- und die Input-Output-Analyse mit dem IOM, welche alle durch Verflechtungseffekte anfallenden indirekten Emissionen berücksichtigt.⁷ Die LCA wird in das Input-Output-Modell meist mittels der hybriden Methodik **Econometric Input-Output LCA (EIO-LCA)** eingebunden.

Betrachtet man nur Emissionen pro End-Nachfrage (in Geldeinheiten) nach Produkten oder Dienstleistungen eines gewissen Sektors, kann man die EIO-LCA vereinfachen zur IOM-LCA.

⁶Beispielsweise LCA nach DIN EN ISO 14040. Die GEMIS Datenbank enthält viele Sachbilanzen und das HBEFA "Handbuch für Emissionsfaktoren" die dazugehörigen Emissionsfaktoren.

⁷Die Abgrenzung ist nicht immer scharf. Beispielsweise werden bei der LCA meist nicht nur die *Tank-to-Wheel* Emissionen durch die direkte Verbrennung berücksichtigt, sondern auch die *Well-to-Tank* Emissionen durch die Produktion und den Transport des Kraftstoffs bis in den Tank. Letztere könnte man auch mittels des IOM durch die Verflechtung des Sektors "Treibstoffe" mit "Transport", "Verarbeitung" und "Bergbau" modellieren.

5 Ökonometrische Verflechtungsmodelle

Abbildung 5.3: Vereinfachtes Flussdiagramm der Lebenszyklusanalyse eines Kraftfahrzeugs bezüglich Emissionen (noch nicht vorhanden, vgl. Vorlesungsmitschrift)

5.4.1 Lebenszyklusanalyse (LCA)

Die Lebenszyklusanalyse verläuft in mehreren Schritten (vgl. Abbildung 5.3):

1. Definition der Lebensphasen: Bei einem Fahrzeug sind das im Wesentlichen

- die Herstellung,
- der Betrieb
- und die Entsorgung.

2. Aufstellung der Sachbilanz: In jeder der Lebensphasen verbraucht einer betrachtete Einheit (hier: ein Auto) gewisse Mengen verschiedener Ressourcen wie

- y_1 : Masse (kg) an Eisen/Stahl,
- y_2 Menge (Liter) an Treibstoff (z.B. Benzin),
- y_3 : Masse (kg) an Aluminium,
- y_4 : Masse (kg) an Kunststoffen,
- y_5 : Masse (kg) an Gummi

und weitere Materialien wie Gummi, Lack, Textilien usw. Über alle Lebensphasen summiert ergibt dies den Vektor y der **Sachbilanz**.

3. Aufstellung der Emissionsfaktoren: Die Bereitstellung der benötigten Materialien und Ressourcen ist mit Emissionen verschiedenster Schadstoffe verbunden. Dies definiert die sachbilanzbezogene Emissionsfaktoren-Matrix **C** :

Bei n betrachteten Emissionsarten (z.B. 1=CO₂, 2=Feinstaub, 3=NO_x) und m Posten der Sachbilanz (z.B. 1=Stahl, 2=Benzin) ist die sachbilanzbezogene Emissionsfaktoren-Matrix **C** eine unsymmetrische reellwertige $n \times m$ -Matrix. Die Komponente C_{ij} gibt an, wieviele Einheiten des Schadstoffes i bei der Herstellung einer Einheit des Postens j der Sachbilanz *einschließlich der kompletten Kette der Vorprodukte* (vgl. Abb. 5.4) im Mittel anfallen.

5 Ökonometrische Verflechtungsmodelle

Die erste Zeile der Matrix betrifft also den ersten Schadstoff und die erste Spalte den ersten Postens der Sachbilanz. Bei der Bereitstellung eines Liters Benzin ($m = 2$) von der Förderung bis zur Tankstelle (*Well-to-Tank*) fallen beispielsweise etwa 0.3 kg an CO_2 an. Das entsprechende Matrixelement ist also (ohne die eigentliche Verbrennung im Motor, siehe Punkt 4) gleich

$$C_{12} = C_{WtT} = 0.3 \text{ kgCO}_2/\text{l}.$$

Meist wird hier nicht direkt CO_2 , sondern das sogenannte **CO₂-Äquivalent** angegeben. Dabei wird die auf einen kg CO_2 bezogene Klimawirksamkeit aller Emissionen betrachtet. Auf einen 100-Jahres-Zeitraum bezogen, wird beispielsweise Methan 25-fach⁸ und Lachgas 298 fach gewichtet.

4. Ermittlung der Gesamtemissionen während der Lebenszeit: Direkt anhand der Definition der Emissionsfaktorenmatrix ist offensichtlich, dass die mit der Herstellung der benötigten Materialien und Verbrauchsstoffe verbundenen Emissionen durch die Multiplikation der Emissionsfaktoren-Matrix mit der Materialbilanz gegeben sind. Hinzu kommen noch die direkt mit dem Verbrauch (hier von Benzin) verursachten Emissionen:

$$\mathbf{e}_{\text{LCA}} = \mathbf{e}^{\text{dir}} + \mathbf{C} \cdot \mathbf{y}^{\text{s}} \quad (5.10)$$

Bei obiger Definitionen der Sachbilanz sind die CO_2 -Emissionen durch die erste Komponente gegeben:

$$e_{\text{CO}_2} = e_1 = e_1^{\text{dir}} + \sum_{j=1}^n C_{1j} y_j^{\text{s}} \quad (5.11)$$

Die direkten Emissionen sind bei Verbrennungsmotorfahrzeugen durch die **Tank-to-Wheel** Emissionen, multipliziert mit der Lauflistung L während der Lebenszeit, gegeben:

$$e_1^{\text{dir}} = C_{\text{TtW}} L \quad (5.12)$$

Die Tank-to-Wheel Emissionen ergeben sich dadurch, dass Benzin oder Diesel im Wesentlichen vollständig verbrennen, dass heißt, jedes Kohlenstoffatom im Treibstoff verbindet sich mit dem Luftsauerstoff zu einem CO_2 -Molekül. Damit ergibt sich

$$C_{\text{TtW}} = \rho f_c \frac{M_{\text{CO}_2}}{M_C} \quad (5.13)$$

mit der Dichte ρ in kg pro Litern, dem Gewichtsanteil f_c an Kohlenstoff im Treibstoff und den Molmassen $M_{\text{CO}_2} = 44 \text{ g}$ sowie $M_C = 12 \text{ g}$. Dies ergibt für Benzin etwa 2.4 kg/l und für Diesel etwa 2.65 kg/l.

⁸Damit wird auch der Grund klar, warum ein Kilogramm Rindfleisch so einen hohen Emissionsfaktor bezüglich CO_2 hat: Vornehm ausgedrückt, sind es die Flatulenzen der Kühe.

5 Ökonometrische Verflechtungsmodelle

Abbildung 5.4: Die gesamte Kette der Well-to-Tank Emissionen bei der Bereitstellung eines Liters an Benzin (noch nicht vorhanden, vgl. Vorlesungsmitschrift).

5. Bezug auf eine Leistungseinheit: Schließlich werden die Lebenszeit-Emissionen auf eine Leistungseinheit, hier also auf einen Kilometer Laufleistung, bezogen:

$$e_{\text{spez}} = \frac{1}{L} e \quad (5.14)$$

Mit üblichen Annahmen ergibt sich bei Kompaktklasse-Benzinfahrzeugen etwa $e_{1,\text{spez}} = 180 \text{ g/km}$, während die direkten Emissionen bei etwa 140 g/km (entspricht einem Verbrauch von 5.8 Liter Benzin auf 100 km) liegen.

Verständnisfrage:

Wie würde man bei der LCA-Analyse eines Elektrofahrzeugs vorgehen? Was sind hier die wichtigsten Posten der Sachbilanz? Wie hoch sind hier die direkten CO_2 -Emissionen e_1^{dir} ?

5.4.2 Econometric Input-Output LCA (EIO-LCA)



Im Allgemeinen wird das Input-Output-Modell für volkswirtschaftliche Fragestellungen angewandt und dabei die gesamte Volkswirtschaft in Sektoren gegliedert. Gelingte es jedoch, die Sachbilanz auf diese Sektoren abzubilden, ist auch eine **produktbezogene Input-Output-Analyse** möglich. Das EIO-LCA verbindet diese Abbildungen mit einer Lebenszyklusanalyse und umfasst folgende Schritte:

1. Definition der Lebensphasen: Wie bei der LCA.

2. Aufstellung der Sachbilanz: Im Wesentlichen wie bei der LCA. Allerdings sollte man zusätzliche rein dienstleistungsorientierte Aktivitäten während eines Autolebens zusätzlich berücksichtigen, beispielsweise Reparatur-Dienstleistungen, was hier als **erweiterte Sachbilanz** bezeichnet wird.

3. Abbildung der Sachbilanz auf den IOM-Nachfragevektor: Aufgrund der Datenlage kann man die volkswirtschaftlichen Sektoren nicht frei wählen, sondern zieht beispielsweise die Einteilung des Statistischen Bundesamtes in 71 Sektoren heran.

Knackpunkt ist es, die m Posten der Sachbilanz \mathbf{y} in die M Einträge des sektorbezogenen Nachfragevektors \mathbf{y} (bzw. in eine Untermenge davon) abzubilden. Im Allgemeinen bezieht sich ein Posten der Sachbilanz auf mehrere Sektoren und mehrere Posten der

5 Ökonometrische Verflechtungsmodelle

Sachbilanz können zur Nachfrage ein- und desselben Sektors beitragen. Nimmt man an, dass der Zusammenhang linear-statisch ist, entspricht die Abbildung daher einer reellwertigen $M \times m$ -Matrix \mathbf{T} :

$$\mathbf{y} = \mathbf{T} \cdot \mathbf{y}^s \quad (5.15)$$

Neben der eigentlichen Abbildung beinhaltet diese Matrix auch eine Transformation auf einheitliche Einheiten, also auf Geldeinheiten.

Im obigen Beispiel bildet die erste Spalte von \mathbf{T} den Sachbilanzposten y_1^s : "Eisen und Stahl" auf die volkswirtschaftlichen Sektoren ab. Verwendet man die 73 Sektoren des Bundes, gibt es im Wesentlichen nur einen von null signifikant unterschiedlichen Eintrag: den Sektor 27.10: "Roheisen, Stahl und Ferrol". Diese Spalte hat also nur einen von null signifikant unterschiedlichen Eintrag, z.B. 3 €/kg, wenn dies der Preis von einem kg Stahl ist. Die zweite Spalte bildet den Sachbilanzposten y_2^s : "Benzin" auf den Sektor 11.00: "Erdöl, Erdgas und -produkte" ab. Der entsprechende Wert des \mathbf{T} -Elements ist der aktuelle Treibstoffpreis in Euro pro Liter. Schließlich werden die Sachbilanzposten Aluminium, Kupfer und Lithium (wichtig als Material für die Batterie von Elektroautos) alle auf den Sektor 27.00: "NE-Metalle" abgebildet. Hier sieht man einen Nachteil des Verfahrens: Materialien mit unterschiedlichen Emissionen bei der Herstellung (die Herstellung eines kg an Kupfer impliziert weniger CO₂ Emissionen als Aluminium oder gar Lithium) werden im Wesentlichen auf den selben Sektor abgebildet und damit gleichbehandelt.⁹

4. Berechnung des Produktionsvektors mit der Standard-IOM: Mit der Abbildung (5.15) der erweiterten Sachbilanz auf den Nachfragevektor ist der Übergang zur IOM vollzogen und die Abbildung auf den Produktionsvektor entspricht der Standardabbildung der IOM mit der Matrix des vollen Aufwandes:

$$\mathbf{x} = \mathbf{B} \mathbf{y} = (\mathbf{1} - \mathbf{A})^{-1} \mathbf{y}$$

Bei Fahrzeugen koppelt die Matrix \mathbf{B} den Nachfragevektor insbesondere stark an den Sektor 38.00: "elektrische Energie", auch wenn dieser bei Benzin-Fahrzeugen nicht direkt in der erweiterten Sachbilanz vorkommt.

5. Ermittlung der Gesamtemissionen: Der Emissionsvektor berechnet sich aus dem Produktionsvektor mittels der sektorbezogenen Emissionsfaktorenmatrix:

⁹Die Skalierung auf den Preis anstelle auf die Masse lindert diesen Nachteil allerdings etwas: Ist Lithium beispielsweise pro kg 10 mal so teuer wie Kupfer, würden auch die Emissionen pro kg um den Faktor 10 höher berechnet.

5 Ökonometrische Verflechtungsmodelle

Die sektorbezogene Emissionsfaktormatrix \mathbf{D} ist eine $n \times M$ -Matrix. Die Komponente D_{ij} gibt an, wieviele Einheiten des Schadstoffes i bei der Herstellung eines Gegenwertes von einem Euro des Gutes j *direkt* anfallen. Im Gegensatz zur Matrix \mathbf{C} werden die Vorketten *nicht* berücksichtigt, da diese ja in viel vollständigerer Form einschließlich aller Rückkopplungen durch die Verflechtungsmatrix \mathbf{B} abgebildet werden.

Nach Definition ergibt dies

$$\mathbf{e}_{\text{IO-LCA}} = \mathbf{e}^{\text{dir}} + \mathbf{D} \mathbf{x} = \mathbf{e}^{\text{dir}} + \mathbf{D} \mathbf{B} \mathbf{y} = \mathbf{e}^{\text{dir}} + \mathbf{D} (\mathbf{1} - \mathbf{A})^{-1} \mathbf{y} \quad (5.16)$$

Diese Hauptformel sieht ähnlich aus wie Gl. (5.10) der konventionellen LCA. Insbesondere werden die direkten Emissionen \mathbf{e}^{dir} gleich berechnet. Es gibt aber folgende Unterschiede:

- Der Nachfragevektor \mathbf{y} bezieht sich auf die volkswirtschaftlichen Sektoren in Euro anstelle auf den Sachbilanzvektor \mathbf{y}^{S} in physikalischen Einheiten
- Die Emissionsfaktormatrix \mathbf{D} enthält nur die direkten Beiträge, nicht die Vorketten wie die in \mathbf{C} enthaltenen produktbezogenen Emissionsfaktoren
- Dafür ist statt der Einheitsmatrix bei der konventionellen LCA die Kopplung der Sektoren untereinander durch die Matrix $\mathbf{B} = (\mathbf{1} - \mathbf{A})^{-1}$ berücksichtigt.

5.4.3 Emissionen bei direkter Verwendung des IOM (IOM-LCA)

Betrachtet man keine Einzelprodukte sondern die durch eine Geldeinheit an Nachfrage in den verschiedenen Sektoren verursachten Emissionen, ist es meist günstiger, alle direkten Emissionen in die Emissionsfaktorenmatrix \mathbf{D} zu stecken. Dazu ist es allerdings notwendig, dass die Ergebnisse von aggregierten Lebenszyklusanalysen für die Produkte bzw. Dienstleistungen der betrachteten Sektoren in Form einer sektorbezogenen direkten Emissionsfaktorenmatrix \mathbf{D} vorliegen. Betrachtet man beispielsweise den Sektor des ÖPNV, werden die direkten Emissionen beim Betrieb der Busse oder Diesel-Loks (elektrifizierte Verkehrsmittel haben ja keine direkten Emissionen sondern nur indirekte über die Verflechtungseffekte bzw. Vorketten!) auf den Fahrpreis umgelegt. Formel (5.16) vereinfacht sich dann zu

$$\mathbf{e}_{\text{OM}} = \mathbf{D} \mathbf{x} = \mathbf{D} \mathbf{B} \mathbf{y} = \mathbf{D} (\mathbf{1} - \mathbf{A})^{-1} \mathbf{y} \quad (5.17)$$

5.4.4 Konsistenz der Methode bei Preissteigerungen

Da man, wie bereits in Abschnitt 5.2 erwähnt, alle Güter und Dienstleistungen in kommensurable Einheiten darstellen muss, werden alle Größen auf den "kleinsten gemeinsamen Nenner" einer Geldeinheit (€) gebracht. Nun kümmert sich die Physik der Emissionen nicht um etwas so Menschliches und Veränderliches wie Preise.¹⁰ Folglich müssen die

¹⁰ Abgesehen davon, dass der Mensch selbst als Reaktion auf Preissteigerungen andere Produktionsmethoden verwendet, beispielsweise bei starkem Ölpreisanstieg eine vermehrte Nutzung sogenannter

5 Ökonometrische Verflechtungsmodelle

nach (5.16) berechneten Emissionen unabhängig von den Preisen der einzelnen Waren und Dienstleistungen sein. Nun sind aber alle Matrizen und Vektoren des IOM und alle auf den rechten Seiten von (5.16) vorkommenden Größen auf Preiseinheiten bezogen, so dass sich diese mit den Preisen ändern. Im Folgenden zeigen wir, dass die Emissionen aber gegenüber Preisänderungen invariant sind, das Modell (5.16) also in dieser Hinsicht konsistent ist.

Wir beschreiben die Preisänderungen in den verschiedenen Sektoren durch einen Faktorvektor \mathbf{f} bzw. eine quadratisch-diagonale Faktormatrix \mathbf{F} ,

$$\mathbf{f} = \begin{pmatrix} \text{Preissteigerungsfaktor Sektor 1} \\ \text{Preissteigerungsfaktor Sektor 2} \\ \vdots \end{pmatrix}, \quad \mathbf{F} = \text{diag } \mathbf{f} = \begin{pmatrix} f_1 & 0 & \dots \\ 0 & f_2 & \\ \vdots & & \ddots \end{pmatrix}, \quad (5.18)$$

welche die Preissteigerungen einer physikalischen bzw. organisatorischen Einheit der verschiedenen Sektoren beschreiben. Bei gleichem Verbrauch ändert sich der Endverbrauchsvektor \mathbf{y} gemäß

$$\mathbf{y} \rightarrow \mathbf{y}^* = \mathbf{F} \mathbf{y},$$

während man direkt aus der Definition folgern kann, dass die neue Koeffizientenmatrix des direkten Aufwands aus der alten hervorgeht, indem man die Zeilen i mit f_i und die Spalten j mit $1/f_j$ multipliziert:

$$\begin{aligned} A_{ij} &= \frac{x_{ij}}{x_j} \Rightarrow \\ A_{ij}^* &= \frac{x_{ij}^*}{x_j^*} \\ &= \frac{f_i x_{ij}}{f_j x_j} = \frac{f_i}{f_j} A_{ij} \end{aligned}$$

Dies kann kompakt durch

$$\mathbf{A} \rightarrow \mathbf{A}^* = \mathbf{F} \mathbf{A} \mathbf{F}^{-1}$$

ausgedrückt werden (Beweis durch Nachrechnen). Zur Berechnung der neuen Koeffizientenmatrix des vollen Aufwands wendet man sinnvollerweise die Reihenentwicklung (5.6) an:

$$\mathbf{B}^* = (\mathbf{1} - \mathbf{A}^*)^{-1} = \mathbf{1} + \mathbf{A}^* + (\mathbf{A}^*)^2 + (\mathbf{A}^*)^3 + \dots$$

Nun gilt

$$\begin{aligned} (\mathbf{A}^*)^2 &= \mathbf{F} \mathbf{A} \mathbf{F}^{-1} \mathbf{F} \mathbf{A} \mathbf{F}^{-1} = \mathbf{F} \mathbf{A}^2 \mathbf{F}^{-1}, \\ (\mathbf{A}^*)^3 &= \mathbf{F} \mathbf{A} \mathbf{F}^{-1} \mathbf{F} \mathbf{A}^2 \mathbf{F}^{-1} = \mathbf{F} \mathbf{A}^3 \mathbf{F}^{-1}, \\ \vdots &= \vdots \end{aligned}$$

“nichtkonventioneller” Förderungsmethoden wie Öl aus Teersänden, welche deutlich mehr Energie und CO₂ Emissionen pro Liter implizieren. Dies würde aber im Rahmen des IOM durch eine Änderung der \mathbf{A} -Matrixelemente berücksichtigt werden.

5 Ökonometrische Verflechtungsmodelle

also (unter Berücksichtigung der Assoziativität)

$$\begin{aligned}\mathbf{B}^* &= \sum_{j=0}^{\infty} (\mathbf{A}^*)^j \\ &= \sum_{j=0}^{\infty} \mathbf{F} \mathbf{A}^j \mathbf{F}^{-1} \\ &= \mathbf{F} \left(\sum_{j=0}^{\infty} \mathbf{A}^j \right) \mathbf{F}^{-1} \\ &= \mathbf{F} \mathbf{B} \mathbf{F}^{-1}.\end{aligned}$$

Schließlich müssen als Folge der Preisänderungen alle Spalten j der Emissionsfaktormatrix mit $1/f_j$ multipliziert werden, da man bei Preissteigerungen pro Euro weniger vom Gut j produziert/verbraucht und damit auch weniger pro Euro emittiert:

$$\mathbf{D}^* = \mathbf{D} \mathbf{F}^{-1}$$

Damit ergibt sich nach (5.16) der Emissionsvektor nach Preissteigerungen zu

$$\begin{aligned}e^* - e^{\text{dir}} &= (\mathbf{D}^*) \mathbf{B}^* \mathbf{y}^* \\ &= \mathbf{D} \mathbf{F}^{-1} \mathbf{F} \mathbf{B} \mathbf{F}^{-1} \mathbf{F} \mathbf{y} \\ &= \mathbf{D} \mathbf{B} \mathbf{y}.\end{aligned}$$

Die Emissionen sind also, wie gefordert, invariant bezüglich Preisänderungen, obwohl sich alle beteiligten Größen ändern.

5.5 Dynamische Verflechtungsmodelle



Das klassische Verflechtungsmodell setzt instantane Anpassungen der Leistungen der einzelnen Sektoren an eine veränderte Nachfrage voraus, so dass zu jedem Zeitpunkt das System bezüglich der variablen Nachfrage im Gleichgewicht ist, also durch (5.1) beschrieben wird. Dies ist im Allgemeinen unrealistisch:

- Für höhere Produktion muss man neue Mitarbeiter einstellen, bei geringer Nachfrage entlassen,
- Maschinen und andere Investitionsgüter können ebenfalls nicht “über Nacht” besorgt oder verkauft/abgeschrieben werden. Während einer Übergangszeit muss deshalb eine gestiegene Nachfrage durch Abbau von Lagerbeständen befriedigt werden.

5 Ökonometrische Verflechtungsmodelle

Bestenfalls können die Verantwortlichen über die *Steigerungsrate* $\frac{dx_i}{dt}$ des Produktionsausstoßes disponieren. Ist das System nachfrageorientiert, spielt also eine eventuelle Knappheit an Zulieferprodukten oder Rohstoffen keine Rolle, wird die Produktion hochgefahren (die Steigerungsrate ist positiv), wenn die Diskrepanz zwischen der externen Nachfrage y_i und dem beim momentanen Systemzustand für diese Nachfrage im Fließgleichgewicht zur Verfügung stehenden Angebot $y_i^{\text{eff}} = x_i - \sum_j A_{ij}x_j$ positiv ist: "Die Nachfrage übersteigt das Angebot". Andernfalls wird die Produktion heruntergefahren.

Im einfachsten Fall wird dies linear modelliert und für jeden Sektor eine charakteristische Anpassungszeit τ_i angenommen, welche in der Größenordnung von Monaten oder Jahren liegt und von der "Schnellebigkeit" der Investitionsgüter des jeweiligen Sektors abhängt. Dies führt zu folgendem linearen, dynamischen und gekoppelten Modell¹¹

$$\frac{dx_i}{dt} = \frac{1}{\tau_i} \left[y_i - \left(x_i - \sum_{j=1}^n A_{ij}x_j \right) \right] = \frac{1}{\tau} (y_i - y_i^{\text{eq}}) \quad (5.19)$$

In der Matrix-Vektor-Notation wird dies mit der Anpassungszeitmatrix

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_1 & 0 & \cdots & \\ 0 & \tau_2 & & \vdots \\ \vdots & & \ddots & 0 \\ \cdots & \cdots & 0 & \tau_n \end{pmatrix} \quad (5.20)$$

zu der kompakten Modelldarstellung

$$\boldsymbol{\tau} \cdot \frac{d\mathbf{x}}{dt} = \mathbf{y} - (\mathbf{1} - \mathbf{A}) \cdot \mathbf{x}. \quad (5.21)$$

Eigenschaften

- Bei konstanter Nachfrage, $\frac{dy_i}{dt} = 0$, nähert sich das System dem durch das klassische statische Modell (5.1) gekennzeichnete Gleichgewicht nach einer Zeit an, welche in etwa der längsten Einzelzeitkonstante, $\max(\tau_i)$, entspricht.
- Bei einen und zwei Sektoren sind keine Schwingungen möglich, auch keine gedämpften.

Mathematisch stellt das Modell ein System von linearen Differenzialgleichungen dar, welches mit den Methoden der linearen Algebra unschwer, aber häufig mühsam lösbar ist. Im Rahmen dieser Vorlesung werden allgemeine analytische Lösungen nur für der eindimensionalen Fall vorgestellt und ansonsten einige fertig gerechnete zweisektorale Lösungen als Beispiel vorgestellt.

¹¹Dies ist nicht das klassische dynamische Leontief-Modell (LIOM), welches nur bei exponentiellem Wachstum funktioniert und ansonsten unrealistische instabile Lösungen liefert. Für das LIOM, siehe z.B. Holub/Schnabl, "Input-Output-Rechnung", Oldenbourg (1994), S. 556 ff, insbesondere S. 592.

5 Ökonometrische Verflechtungsmodelle

5.5.1 Spezialfall: Dynamisches Ein-Sektor-Modell

Im dynamischen Ein-Sektor-Modell gibt es nur einen einzigen Waren und Dienstleistungen produzierenden Sektor (welcher die gesamte Wirtschaft darstellt), sowie die Verbraucher. Mit $x_1 = x$, $y_1 = y$ und $A_{11} = a$ reduziert sich das Modell (5.19) zu

$$\frac{dx(t)}{dt} = \frac{1}{\tau} [y(t) - (1 - a)x(t)].$$

Setzt man den eindimensionalen Koeffizient des vollen Aufwandes $b = 1/(1 - a)$ sowie $\tilde{\tau} = b\tau$ und $\tilde{y} = by$, so wird dies zu

$$\frac{dx(t)}{dt} = \frac{1}{\tilde{\tau}} [\tilde{y}(t) - x(t)]. \quad (5.22)$$

Hat man zur Zeit $t = 0$ den Ausgangszustand $x(0) = x_0$, so ist die Lösung dieser Differenzialgleichung in Abhängigkeit der Nachfragekurve $\tilde{y}(t) = by(t)$ für $t \geq 0$ gegeben durch

$$x(t) = x_0 e^{-\frac{t}{\tilde{\tau}}} + \frac{1}{\tilde{\tau}} \int_0^t e^{-\frac{t-t'}{\tilde{\tau}}} \tilde{y}(t') dt'. \quad (5.23)$$

Insbesondere gilt für eine zur Zeit $t = 0$ sprunghaft vom Wert y_a auf y_e sich ändernde Nachfrage die Lösung

$$x(t) = (x_0 - by_e) e^{-\frac{t}{\tilde{\tau}}} + by_e. \quad (5.24)$$

Die Produktion passt sich also mit der Zeitkonstante $\tilde{\tau} = b\tau \geq \tau$ dem neuen, durch das klassische Ein-Sektor IOM gegebenen Gleichgewichtswert by_e an. Die Zeit $\tilde{\tau}$ ist etwas größer als τ , da sich die durch die erhöhte Produktion erhöhte interne Nachfrage mittels der Strategie "Steigerung proportional zur Differenz zwischen externer Nachfrage und Nettoangebot" sozusagen "herumsprechen" muss. Herrscht insbesondere bereits vor dem Sprung Gleichgewicht, $x_0 = by_a$, so ist das Ergebnis durch

$$x(t) = b \left(y_e + (y_a - y_e) e^{-\frac{t}{\tilde{\tau}}} \right)$$

gegeben (vgl. Abbildung 5.5).

5.5.2 Varianten dynamischer Verflechtungsmodelle

Begrenzung durch das Angebot

Das durch Gl. (5.19) beschriebene dynamische IOM ist nachfragebestimmt, d.h. die Produktionssteigerungen sind proportional der Differenz aus Nachfrage y_i und Nettoangebot $((\mathbf{1} - \mathbf{A}) \cdot \mathbf{x})_i$. Bisweilen ist jedoch die Produktion nicht durch die Nachfrage begrenzt, sondern durch den Mangel an Rohstoffen oder anderen benötigten Zulieferprodukten.

5 Ökonometrische Verflechtungsmodelle

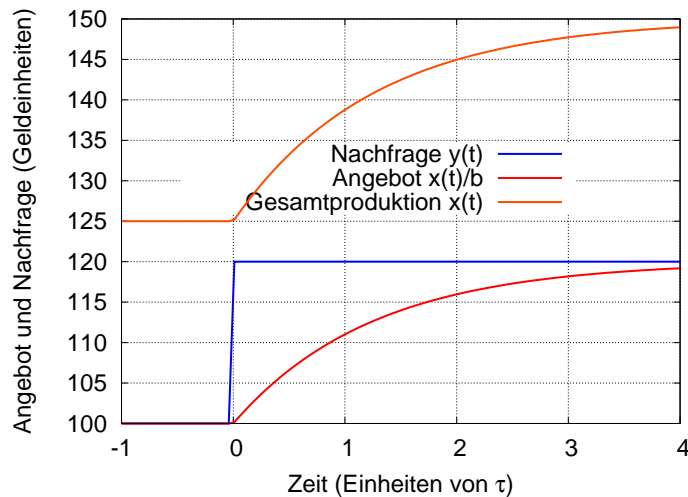


Abbildung 5.5: Dynamisches Einsektormodell: Anpassung der Gesamtproduktion und des der Nachfrage zur Verfügung Angebotes, wenn die Nachfrage sich zum Zeitpunkt $t = 0$ plötzlich um 20% ändert. ($A_{11} = a = 0.2$, $\tau = 1$).

Beispiel für zwei Sektoren Zum Herstellen von Solarzellen benötigt man Siliziumkristalle einer bestimmten Spezifikation. Obwohl es Silizium (ziemlich wortgetreu) “wie Sand am Meer” gibt, ist die Verarbeitungsstufe zum Zustand, wie er bei Solarzellen benötigt wird, nur mit aufwändigen Investitionsgütern durchzuführen. Bezeichnet man den Rohstoff verarbeitenden Sektor (Silizium) als Sektor 1 und den Sektor alternativer Energieerzeugung (Solarzellen) als Sektor 2, hat Sektor 1 also eine hohe Zeitkonstante τ_1 . Folgende, das Prinzip des Mechanismus nicht beeinflussende weitere Annahmen werden getroffen: Weder der Endverbraucher noch der Siliziumsektor benötigen selbst Silizium, $y_1 = A_{11} = 0$. Solarzellen werden nur vom Endverbraucher nachgefragt (die Sektoren benutzen für ihre Energiebedarf konventionelle Kraftwerke), also $A_{21} = A_{22} = 0$. Damit gilt im nachfragebegrenzten Zustand das dynamische IOM,

$$\begin{aligned} \tau_1 \frac{dx_1}{dt} &= -x_1 + A_{12}x_2, \\ \tau_2 \frac{dx_2}{dt} &= -x_2 + y_2. \end{aligned} \quad (5.25)$$

In der Gleichung für $\frac{dx_2}{dt}$ kommt x_1 gar nicht vor, da der Sektor x_1 die Produkte von Sektor 2 nicht benötigt, also für die Nachfrage irrelevant ist. Dass Sektor 2 wiederum das *Angebot* von Sektor 1 zu seiner Produktion benötigt, wird nicht berücksichtigt, woran man die Nachfrageorientierung dieses Ansatzes sieht.

Im angebotsbegrenzten Zustand, wie 2006 und 2007 bei den Solarzellen, ist die Produktion x_2 jedoch von der Verfügbarkeit geeigneten Siliziums begrenzt, so dass nur über kurze Zeit (Verwendung von Lagerbeständen) der Verbrauch $A_{12}x_2$ an Silizium den Zufluss $x_{12} = x_1$ (alles Silizium geht hier an die Solarzellenhersteller, da es für Sektor 1 keinen Endabnehmer gibt) überschreiten kann. Damit kann der Solarzellenausstoß x_2 nur kurzfristig (Zeitskala τ_2) den Wert x_1/A_{12} übersteigen und Gleichung (5.25) für x_2

5 Ökonometrische Verflechtungsmodelle

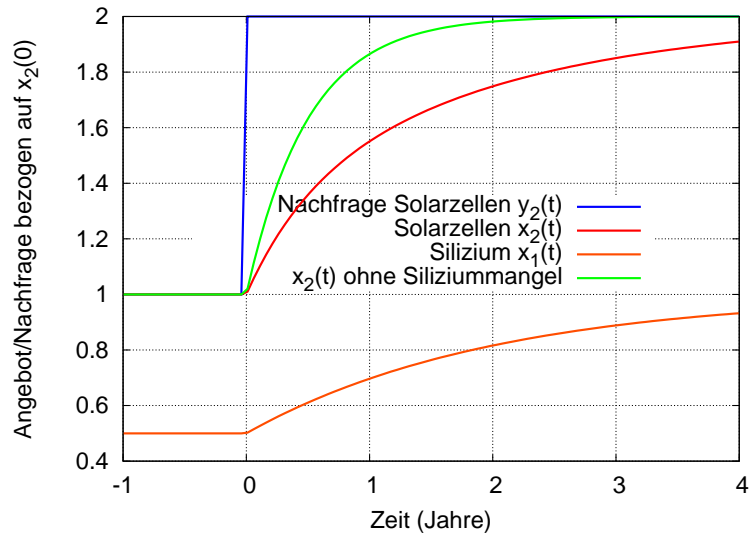


Abbildung 5.6: Reaktion der Sektoren 1 (Silizium) und 2 (Solarzellen) auf eine plötzliche Nachfragesteigerung um 100 %.

wird zu

$$\tau_2 \frac{dx_2}{dt} = -x_2 + \frac{1}{A_{12}} x_1. \quad (5.26)$$

Die Nachfrage y_2 der Endabnehmer geht nun (über das Drängen von Sektor 2 an Sektor 1, doch mehr Silizium zu erzeugen) indirekt auf Sektor 1 über, so dass für diesen Sektor, der nachfragegetrieben bleibt, nun bei der Berücksichtigung der Nachfrage von Sektor 2 nicht dessen aktuellen Produktion x_2 , sondern die nach dem statischen IOM *gewünschte* Produktion $x_2 = y_2$ relevant ist. Damit wird die Gleichung (5.25) für x_1 zu

$$\tau_1 \frac{dx_1}{dt} = -x_1 + A_{12} y_2. \quad (5.27)$$

Die kombinierte Dynamik kann jederzeit zwischen den für $y_2 \leq x_1/A_{12}$ gültigen nachfragebegrenzten Zustand (5.25) und den sonst gültigen und durch (5.26) und (5.27) definierten angebotsbegrenzten Zustand wechseln.

Zahlenbeispiel Es sei $\tau_1 = 2$ und $\tau_2 = 0.5$ (Zeiteinheiten, z.B. Jahre). Die Hälfte des Wertes der Solarzellen bestehe in den eingekauften Siliziumkristallen, also $A_{12} = 0.5$. Im Ausgangszustand sei das System bei einer Nachfrage $y_2 = 1$ im durch das statische IOM gegebenen Gleichgewicht, also $x_2 = 2x_1 = y_2 = 1$. Zum Zeitpunkt $t = 0$ steigt die Nachfrage sprunghaft von 1 auf 2. Da $\tau_1 > \tau_2$ will Sektor 2 die Produktion schneller als Sektor 1 hochfahren, so dass er in den angebotsbegrenzten Zustand gerät. Die Lösungen lauten (vgl. Abb. 5.6)

$$\begin{aligned} x_1 &= 1 - \frac{1}{2} e^{-\frac{t}{2}}, \\ x_2 &= 2 - \frac{1}{3} e^{-2t} - \frac{2}{3} e^{-\frac{t}{2}}. \end{aligned}$$

5 Ökonometrische Verflechtungsmodelle

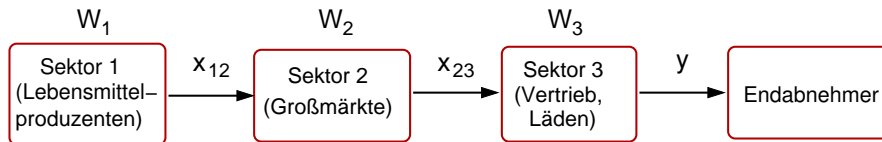


Abbildung 5.7: Materialfluss vom Produzenten über den Vertrieb zum Verbraucher.

Zum Vergleich lautet die Lösung für x_2 beim rein nachfragegetriebenen dynamischen IOM

$$x_2 = 1 - \frac{1}{2}e^{-2t}.$$

Verallgemeinerung auf mehr als zwei Sektoren

Bei mehreren Sektoren hat der Hersteller des knappen Gutes (wieder als Sektor 1 bezeichnet) im Allgemeinen die Wahl, mit welchem maximalen Anteil P_{1j} seiner Produktion er die Sektoren j und mit welchem maximalen Anteil P_{1y} er die externe Nachfrage bedient. (natürlich gilt $\sum_j P_{1j} + P_{1y} = 1$). Damit können die anderen Sektoren j ihre Brutto-Nachfrage $\sum_k A_{jk} + y_j$ nur dann bedienen, wenn diese die Zulieferbegrenzung $P_{1j}/A_{1j}x_1$ nicht übersteigt.

Generell kann sich jeder Sektor getrennt von den anderen im Nachfrage- oder Angebotsbestimmten Zustand befinden und sich dies auch jederzeit ändern, was das allgemeine System schnell unübersichtlich macht.¹²

5.5.3 Produktionsketten (Supply Chains)

Oft hat man Produktionsketten, bei denen der Endverbraucher erst über eine Reihe von Zwischenstufen erreicht wird. Da bei den Zwischenstufen (Vertrieb) nichts produziert sondern nur zwischengelagert wird sowie die Lagerkapazität endlich ist, ist es sinnvoll, die x_{ij} als Warenströme x_{ij} aufzufassen und darüber hinaus noch die Menge W_i des in jeder Stufe i vorrätigen Bestands an Waren bzw. Gütern in Lagern, Verkaufsräumen etc. als weitere dynamische Variable einzuführen. Die folgenden Überlegungen werden einfacher, wenn wir W_i nicht als die Warenmenge, sondern als *Differenz* zum angezielten Lagerbestand auffassen. Positive Werte bezeichnen also zu hohe und negative Werte zu geringe Lagerbestände.

Wir betrachten nun der Übersichtlichkeit halber die konkreten lineare Warenflussbeziehungen der Abb. 5.7. Zunächst gelten folgende unmittelbar anschauliche Kontinuitätsbeziehung zwischen den Waren W_i in den Lagern des Sektor i und den Warenströmen:

$$\frac{dW_2}{dt} = x_{12}(t) - x_{23}(t), \quad \frac{dW_3}{dt} = x_{23}(t) - y(t). \quad (5.28)$$

Des Weiteren versuchen die Warendisponenten der jeweiligen Sektoren, zu niedrige oder zu hohe Lagerbestände durch Hoch- bzw. Herunterfahren der Warenbestellungen vom

¹²Vielleicht ist es deshalb so schwierig, zu bestimmen, ob der Ansatz von Keynes (die Nachfrage ist marktbestimmend) zutrifft oder nicht.

5 Ökonometrische Verflechtungsmodelle

jeweils vorgelagerten Sektor zu kontrollieren, während die nachgelagerten Sektoren als nicht weiter beeinflussbare Nachfrager auftreten. Die Änderung des Warenflusses vom vorgelagerten Sektor kann jedoch nicht augenblicklich geschehen, da beim Hochfahren erst Ressourcen für die erhöhten Warenflüsse (bzw. beim ersten Sektor für die Produktionsrate x_1) bereitgestellt werden müssen und außerdem zwischen Bestellung und Lieferung eine endliche Zeitspanne liegt.

Damit kann man die Warenströme nur allmählich hoch- oder herunterfahren. Es wird nun folgendermaßen disponiert:

- Ist zu wenig in den Lagern ($W_i < 0$), wird der eingehende Warenstrom hochgefahren, und zwar umso mehr, je höher der Fehlbestand ist.
- Analog werden bei zu hohem Bestand die Bestellungen heruntergefahren.

In der Sprache der Mathematik lautet dies:

$$\frac{dx_{12}}{dt} = -\beta_2 W_2, \quad \frac{dx_{23}}{dt} = -\beta_3 W_3. \quad (5.29)$$

Hierbei geben die Parameter β_i an, wie schnell die Sektoren auf Bestellungen reagieren können. Leitet man nun die Gleichungen von (5.29) noch einmal nach der Zeit ab, so kann man mit (5.28) die Variablen der Warenbestände eliminieren und erhält ein dynamisches Verflechtungsmodell, was für diesen speziellen Fall **Supply-Chain-Modell** genannt wird:

$$\frac{d^2 x_{12}}{dt^2} = -\beta_2(x_{12} - x_{23}), \quad \frac{d^2 x_{23}}{dt^2} = -\beta_3(x_{23} - y). \quad (5.30)$$

Nimmt man nun an, dass die Nachfrage $y(t)$ mit der Amplitude A_y und der Periode $2\pi/\omega$ schwankt, also

$$y(t) = \bar{y} + A_y \cos \omega t,$$

und naturgemäß alle Warenströme mit derselben Periode schwanken (das System ist ja nachfragegetrieben!), so zeigt direktes Einsetzen, dass sich die Amplitude der Schwankungen aufschaukelt, wenn man in der Produktionskette (*Supply Chain*) “nach vorne”, also Richtung Hersteller geht. Insbesondere haben die Warenströme x_{23} und x_{12} die Amplituden

$$A_{23} = \left| \frac{A_y}{1 - \frac{\omega^2}{\beta_3}} \right|, \quad A_{12} = \left| \frac{A_{23}}{1 - \frac{\omega^2}{\beta_2}} \right| = \left| \frac{A_y}{\left(1 - \frac{\omega^2}{\beta_2}\right) \left(1 - \frac{\omega^2}{\beta_3}\right)} \right|.$$

Abbildung 5.8 zeigt dies am Beispiel der Supply Chain des Bieres von der Brauerei über den Großhandel zur Kneipe bzw. Biergarten bis zur durstigen Kehle des “Endverbrauchers”. Der Ausstoß der Endabnehmer schwankt wöchentlich (z.B. mit Peaks an den Freitagen oder Samstagen). Die Lieferzeit $T_3 = 2\pi/\omega_3$ des Großhandels an die Kneipe beträgt im Beispiel 5 Tage und die der Brauerei an den Großhandel $T_2 = 2\pi/\omega_2 = 9$ Tage. Man sieht, dass bei Lieferzeiten länger als die Schwankungsperiode der Nachfrage sich die Schwankungen im Vorzeichen umkehren, ansonsten nicht.

5 Ökonometrische Verflechtungsmodelle

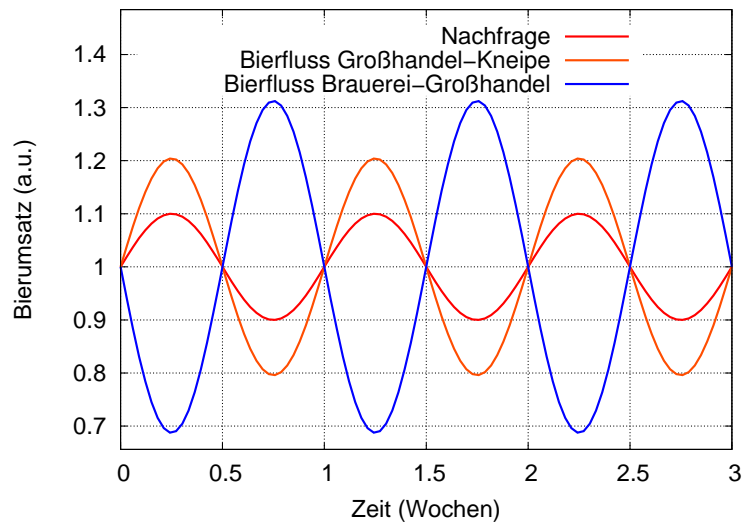


Abbildung 5.8: Der Bullwhip-Effekt bei der Distribution von Bier an die Biergärten

Diskussion

- Sind die Nachfrageschwankungen sehr langsam, ist ω sehr viel kleiner als die durch $\sqrt{\beta_i}$ charakterisierten Reaktionsschnelligkeiten der “Kettenglieder” der Supply Chain ($1/\sqrt{\beta_i}$ sind die Reaktionszeiten). Dann kommt die Produktionskette mit den Schwankungen mit und die Amplituden vergrößern sich kaum.
- Sind die Nachfrageschwankungen sehr schnell, ω also deutlich größer als alle $\sqrt{\beta_i}$, nehmen die Schwankungen ab: Ehe die Disponenten ihre Warenströme deutlich hoch- oder herunterfahren können, hat sich die Nachfrageschwankung schon ins Gegenteil verkehrt.
- Sind die Nachfrageschwankungen hingegen von der Größenordnung der Reaktionszeiten der Sektoren, ω also von ähnlicher Größe wie die $\sqrt{\beta_i}$, steigen die Schwankungen rapide an. Dies bezeichnet man auch als **Bullwhip-Effekt**.

Aufgabe: Geht es besser?

Was könnte der Lagerdisponent unternehmen, um den Bullwhip-Effekt zu vermeiden?

Antwort: (1) vorhersagbare Zyklen (im Bier-Beispiel wie der Wochenzyklus und auch ein Jahreszyklus) durch sogenannte *Ganglinien* historischer Daten antizipieren. (2) Auf unregelmäßigere Schwankungen (beispielsweise durch Schön- bzw. Schlechtwetter) antizipativer disponieren und beispielsweise die Bestellungen nicht nur nach

5 Ökonometrische Verflechtungsmodelle

dem Lagerbestand, sondern zusätzlich in Abhängigkeit der aktuellen *Änderungen* des Bestands hoch oder herunterfahren. (Dies ergibt in den Gleichungen zusätzliche "Dämpfungsterme", welche die Schwankungen mäßigen.)